Mathematical Programming (2021) 190:427–466
https://doi.org/10.1007/s10107-020-01541-x

FULL LENGTH PAPER

Series A

A new framework to relax composite functions in nonlinear programs

Taotao He[1] · **Mohit Tawarmalani**[2]

Received: 27 August 2018 / Accepted: 2 July 2020 / Published online: 13 July 2020

Abstract

In this paper, we devise new relaxations for composite functions, which improve the prevalent factorable relaxations, without introducing additional variables, by exploiting the inner-function structure. We outer-approximate inner-functions using arbitrary under- and over-estimators and then convexify the outer-function over a polytope P, which models the ordering relationships between the inner-functions and their estimators and utilizes bound information on the inner-functions as well as on the estimators. We show that there is a subset Q of P, with significantly simpler combinatorial structure, such that the separation problem of the graph of the outer-function over P is polynomially equivalent, via a fast combinatorial algorithm, to that of its graph over Q. We specialize our study to consider the product of two inner-functions with one non-trivial underestimator for each inner-function. For the corresponding polytope P, we show that there are eight valid inequalities besides the four McCormick inequalities, which improve the factorable relaxation. Finally, we show that our results generalize to simultaneous convexification of a vector of outer-functions.

Keywords Mixed-integer nonlinear programs · Factorable programming · Abstractions of outer approximations · Polynomial equivalence of separations · Computational geometry

Mathematics Subject Classification 90C11 · 90C30 · 90C26

This work was supported in part by NSF CMMI Grant 1727989.

✉ Mohit Tawarmalani
mtawarma@purdue.edu
http://web.ics.purdue.edu/~mtawarma/

Taotao He
hetaotao@sjtu.edu.cn

1 Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China

2 Krannert School of Management, Purdue University, West Lafayette, USA

1 Introduction

Mixed-integer nonlinear programs (MINLPs) are typically solved to global optimality using branch-and-bound (B&B) techniques, which construct successively tighter relaxations over refined partitions of the feasible domain [4,49]. The prevalent techniques adopted by most state-of-the-art solvers for relaxing mixed-integer nonlinear programs are inspired by the factorable programming (FP) scheme [30]. Given a library of univariate functions and their relaxations, factorable programs are MINLPs that only use functions that can be expressed as recursive sums and products of library functions. Factorable programs are relaxed by introducing a new variable to replace each univariate function, which is then required to satisfy constraints that model some relaxation of the graph of the associated univariate function. Variable products are relaxed by introducing a new variable for each product restricted to satisfy McCormick constraints, which describe the tightest convex outer-approximation of a bilinear term over a rectangle [1,30].

The following three features of the FP scheme have resulted in its widespread adoption. First, as the partition size is refined, the relaxation converges asymptotically to the original function, a property needed for its successful use in convergent B&B algorithms [24]. Second, the scheme imposes few restrictions on the types of functions that can be relaxed, besides boundedness, making it suitable for automatically relaxing large classes of MINLPs. Third, the number of variables introduced in the relaxation is in direct correspondence with the nonlinearities in the problem, and, as a result, the size of the relaxation is not much larger than the original MINLP formulation.

Nevertheless, the primary deficiency of FP is that it often produces weak relaxations [7,29]. Significant research has been devoted to improving the quality of factorable relaxations by exploiting function structure. For various types of multilinear functions [5,9,21,31,39,40], the fractional terms [46], and other useful functions [25–27,45], envelopes have been derived. Tighter relaxations and cutting planes for multilinear functions/sets have been proposed in [3,12–15,33,35,38,42]. Strong convex relaxations for structured sets have been derived in [2,6,8,20,28,34,37,44]. These techniques, however, do not directly improve the factorable programming scheme itself.

Our framework shares the key properties of the FP but produces tighter relaxations. We consider the composite function $\phi \circ (f_1, \ldots, f_d)$ where, we refer to $\phi : \mathbb{R}^d \to \mathbb{R}$ as the outer function and $(f_1, \ldots, f_d)$ as the inner-functions. Nonlinear expressions in optimization problems are typically stored using expression trees, where each node models a composition of a function and its children. Composite functions where the outer-function is a product, radical, fraction, $\exp(\cdot)$, $\log(\cdot)$, and/or a trignometric function are commonly used to model nonlinear optimization problems [11,50]. We outer-approximate the graph of $f_i(\cdot)$ with n estimators and derive bounds for estimators. Then, we relax $\phi(\cdot)$ over a polytope, P, that models the ordering relationship between the functions, their estimators, and the bounds. Although P's structure is quite complex, it has as its subset a product of simplices, Q, which captures all the interesting structure of the convex hull of the graph of $\phi(\cdot)$ over P. We give a fast combinatorial algorithm, with complexity $O(dn)$ and a separation oracle call for the graph of $\phi(\cdot)$ over Q, that solves the separation problem for the graph of $\phi(\cdot)$ over P.

Moreover, when d is constant, and the convex hull of the graph of $\phi(\cdot)$ is determined by its value at extreme points of Q, we derive an LP formulation for the separation problem, which is polynomially-sized in terms of the number of outer-approximators.

We derive the inequalities explicitly for the case when the outer-function is a bilinear term, where $d = 2$, and there is only one non-trivial underestimator for each inner-function. Instead of the four inequalities that describe the McCormick envelopes, we obtain 12 non-redundant inequalities for the bilinear term. This result shows factorable programming relaxations can be easily tightened without introducing new variables. We also show that access to a separation oracle that generates facet-defining inequalities for the graph of $\phi(\cdot)$ over Q and its projections, which are affinely isomorphic to Q and obtained by ignoring some of the estimators, can be used to generate facet-defining inequalities that separate a point from the convex hull of graph of $\phi(\cdot)$ over P. Finally, we show that these results extend to simultaneous convexification of composite functions.

Notation Throughout this paper, we use the following notation. We shall denote the convex hull of a set S by $\mathrm{conv}(S)$, the projection of S to the space of x variables by $\mathrm{proj}_x(S)$, the extreme points of S by $\mathrm{vert}(S)$, the dimension of the affine hull of S by $\dim(S)$, and the relative interior of S by $\mathrm{ri}(S)$. We denote the graph of a function f by $\mathrm{gr}(f)$, the convex (resp. concave) envelope of $f(x)$ over a convex set X by $\mathrm{conv}_X(f)$ (resp. $\mathrm{conc}_X(f)$), and the conjugate of $f(x)$ over a set X by $f^*_X(x^*)$, *i.e.*, $f^*_X(x^*) := \sup\bigl\{\langle x, x^*\rangle - f(x) \bigm| x \in X\bigr\}$.

2 Improving factorable relaxations using outer-approximations

FP expresses each function as a recursive sum and product of constituent functions, where the key step involves relaxing a product of two functions. Consider the product of two functions $f_1(x)f_2(x)$, where $f_i : X \to \mathbb{R}$ and X is a convex set in $\mathbb{R}^m$. FP assumes that, for $i = 1, 2$, there are a convex function $c_{f_i}(x)$ and a concave function $C_{f_i}(x)$ and constants f_i^L and f_i^U so that, for $x \in X$, $f_i^L \le f_i(x) \le f_i^U$ and $c_{f_i}(x) \le f_i(x) \le C_{f_i}(x)$. Assume, without loss of generality, that $f_i^L \le c_{f_i}(x)$ and $C_{f_i}(x) \le f_i^U$. Then, FP relaxes the epigraph $\bigl\{(x, \mu) \in X \times \mathbb{R} \bigm| \mu \ge f_1(x)f_2(x)\bigr\}$, using additional variables f_1 and f_2, as follows:

$$\Bigl\{(x, f, \mu) \Bigm| \mu \ge \max\bigl\{f_1^L f_2 + f_1 f_2^L - f_1^L f_2^L,\ f_1^U f_2 + f_1 f_2^U - f_1^U f_2^U\bigr\},\quad c_{f_i}(x) \le f_i \le C_{f_i}(x),\ x \in X\Bigr\}.$$

Observe that, in order to make the relationship between variables and functions transparent, we will use the same name for the variable and the function, when the variable models the graph of the function. The relaxation technique currently used in most global optimization solvers augments the above relaxation with results for specially structured problems; see [4,16,32,48].

We begin by showing that the FP relaxation of $f_1(x)f_2(x)$ can be improved using a two-step procedure. Assume that, besides bounds on $f_i(x)$, we may have access to an underestimator for $f_i(x)$. The factorable relaxation ignores this information, which, as we show next, can be used to derive tighter relaxations.

Example 1 Consider the function $x_1^2x_2^2$ over $[0,2]^2$ and, for $i=1,2$ and for $j\in J_i$, let $p(j_i)\in[0,2]$ for $j_i\in J_i$. The epigraph of x_i^2 satisfies the tangent inequality $x_i^2\geq v_{ij_i}:=p(j_i)^2+2p(j_i)\big(x_i-p(j_i)\big)$. The inequality $x_1^2x_2^2\geq v_{1j_1}(x_1)v_{2j_2}(x_2)$ does not hold, in general. For example, $x_1^2\geq 2x_1-1$ and $x_2^2\geq 2x_2-1$ but $x_1^2x_2^2\geq(2x_1-1)(2x_2-1)$ is violated at $(0,0)$. Nevertheless, if $u_{ij_i}(x)=\max\{0,v_{ij_i}(x)\}$. then $x_1^2x_2^2\geq u_{1j_1}(x)u_{2j_2}(x)$. From the definition of $u_{ij_i}(x)$, $0\leq u_{ij_i}(x)\leq p(j_i)\big(4-p(j_i)\big)\leq 4$. Let $a_{i,j_i}=p(j_i)\big(4-p(j_i)\big)$. Then, McCormick inequalities can be used to underestimate $u_{1j_1}(x)u_{2j_2}(x)$ over $[0,2]^2$ to obtain:

$$\begin{aligned}x_1^2x_2^2 &\geq u_{1j_1}(x)u_{2j_2}(x)\geq\max\big\{0,a_{1j_1}u_{2j_2}(x)+a_{2j_2}u_{1j_1}(x)-a_{1j_1}a_{2j_2}\big\}\\ &\geq a_{1j_1}v_{2j_2}(x)+a_{2j_2}v_{1j_1}(x)-a_{1j_1}a_{2j_2}.\end{aligned}$$

For $p(j_1)=p(j_2)=1$, this reduces to $x_1^2x_2^2\geq 6x_1+6x_2-15$. It can be easily verified that $x_1=1.5$, $x_2=1.5$ and $\mu=2$ satisfies the factorable relaxation, $\{(x,y,\mu)\mid 0\leq\mu,\ 4x^2+4y^2-16\leq\mu,\ 0\leq x\leq 2,\ 0\leq y\leq 2\}$, but not $\mu\geq 6x_1+6x_2-15$. $\square$

Example 1 uses a *two-step procedure*, where we first relaxed x_i^2 using its lower bound and a tangent inequality, and, in the second step, we relaxed the product of underestimators using McCormick envelopes. We concluded that the resulting inequality is not implied by using McCormick envelopes on the original product $x_1^2x_2^2$, the *one-step procedure* typically used in factorable relaxations. It may appear surprising that the first relaxation step in the two-step procedure could strengthen the relaxation obtained by directly using McCormick inequalities. However, this can be explained by the fact, that the upper bounds, a_{ij_i}, are in general strictly smaller than the upper bounds on x_i^2, and this helps strengthen the generated inequalities.

Now, we consider a generalization, where we no longer assume that the inner-functions are nonnegative. More specifically, we consider the product $f_1(x)f_2(x)$, where each $f_i(x)$ has a non-trivial underestimator $u_i(x)$ so that for all $x\in X$, $f_i^L\leq f_i(x)\leq f_i^U$ and $f_i^L\leq u_i(x)\leq a_i$. For $i=1,2$, we introduce variables f_i and u_i for functions $f_i(x)$ and $u_i(x)$, respectively.

Theorem 1 *Let $f_1^L\leq a_1\leq f_1^U$ and $f_2^L\leq a_2\leq f_2^U$. Then, consider the set:*

$$\begin{aligned}P=\big\{(u_1,f_1,u_2,f_2)\ \big|\ &f_1^L\leq u_1\leq\min\{f_1,a_1\},\ f_1\leq f_1^U,\\ &f_2^L\leq u_2\leq\min\{f_2,a_2\},\ f_2\leq f_2^U\big\}.\end{aligned}$$

The following linear inequalities are valid for the epigraph of $f_1 f_2$ over P:

$$f_1 f_2 \geq \max \left\{ \begin{aligned} & e_1 := f_1 f_2^U + f_2 f_1^U - f_1^U f_2^U \\ & e_2 := (f_2^U - a_2)u_1 + (f_1^U - a_1)u_2 + a_2 f_1 + a_1 f_2 \\ & \quad\;\; + a_1 a_2 - a_1 f_2^U - f_1^U a_2 \\ & e_3 := (f_2^U - f_2^L)u_1 + f_2^L f_1 + a_1 f_2 - a_1 f_2^U \\ & e_4 := (f_1^U - f_1^L)u_2 + a_2 f_1 + f_1^L f_2 - f_1^U a_2 \\ & e_5 := (a_2 - f_2^L)u_1 + (a_1 - f_1^L)u_2 + f_2^L f_1 + f_1^L f_2 - a_1 a_2 \\ & e_6 := f_1^L f_2 + f_1 f_2^L - f_1^L f_2^L \end{aligned} \right\}.$$

Proof We show that e_3, e_4, and e_5 are valid underestimators for $f_1 f_2$ over P using the procedure in Example 1. That e_3 underestimates $f_1 f_2$ follows from:

$$\begin{aligned} f_1 f_2 &= (f_1 - f_1^L)(f_2 - f_2^L) + f_1^L f_2 + f_2^L f_1 - f_1^L f_2^L \\ &\geq (u_1 - f_1^L)(f_2 - f_2^L) + f_1^L f_2 + f_2^L f_1 - f_1^L f_2^L \\ &\geq (u_1 - f_1^L)(f_2^U - f_2^L) + (a_1 - f_1^L)(f_2 - f_2^L) \\ &\quad - (a_1 - f_1^L)(f_2^U - f_2^L) + f_2^L f_1 + f_1^L f_2 - f_1^L f_2^L \\ &= e_3, \end{aligned}$$

where the first equality shifts f_1 and f_2 so that we relax a product of non-negative functions, the first inequality underestimates the product $(f_1 - f_1^L)(f_2 - f_2^L)$, and the second inequality uses the McCormick relaxation for $(u_1 - f_1^L)(f_2 - f_2^L)$. The derivation for e_4 is symmetric, with the role of f_1 and f_2 is interchanged. Similarly, to show that e_5 is an underestimator:

$$\begin{aligned} f_1 f_2 &= (f_1 - f_1^L)(f_2 - f_2^L) + f_1^L f_2 + f_2^L f_1 - f_1^L f_2^L \\ &\geq (u_1 - f_1^L)(u_2 - f_2^L) + f_1^L f_2 + f_2^L f_1 - f_1^L f_2^L \\ &\geq (u_1 - f_1^L)(a_2 - f_2^L) + (a_1 - f_1^L)(u_2 - f_2^L) - (a_1 - f_1^L)(a_2 - f_2^L) \\ &\quad + f_2^L f_1 + f_1^L f_2 - f_1^L f_2^L \\ &= e_5. \end{aligned}$$

Observe that e_1 and e_6 are derived by using the functions f_1 and f_2 themselves as their underestimators.

To show the second inequality is valid, define $s_2 = \max\left\{u_2, f_2^L + \frac{a_2 - f_2^L}{f_2^U - f_2^L}(f_2 - f_2^L)\right\}$ and note that $a_2 - s_2 + f_2 - f_2^U = \frac{f_2 - f_2^U}{f_2^U - f_2^L}(f_2^U - a_2) \leq 0$. Then, $f_1 f_2 = e_2 + (s_2 - u_2)(f_1^U - a_1) + (u_1 - a_1)(a_2 - s_2 + f_2 - f_2^U) + (f_1 - u_1)(f_2 - s_2) + (f_1 - f_1^U)(s_2 - a_2) \geq e_2$, completing the proof. □

If $a_i \in \{f_i^L, f_i^U\}$ for $i \in \{1, 2\}$, the factorable relaxation is the convex hull of the epigraph of $f_1 f_2$ over P. Thus, to improve the factorable relaxation using Theorem 1, we require $a_i < f_i^U$ for at least some i. Observe that the inequalities in Theorem 1 share many properties of the FP relaxations. First, they apply to all factorable programming problems. Second, the coefficients of u_i in the inequalities are non-negative. Therefore, u_i can be substituted with any convex underestimator of their defining relation, $u_i(x)$, to yield convex inequalities. Third, if there are n_i estimators of $f_i(x)$, Theorem 1 yields $2n_1n_2+n_1+n_2+2$ inequalities underestimating $f_1 f_2$ instead of the two inequalities typically used in FP.

Observe that the underestimator e_2 is not obtained using the two-step procedure in Example 1. In the next example, we demonstrate that underestimator e_2 is not dominated by other inequalities obtained using the two-step procedure. In other words, existence of e_2 shows that the epigraph of $f_1 f_2$ over P satisfies inequalities besides those obtained using the two-step procedure.

Example 2 Consider the monomial $x_1^2x_2^2$ over $[0, 2]^2$. Then, let $u_i = \max\{0, 2x_i - 1\}$. Then, Theorem 1 yields the following relaxation, after substitutions:

$$\left\{ (x, y, \mu) \in [0, 2]^2 \times \mathbb{R}_+ \;\middle|\; \begin{array}{l} \mu \geq e_1 = 4x_1^2 + 4x_2^2 - 16 \\ \mu \geq e_2 = 2x_1 + 2x_2 + 3x_1^2 + 3x_2^2 - 17 \\ \mu \geq e_3 = 8x_1 + 3x_2^2 - 16 \\ \mu \geq e_4 = 3x_1^2 + 8x_2 - 16 \\ \mu \geq e_5 = 6x_1 + 6x_2 - 15 \\ \mu \geq e_6 = 0 \end{array} \right\}.$$

Observe that $\mu \geq e_3$, $\mu \geq e_4$, and $\mu \geq e_5$ are the inequalities obtained by first underestimating x_1^2 and x_2^2 and then using the McCormick inequalities. However, the inequality $\mu \geq e_2$ is not obtained using the two-step procedure and is not redundant. For example, at $(x_1, x_2) = (1.6, 1.6)$, the highest underestimator is e_2 which equals 4.76, while the remaining underestimators are below 4.5. We remark here that the new inequalities we have derived still fail to describe the convex envelope of $x_1^2x_2^2$ over $[0, 2]^2$ since, at $(x_1, x_2) = (1.6, 1.6)$, a convex underestimator $9.6x_1 + 9.6x_2 - 24.96$ equals 5.76 while as mentioned, the tightest inequality in the set above evaluates to 4.76 at this point. For completeness, we show that $x_1^2x_2^2 \geq 9.6x_1 + 9.6x_2 - 24.96$. First, assume that $x_1x_2 + 2x_1 + 2x_2 \geq 4$. Then, it follows that $x_1^2x_2^2 - 9.6x_1 - 9.6x_2 + 24.96 = (x_1x_2 + 2x_1 + 2x_2 - 4)(x_1 - 2)(x_2 - 2) + (2x_1 + 2x_2 - 6.4)^2 \geq 0$. Now, consider the case when $x_1x_2 + 2x_1 + 2x_2 < 4$. Then, $x_1^2x_2^2 - 9.6x_1 - 9.6x_2 + 24.96 = x_1^2x_2^2 - 4.8(x_1x_2 + 2x_1 + 2x_2 - 4) + 4.8x_1x_2 + 5.76 \geq 0$. Therefore, we have $x_1^2x_2^2 \geq 9.6x_1 + 9.6x_2 - 24.96$. □

In the next example, we illustrate that inequalities obtained using Theorem 1 are not implied by reformulation-linearization techniques (RLT) [41].

Example 3 Consider the following system of valid inequalities:

$$\left. \begin{array}{rr} x_i \geq 0 & 2 - x_i \geq 0 \\ x_i^2 \geq 0 & 4 - x_i^2 \geq 0 \\ x_i^2 - (2x_i - 1) \geq 0 & 3 - (2x_i - 1) \geq 0 \end{array} \right\} \quad i = 1, 2. \tag{1}$$

Clearly, applying RLT over these inequalities yields a relaxation for $x_1^2x_2^2$ over $[0, 2]^2$. However, RLT does not exploit the inequality $\max\{0, 2x_i - 1\} \le x_i^2$, which we used in Examples 1 and 2 to derive $6x_1 + 6x_2 - 15 \le \mu$. Therefore, it is perhaps not surprising that RLT is unable to generate this inequality. Moreover, minimizing an affine function $\mu - (6x_1 + 6x_2 - 15)$ over the degree-4 RLT relaxation of (1) yields -0.36, while it has been shown that $\mu - (6x_1 + 6x_2 - 15) \ge 0$ is valid for the epigraph of $x_1^2x_2^2$ over $[0, 2]^2$. Perhaps, more importantly, the inequality $\mu \ge 2x_1 + 2x_2 + 3x_1^2 + 3x_2^2 - 17$ does not have a two-step derivation and exploits $2y - 1 \le s_2 := \max\left\{2y - 1, \frac{3}{4}y^2\right\} \le 3$ in the derivation, a construction that will play a central role in the subsequent development. □

3 A relaxation framework for composite functions

In this section, we introduce a generalized version of the setup in Theorem 1, which will be subject of study for most of the remaining paper. Our setup will generalize that of Theorem 1 in the following way. First, we replace the bilinear term with an arbitrary function, $\phi : \mathbb{R}^d \to \mathbb{R}$, and consider relaxations of $\phi \circ f : X \subseteq \mathbb{R}^m \to \mathbb{R}$, where $f : X \to \mathbb{R}^d$ is a vector of *bounded* functions over X. We shall write $f(x) := \big(f_1(x), \dots, f_d(x)\big)$ and refer to $f(\cdot)$ as inner-functions while $\phi(\cdot)$ will be referred to as the outer-function. Second, we allow a vector of under- and over-estimators, instead of one underestimator, for each inner-function $f_i(\cdot)$. Last, we will derive convex relaxations for the graph of $\phi \circ f$ (instead of just the epigraph). Formally, we will relax the set:

$$\operatorname{gr}(\phi \circ f) = \left\{(x, \phi) \;\middle|\; \phi = \phi\big(f(x)\big),\ x \in X\right\}.$$

3.1 Polyhedral abstraction of outer-approximation

In this subsection, we formally generalize the construction in Theorem 1. First, instead of using a single underestimator for each function, we consider a vector of bounded under- and over-estimators for the inner-function $f_i(\cdot)$. More specifically, let $(n_1, \dots, n_d) \in \mathcal{Z}^d$, and consider a vector of functions $u : \mathbb{R}^m \to \mathbb{R}^{\sum_{i=1}^d (n_i+1)}$ defined as $u(x) = \big(u_1(x), \dots, u_d(x)\big)$, where $u_i(x) : \mathbb{R}^m \to \mathbb{R}^{n_i+1}$, and consider a vector $a = (a_1, \dots, a_d) \in \mathbb{R}^{\sum_{i=1}^d (n_i+1)}$, where $a_i \in \mathbb{R}^{n_i+1}$. Moreover, for each $i \in \{1, \dots, d\}$ and for every $x \in X$, the pair $\big(u_i(x), a_i\big)$ is assumed to satisfy

$$\begin{aligned}
&u_{i0}(x) = a_{i0}, \quad a_{i0} \le u_{in_i}(x) = f_i(x) \le a_{in_i},\\
&\text{for each } j \in A_i : \quad u_i^L \le u_{ij}(x) \le \min\big\{f_i(x), a_{ij}\big\},\\
&\text{for each } j \in B_i : \quad \max\big\{f_i(x), a_{ij}\big\} \le u_{ij}(x) \le u_i^U,\\
&u_i^L \le a_{i0} \le \cdots a_{in_i} \le u_i^U,
\end{aligned} \tag{2}$$

where the pair (A_i, B_i) is a partition of $\{0, \dots, n_i\}$ so that $\{0, n_i\} \subseteq A_i$. The first requirement states that $u_{in_i}(\cdot)$ is the inner-function $f_i(\cdot)$, which is bounded from below (resp. above) by a_{i0} (resp. a_{in_i}), and $u_{i0}(\cdot)$ is a constant function matching the lower bound a_{i0}. The second (resp. third) requirement states that for $j \in A_i$ (resp. $j \in B_i$), $u_{ij}(\cdot)$ is an underestimator (resp. overestimator) for $f_i(\cdot)$, which is bounded from above (resp. below) by a_{ij} and bounded from below (resp. above) by u_i^L (resp. u_i^U). The last requirement that the elements of a_i are ordered in a non-decreasing order and contained within $[u_i^L, u_i^U]$ is only for notational convenience. Notice that we do not explicitly specify different lower bounds for the underestimators and upper bounds for the overestimators. This is because they are not important in constructing the relaxations and do not change the quality of the relaxations. In particular, we show in Proposition 1 that, without loss of generality, u_i^L and u_i^U can be set to be the lower bound a_{i0} and the upper bound a_{in_i} of $f_i(\cdot)$ respectively. We also remark that, for $j \in A_i$ (resp. $j \in B_i$), a_{ij} need not be the tightest possible upper (resp. lower) bound of $u_{ij}(\cdot)$ over X although better bounds improve the quality of the relaxation. We mention that there are several techniques to derive bounds for expressions defining inner-functions and estimators including bound tightening and relaxation techniques; see [17,36] for example. Last, although the number of estimators for each function can be different, we will assume without loss of generality and for notational simplicity that $n_1 = \cdots = n_d$. Since all n_i are equal, we will use n to denote n_i.

We now formally describe a generalization of P that was defined as in Theorem 1. The polytope P is denoted, in general, as

$$P(a, u^L, u^U, B) := \prod_{i=1}^{d} P_i(a_i, u_i^L, u_i^U, B_i),$$

where $B := \prod_{i=1}^{d} B_i$, $u^L := (u_1^L, \dots, u_d^L)$ and $u^U := (u_1^U, \dots, u_d^U)$, and

$$P_i(a_i, u_i^L, u_i^U, B_i) := \left\{ u_i \in \mathbb{R}^{n+1} \;\middle|\; \begin{array}{l} \forall j \in A_i : u_{ij} \le u_{in},\ u_i^L \le u_{ij} \le a_{ij} \\ \forall j \in B_i : u_{in} \le u_{ij},\ a_{ij} \le u_{ij} \le u_i^U \\ u_{i0} = a_{i0},\ a_{i0} \le u_{in} \le a_{in} \end{array} \right\}. \tag{3}$$

We will typically not write the arguments of P_i and P since they will be apparent from the context. We will refer to the polytope P_i as the *abstraction* of outer-approximators of function $f_i(\cdot)$ and P as the *abstraction* of outer-approximators for the vector of functions $f(\cdot)$. Essentially, the polytope P_i is obtained by introducing a variable u_{ij} for the estimator $u_{ij}(\cdot)$ and replacing the ordering relationships between the functions with those between the introduced variables.

Last, we consider the graph, Φ^P, of the outer-function $\phi(u_{1n}, \dots, u_{dn})$ over the abstraction P given as follows:

$$\Phi^P = \big\{ (u, \phi) \;\big|\; \phi = \phi(u_{1n}, \dots, u_{dn}),\ u \in P \big\}.$$

In the rest of the paper, we are interested in studying the graph Φ^P and its convex hull $\operatorname{conv}(\Phi^P)$. We remark that our construction treats estimators of inner-functions abstractly, and exploits various bounding relationships while being oblivious of the precise dependence of the inner-functions $f(\cdot)$ on x. Thus, our methods apply to general nonlinear programming problems while providing flexibility, which can be tailored to exploit specific problem structure, *e.g.* properties of the outer-function $\phi(\cdot)$.

3.2 An overview of the main polynomial-time equivalence result

Since P generalizes the standard hypercube $[0, 1]^d$, by considering the special case where $\phi(\cdot)$ is a bilinear function, it follows that there does not exist a polynomial time algorithm to solve the separation problem for Φ^P, unless P $=$ NP. But, Theorem 1 shows that special instances of this problem are solvable, sometimes in closed-form, which improve relaxations for nonlinear programs. We now describe the main structural insights, we derive for Φ^P, in this paper. We will show that separating Φ^P is polynomially equivalent to separating a simpler object Φ^Q, the graph of $\phi(\cdot)$ over a subset of P, which is termed Q here onwards and described formally in Sect. 4. We remark that this equivalence is not just a mapping of hard instances of Φ^P to hard instances of Φ^Q. Clearly, existence of such a mapping follows from NP-Hardness of these problems. Rather, this equivalence is derived from an algorithm that separates an instance of Φ^P using an oracle to separate a specific related instance of Φ^Q.

To show this polynomial equivalence, we devise a separation oracle for Φ^P by augmenting the separation oracle for Φ^Q with a fast-polynomial-time combinatorial algorithm. The key ingredient of the combinatorial algorithm is a lifting procedure that lifts Φ^P into a higher dimensional space. We show such lifting procedure is equivalent to solving d two-dimensional convexification problems separately. A direct consequence of our results will be that for any fixed number of inner-functions d, the problem of separating $\mathrm{conv}(\Phi^P)$, when its extreme points project to the extreme points of P, is polynomially solvable in the number of estimators n. In contrast, a direct LP-based separation formulation of $\mathrm{conv}(\Phi^P)$ would be exponential because, even for a fixed d, the number of extreme points of P is exponential in n. This result is interesting for practice because compositions often involve only a few functions or can be recursively decomposed as such.

The resulting algorithm has many interesting features besides tractability. Assume, for the purpose of illustration, that $\phi(\cdot)$ is a multilinear function, which would imply that $\mathrm{conv}(\Phi^P)$ and $\mathrm{conv}(\Phi^Q)$ are polyhedral sets [39]. If we have access to a separation oracle that separates $\mathrm{conv}(\Phi^Q)$ by generating facet-defining inequalities and we assume that this separation oracle can also be used when some of the estimators are dropped from the construction, our combinatorial algorithm can be used to separate points from $\mathrm{conv}(\Phi^P)$ using facet-defining inequalities. Moreover, the polynomial equivalence carries over to the problem of simultaneously convexifying a collection of functions and, so does the property of generating facet-defining inequalities.

3.3 Projecting out introduced estimator variables

We mentioned in the discussion following Theorem 1 that u_i variables can be replaced with convex functions, and illustrated this procedure in Example 2. To project out introduced variables, u_1 and u_2, we relied on the fact that u_1 and u_2 have non-negative coefficients in the inequalities derived in Theorem 1. In this subsection, we show a similar result for $\text{conv}(\Phi^P)$, and will utilize it to eventually substitute the underestimator (resp. overestimator) variables u_{ij} with convex functions (resp. concave functions).

The convex hull of Φ^P is the intersection of the epigraph of the convex envelope and hypograph of concave envelope of $\phi(u_{1n}, \ldots, u_{dn})$ over P, that is,

$$\text{conv}(\Phi^P) = \Bigl\{(u, \phi) \Bigm| \text{conv}_P(\phi)(u) \leq \phi \leq \text{conc}_P(\phi)(u),\ u \in P\Bigr\}. \tag{4}$$

Observe that although the function $\phi(\cdot)$ depends only on $(u_{1n}, \ldots, u_{dn})$, the functions $\text{conv}_P(\phi)$ and $\text{conc}_P(\phi)$ depend on all the variables u. In the following lemma, we consider the concave envelope $\text{conc}_P(\phi)(\cdot)$ and establish certain monotonicity properties for it. A similar property, albeit reversed in direction, can be obtained for $\text{conv}_P(\phi)(\cdot)$ since $\text{conv}_P(\phi)(u) = -\text{conc}_P(-\phi)(u)$ for every $u \in P$.

Lemma 1 *If the concave envelope* $\text{conc}_P(\phi)(u)$ *is closed then it is non-increasing in* u_{ij} *for all* i *and* $j \in A_i \backslash \{0, n\}$ *and non-decreasing in* u_{ij} *for all* i *and* $j \in B_i$.

Proof We will only prove that $\text{conc}_P(\phi)(u)$ is non-increasing in u_{ij} for all i and $j \in A_i \backslash \{0, n\}$ since a similar argument shows that it is non-decreasing in u_{ij} for all i and $j \in B_i$. Let $\phi \leq \langle \alpha, u \rangle + b$ be a valid inequality of $\text{conc}_P(\phi)(u)$. Let $J_i := \{j' \in A_i \backslash \{0, n\} \mid \alpha_{ij'} > 0\}$. By considering $(\tilde{\alpha}, b')$, where $\tilde{\alpha}_{ij} = 0$ for all i and $j \in J_i$, $\tilde{\alpha}_{ij} = \alpha_{ij}$ otherwise, and $b' = b + \sum_{i=1}^{d} \sum_{j \in J_i} \alpha_{ij} a_i^L$, it is easy to construct a valid inequality $\phi \leq \langle \tilde{\alpha}, u \rangle + b'$ of $\text{conc}_P(\phi)(u)$ such that $\tilde{\alpha}_{ij} \leq 0$ for all i and $j \in A_i \backslash \{0, n\}$, and $\langle \tilde{\alpha}, u \rangle + b' \leq \langle \alpha, u \rangle + b$ for every $u \in P$.

Now assume that $\text{conc}_P(\phi)(u)$ is closed. We prove that $\text{conc}_P(\phi)(u) = \psi(u)$, where $\psi(u) := \inf_\alpha \bigl\{\langle \alpha, u \rangle + (-\phi)_P^*(-\alpha) \bigm| \alpha_{ij} \leq 0\ \forall i \in \{1, \ldots, d\}\ \forall j \in A_i \backslash \{0, n\}\bigr\}$ and $(-\phi)_P^*$ denotes the Fenchel conjugate of $-\phi(u_{1n}, \ldots, u_{dn})$ with its domain restricted to P. This will show what we seek to prove since, by definition, $\psi(u)$ is non-increasing in u_{ij} for all i and $j \in A_i \backslash \{0, n\}$, being the infimum over α of linear functions $\langle \alpha, u \rangle + (-\phi)_P^*(-\alpha)$, all of which satisfy this property. Since $\text{conc}_P(\phi)(u)$ is assumed to be closed, by Theorem 1.3.5 in [23], we have

$$\text{conc}_P(\phi)(u) = \inf_\alpha \bigl\{\langle \alpha, u \rangle + (-\phi)_P^*(-\alpha)\bigr\}. \tag{5}$$

It follows that $\psi(u) \geq \text{conc}_P(\phi)(u)$ because $\psi(u) \geq \inf_\alpha \bigl\{\langle \alpha, u \rangle + (-\phi)_P^*(-\alpha)\bigr\}$. To show $\text{conc}_P(\phi)(u) \geq \psi(u)$, we consider a point $\bar{u} \in P$. By (5), there exists a sequence $\bar{\alpha}^k$ so that the inequality $\text{conc}_P(\phi)(u) \leq \langle \bar{\alpha}^k, u \rangle + (-\phi)_P^*(-\bar{\alpha}^k)$ is valid for all k and $\lim_{k \to \infty} \langle \bar{\alpha}^k, \bar{u} \rangle + (-\phi)_P^*(-\bar{\alpha}^k) = \text{conv}_P(\phi)(\bar{u})$. If $\bar{\alpha}_{ij}^k > 0$ for some i and $j \in A_i \backslash \{0, n\}$, we have shown that there exists a valid inequality $\phi \leq \langle \tilde{\alpha}^k, u \rangle + b'$ of $\text{conc}_P(\phi)(u)$ such that $\langle \tilde{\alpha}^k, u \rangle + b' \leq \langle \bar{\alpha}^k, u \rangle + (-\phi)_P^*(-\bar{\alpha}^k)$ for all $u \in P$ and $\tilde{\alpha}_{ij}^k \leq 0$ for all i and $j \in A_i \backslash \{0, n\}$. Therefore, we have

$$\begin{aligned}
\text{conc}_P(\phi)(\bar{u}) &= \lim_{k \to \infty} \langle \bar{\alpha}^k, \bar{u} \rangle + (-\phi)_P^*(-\bar{\alpha}^k) \\
&\geq \lim_{k \to \infty} \langle \tilde{\alpha}^k, \bar{u} \rangle + b'
\end{aligned}$$

$$\geq \lim_{k\to\infty} \langle \tilde{\alpha}^k, \bar{u}\rangle + (-\phi)^*_P(-\tilde{\alpha}^k) \geq \psi(\bar{u}),$$

where first equality and first inequality are established above, second inequality follows because the validity of $\phi(u_{1n}, \ldots, u_{dn}) \leq \langle \tilde{\alpha}^k, u\rangle + b'$ for $u \in P$ implies $(-\phi)^*_P(-\tilde{\alpha}^k) = \sup_{u\in P}\big\{\phi(u) - \langle \tilde{\alpha}^k, u\rangle\big\} \leq b'$, and the last inequality holds since $\tilde{\alpha}^k \leq 0$ for all i and $j \in A_i\backslash\{0, n\}$ implies $\tilde{\alpha}$ is feasible in the optimization problem defining $\psi(\bar{u})$. Therefore, the proof is complete. □

We now use monotonicity of $\mathrm{conc}_P(\phi)(u)$ and $\mathrm{conv}_P(\phi)(u)$ to construct relaxations for $\mathrm{gr}(\phi \circ f)$ in the space of $(x, \phi, u_{\cdot n})$ variables, where $u_{\cdot n} = (u_{1n}, \ldots, u_{dn})$, by substituting the convex underestimators and concave overestimators with their defining relationships in the cuts valid for $\mathrm{conv}(\Phi^P)$.

Theorem 2 *Let $\phi \circ f$ be a composite function, where $\phi : \mathbb{R}^d \to \mathbb{R}$ is a continuous function and $f : \mathbb{R}^m \to \mathbb{R}^d$ is a vector of functions which are bounded over $X \subseteq \mathbb{R}^m$. Given a pair $\big(a, u(x)\big)$ satisfying (2), we have $\mathrm{gr}(\phi \circ f) \subseteq \mathrm{proj}_{(x,\phi)}(R)$, where*

$$R := \left\{(x, u_{\cdot n}, \phi) \;\middle|\; \begin{array}{l} \big(\tilde{u}_1(x, u_{1n}), \ldots, \tilde{u}_d(x, u_{dn}), \phi\big) \in \mathrm{conv}(\Phi^P) \\ u_{\cdot n} = f(x),\ x \in X, \end{array}\right\} \tag{6}$$

and $\tilde{u}_i(x, u_{in}) = \big(u_{i0}(x), \ldots, u_{i(n-1)}(x), u_{in}\big)$. The relaxation is convex if $u_{ij}(x)$ is convex for $j \in A_i\backslash\{n\}$ and concave for $j \in B_i$ and $\big\{(x, u_{\cdot n}) \in X \times \mathbb{R} \;\big|\; u_{\cdot n} = f(x)\big\}$ is outer-approximated by a convex set.

Proof See "Appendix A.1". □

3.4 Simplifying the structure of polyhedral abstraction

We start with simplifying $P(a, u^L, u^U, B)$ to show that it suffices to consider the case where $u^L = a_{\cdot 0}$ and $u^U = a_{\cdot n}$ to treat the general case, where $a_{\cdot 0} := (a_{10}, \ldots, a_{d0})$ and $a_{\cdot n} := (a_{1n}, \ldots, a_{dn})$.

Proposition 1 *Define u' so that $u'_{ij} = \min\big\{\max\{u_{ij}, a_{i0}\}, a_{in}\big\}$. Then, for any $u \in P(a, u^L, u^U, B)$ and $\phi \in \mathbb{R}$, $(u, \phi) \in \mathrm{conv}\big(\Phi^{P(a,u^L,u^U,B)}\big)$ if and only if $(u', \phi) \in \mathrm{conv}\big(\Phi^{P(a,a_{\cdot 0},a_{\cdot n},B)}\big)$.*

Proof Let $P := P(a, u^L, u^U, B)$ and $P' := P(a, a_{\cdot 0}, a_{\cdot n}, B)$. In the following, we will show that $\mathrm{conc}_P(\phi)(u) = \mathrm{conc}_{P'}(\phi)(u')$. By considering $-\phi(\cdot)$, a similar argument shows that $\mathrm{conv}_P(\phi)(u) = \mathrm{conv}_{P'}(\phi)(u')$, completing the proof.

First, we argue that $\mathrm{conc}_P(\phi)(u) = \mathrm{conc}_P(\phi)(u')$. For any $j \in B_i$, $u_{ij} \geq u_{in} \geq a_{i0}$. Therefore, $u'_{ij} \leq u_{ij}$. Similarly, for $j \in A_i$, $\max\{u_{ij}, a_{i0}\} \leq u_{in} \leq a_{in}$. Therefore, $u'_{ij} \geq u_{ij}$, and, in particular, $u'_{i0} = u_{i0}$ and $u'_{in} = u_{in}$. It follows from Lemma 1 that $\mathrm{conc}_P(\phi)(u') \leq \mathrm{conc}_P(\phi)(u)$. Now, we argue that $\mathrm{conc}_P(\phi)(u') \geq \mathrm{conc}_P(\phi)(u)$. Let $J_i(u) = \{j \mid u_{ij} < a_{i0}\}$ and $K_i(u) = \{j \mid u_{ij} > a_{in}\}$. We perform induction on $\sum_{i=1}^d\big(|J_i(u)|+|K_i(u)|\big)$. The base case is trivial because $\sum_{i=1}^d\big(|J_i(u)|+|K_i(u)|\big) = 0$

implies $u' = u$ and the inequality is trivially satisfied. Let i' be such that there exists a $j' \in J_{i'}(u) \cup K_{i'}(u)$. Since $a_{i'0} \le u_{i'n} \le a_{i'n}$, it follows that $j' \ne n$. We assume $j' \in J_{i'}(u)$ as a similar argument applies when $j' \in K_{i'}(u)$. Since $u \in P$, by the definition of $\operatorname{conc}_P(\phi)$, there exist convex multipliers λ^k and points $(u^k, \phi^k) \in \Phi^P$ such that $\big(u, \operatorname{conc}_P(\phi)(u)\big) = \sum_k \lambda^k(u^k, \phi^k)$. Define $\bar{u}^k_{i'j'} = a_{i'0}$ and $\bar{u}^k_{ij} = u^k_{ij}$ otherwise. Since $j' \ne n$, it can be verified easily that $(\bar{u}^k, \phi^k) \in \Phi^P$. Then, define $\big(\tilde{u}, \operatorname{conc}_P(\phi)(u)\big) = \sum_k \lambda^k(\bar{u}^k, \phi^k)$ and observe that the representation shows that $\operatorname{conc}_P(\phi)(\tilde{u}) \ge \operatorname{conc}_P(\phi)(u)$, $\tilde{u}_{i'j'} = a_{i'0}$, and $\tilde{u}_{ij} = u_{ij}$ otherwise. However, since $j' \notin J_{i'}(\tilde{u})$, it follows that $\sum_{i=1}^d \big(|J_i(\tilde{u})| + |K_i(\tilde{u})|\big) = \sum_{i=1}^d \big(|J_i(u)| + |K_i(u)|\big) - 1$. Therefore, it follows that $\operatorname{conc}_P(\phi)(u) \le \operatorname{conc}_P(\phi)(\tilde{u}) \le \operatorname{conc}_P(\phi)(u')$, where the last inequality is by the induction hypothesis.

Next, we show that $\operatorname{conc}_P(\phi)(u') = \operatorname{conc}_{P'}(\phi)(u')$. Clearly, $\operatorname{conc}_{P'}(\phi)(u') \le \operatorname{conc}_P(\phi)(u')$ since $P' \subseteq P$. Let $\phi' = \operatorname{conc}_P(\phi)(u')$. Then, there exist $(u^l, \phi^l) \in \Phi^P$ and convex multipliers γ^l so that $\big(u', \phi'\big) = \sum_l \gamma^l(u^l, \phi^l)$. Define

$$\tilde{u}^l_{ij} = \min\big\{\max\{u^l_{ij}, a_{i0}\}, a_{in}\big\},$$

and let $\bar{u} = \sum_l \gamma^l \tilde{u}^l$. It follows that $(\tilde{u}^l, \phi^l) \in \Phi^{P'}$, and thus $(\bar{u}, \phi') \in \operatorname{conv}(\Phi^{P'})$. In other words, $\phi' \le \operatorname{conc}_{P'}(\phi)(\bar{u})$. Moreover, $\bar{u}_{ij} \ge u'_{ij}$ for $j \in A_i \backslash \{0, n\}$, $\bar{u}_{i0} = u'_{i0}$ and $\bar{u}_{in} = u'_{in}$, and $\bar{u}_{ij} \le u'_{ij}$ for $j \in B_i$. Hence, by Lemma 1, we obtain that $\operatorname{conc}_P(\phi)(u') = \phi' \le \operatorname{conc}_{P'}(\phi)(\bar{u}) \le \operatorname{conc}_{P'}(\phi)(u')$. □

By Proposition 1, it suffices to construct $\operatorname{conv}\big(\Phi^{P(a,a_{\cdot 0},a_{\cdot n},B)}\big)$ to characterize $\operatorname{conv}\big(\Phi^{P(a,u^L,u^U,B)}\big)$. The primary role played by u^L and u^U is to ensure that the estimating functions $u(\cdot)$ are bounded. Therefore, without loss of generality, we will assume in the foregoing discussion, unless specified otherwise, that $u^L = a_{\cdot 0}$ and $u^U = a_{\cdot n}$. In other words, we will use Proposition 1 to simplify P. Proposition 1 has another subtle value. It turns out that one of the main results in this paper can be viewed as a sharpening of Proposition 1. This sharper version will show that $(u, \phi) \in \operatorname{conv}(\Phi^P)$ if and only if a certain point $(s, \phi) \in \operatorname{conv}(\Phi^P)$, where s is larger than the u' constructed in the statement of Proposition 1. The transformation from u to s is significantly more involved and is, arguably, one of the cornerstones of the development later. The more general result will have much further reaching consequences since it will enable a significant simplification of P.

We now further simplify the structure of polytope P. First, we show that we may assume that the index set B_i defining P_i in (3) is empty. To do so, we transform overestimators into underestimators (and vice-versa) and thus show that the assumption $B_i = \emptyset$ is without loss of generality. Consider an affine transformation $T : \mathbb{R}^{d\times(n+1)} \to \mathbb{R}^{d\times(n+1)}$ defined as follows:

$$T(u)_{ij} = u_{ij} \quad \text{for } j \in A_i \qquad \text{and} \qquad T(u)_{ij} = a_{ij} - u_{ij} + u_{in} \quad \text{for } j \in B_i. \tag{7}$$

Recall that A_i and B_i are the index sets for underestimators and overestimators of f_i respectively. Let $\big(u(x), a\big)$ be a pair satisfying (2). It follows that $\big(T(u(x))\big)_{ij}$ is an underestimator of $f_i(x)$ bounded from above by a_{ij} because $a_{ij} \le u_{ij}(x)$ for all i and

$j \in B_i$. Clearly, the transformation T_i is invertible. More specifically, given a vector $t \in \mathbb{R}^{d\times(n+1)}$, we have $T^{-1}(t)_{ij} = t_{ij}$ for $j \in A_i$ and $T^{-1}(t)_{ij} = a_{ij} - t_{ij} + t_{in}$ for $j \in B_i$, where we have used $n \in A_i$ to substitute t_{in} for u_{in}. Since the transformation T is such that $T(u)_{ij}$ depends only on u_{ij} and u_{in}, we may write $T(u) := \big(T_1(u_1), \dots, T_d(u_d)\big)$, where $T_i : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$. Similarly, we write $T^{-1}(t) := \big(T_1^{-1}(t_1), \dots, T_d^{-1}(t_d)\big)$. Then,

$$\hat{P}_i := T_i(P_i) = \left\{ t_i \in \mathbb{R}^{n+1} \;\middle|\; \begin{array}{l} \forall j \in A_i : t_{ij} \le t_{in},\; u_i^L \le t_{ij} \le a_{ij} \\ \forall j \in B_i : t_{ij} \le t_{in},\; t_{in} + a_{ij} - u_i^U \le t_{ij} \le a_{ij} \\ t_{i0} = a_{i0},\; a_{i0} \le t_{in} \le a_{in} \end{array} \right\}.$$

We further relax $\hat{P}_i$ to $\bar{P}_i$ as follows:

$$\bar{P}_i = \left\{ t_i \in \mathbb{R}^{n+1} \;\middle|\; t_{ij} \le t_{in},\; \bar{u}_i^L \le t_{ij} \le a_{ij},\; t_{i0} = a_{i0},\; a_{i0} \le t_{in} \le a_{in} \right\},$$

where $\bar{u}_i^L = \min\{u_i^L, a_{i0} + a_{ij} - u_i^U\}$. Let $\hat{P} = \prod_{i=1}^d \hat{P}_i$ and $\bar{P} = \prod_{i=1}^d \bar{P}_i$.

Proposition 2 *Let* $G := \big\{(t,\phi) \;\big|\; t \in \hat{P}, (t,\phi) \in \mathrm{conv}(\Phi^{\bar{P}})\big\}$. *Then,* $\mathrm{conv}(\Phi^{\hat{P}}) = G$. *Moreover,* $\mathrm{conv}(\Phi^P) = \big\{(u,\phi) \;\big|\; u \in P, \big(T(u),\phi\big) \in \mathrm{conv}(\Phi^{\bar{P}})\big\}$. □

The proof of Proposition 2, which is included in "Appendix A.2", considers a point, $t \in \hat{P}$ and $(t,\phi) \in \mathrm{conv}(\Phi^{\bar{P}})$, and shows that $(t,\phi) \in \mathrm{conv}(\Phi^{\hat{P}})$. This is is done by changing each $(t^k, \phi^k) \in \Phi^{\bar{P}}$, which is used to express (t,ϕ) as a convex combination, such that the t_{ij}^k is replaced with $\max\big\{t_{ij}^k, u_i^L, t_{in}^k + a_{ij} - u_i^U\big\}$.

Since $\bar{P}$ is a special case of P with $B_i = \emptyset$ for all i, Proposition 2 shows that this special case is sufficient to treat the general case. Combined with Proposition 1, it suffices to consider $\bar{P}$ with $a_i^L = a_{i0}$. We may further assume that $a_{i0} < a_{i1} < \cdots < a_{in}$, *i.e.*, we may assume that $(a_{ij})_{j=0}^n$ are strictly increasing in j since we may replace with u_{ij} the maximum of all underestimators that share the same bounds. Unless specified otherwise, we will, from here onwards, let $P(a)$ denote $\prod_{i=1}^d P_i(a_i)$ for $a_{i0} < \cdots < a_{in}$, where

$$P_i(a_i) = \left\{ u_i \in \mathbb{R}^{n+1} \;\middle|\; \begin{array}{l} u_{ij} \le u_{in},\; a_{i0} \le u_{ij} \le a_{ij} \\ u_{i0} = a_{i0},\; a_{i0} \le u_{in} \le a_{in} \end{array} \right\}. \tag{8}$$

We showed in Example 1 that redundant underestimators of inner-functions can help improve the relaxation of the composite function. However, when underestimators are obtained using a convex combination derivation, we will show that they do not improve the quality of the relaxation. For a pair $\big(u(x), a\big)$ satisfying (2), we denote by $R\big(u(x), a\big)$ the relaxation of composite function $\phi \circ f$ obtained as in (6). In addition, consider a vector of underestimators $u' : \mathbb{R}^m \to \mathbb{R}^{d\times(n'+1)}$ and their upper bounds $a' \in \mathbb{R}^{d\times(n'+1)}$ obtained by taking convex combinations of $u(x)$ and a, respectively, where $n' \ge 0$. More precisely, let Λ_i be a nonnegative matrix in $\mathbb{R}^{(n+1)\times(n'+1)}$, where rows indexes are in $\{0, \dots, n\}$ and columns indexes are in $\{0, \dots, n'\}$ such that

$$\Lambda_i \geq 0,\ e^{n+1}\Lambda_i = e^{n'+1},\ \Lambda_i(\cdot, 0) = (1, 0, \dots, 0)^\top,\ \Lambda_i(\cdot, n') = (0, \dots, 0, 1)^\top, \tag{9}$$

where e^k is the all-ones row vector in $\mathbb{R}^k$ and $\Lambda_i(\cdot, k)$ is the k^{th} column of the matrix Λ_i. We let $u_i(x)$ (resp. $u'_i(x)$) denote a row vector of $n+1$ (resp. $n'+1$) functions $\big(u_{i0}(x), \dots, u_{in}(x)\big)$ (resp. $\big(u'_{i0}(x), \dots, u'_{in'}(x)\big)$). Then, we define $u'_i(x) : \mathbb{R}^m \to \mathbb{R}^{n'+1}$ such that $u'_i(x) := u_i(x)\Lambda_i$ and $a'_i \in \mathbb{R}^{n'+1}$ such that $a'_i := a_i\Lambda_i$. First two conditions in (9) imply that, for all i and $k \in \{0, \dots, n'\}$, $u'_{ik}(x)$ and the corresponding bound a'_{ik} are obtained by taking a convex combination of $u_{i0}(x), \dots, u_{in}(x)$ and their bounds $a_{i0}, \dots, a_{in}$ respectively. The third (resp. fourth) requirement in (9) ensures that $u'_{i0}(x) = u_{i0}(x)$ and $a'_{i0} = a_{i0}$ (resp. $u'_{in'}(x) = u_{in}(x)$ and $a'_{in'} = a_{in}$). Therefore, the new pair $\big(u'(x), a'\big)$ satisfies the requirements in (2), and thus, by Theorem 2, $R(u'(x), a')$ is a valid relaxation for the composite function $\phi \circ f$. In the following proposition, we show that $R\big(u(x), a\big) \subseteq R\big(u'(x), a'\big)$.

Proposition 3 *Let $\big(u(x), a\big)$ be a pair which satisfies conditions in (2) and Λ_i be a matrix defined in (9). Define $u'(x) \in \mathbb{R}^d \to \mathbb{R}^{d\times(n'+1)}$ such that $u'_{ij}(x) = u_i(x)\Lambda_i(\cdot, j)$ and $a' \in \mathbb{R}^{d\times(n'+1)}$ such that $a'_{ij} = a_i\Lambda_i(\cdot, j)$. Then, $R\big(u(x), a\big) \subseteq R\big(u'(x), a'\big)$.*

Proof See "Appendix A.3". □

4 Polynomial time equivalence of separations

To efficiently utilize $\operatorname{conv}(\Phi^P)$ for constructing relaxations, we must solve the *separation problem* of $\operatorname{conv}(\Phi^P)$, that is, given a vector $(\bar{u}, \bar{\phi})$ we need to determine if $(\bar{u}, \bar{\phi}) \in \operatorname{conv}(\Phi^P)$ and, if not, find a hyperplane that separates $(\bar{u}, \bar{\phi})$ from $\operatorname{conv}(\Phi^P)$. The main goal of this section is to prove that the separation problem of $\operatorname{conv}(\Phi^P)$ can be solved in polynomial time, given a polynomial time separation oracle for $\operatorname{conv}(\Phi^Q)$, where $Q := \prod_{i=1}^d Q_i$ for a certain subset Q_i of P_i, which will be formally defined in Sect. 4.1, and

$$\Phi^Q := \big\{(s, \phi) \;\big|\; \phi = \phi(s_{1n}, \dots, s_{dn}),\ s = (s_1, \dots, s_d) \in Q\big\}.$$

4.1 A simplicial structure of polyhedral abstraction

Recall that a_i is a strictly increasing vector in $\mathbb{R}^{n+1}$. Now, we consider a set of points $\{v_{ij}\}_{j=0}^n \subseteq \mathbb{R}^{n+1}$, where $v_{i0} = (a_{i0}, \dots, a_{i0})$ and:

$$v_{ij} = \big(a_{i0}, \dots, a_{ij-1}, \underbrace{a_{ij}, \dots, a_{ij}}_{n-j+1 \text{ terms}}\big) \qquad j = 1, \dots, n. \tag{10}$$

Clearly, $Q_i := \operatorname{conv}\big(\{v_{ij}\}_{j=0}^n\big)$ is a simplex in $\mathbb{R}^{n+1}$. Moreover, Q_i is a subset of P_i since $\{v_{ij}\}_{j=0}^n \subseteq P_i$. In the following, we first characterize the vertex set and facet-defining inequalities of Q_i. Then, we study the relation between $\operatorname{conv}(\Phi^P)$ and $\operatorname{conv}(\Phi^Q)$. In particular, we show that a convex hull description of Φ^Q, together with $\mathcal{O}(dn)$ number of inequalities, yields a convex hull description of Φ^P in the space of (u, s, ϕ) variables.

To characterize the simplex Q_i, we argue that it can be seen as an invertible affine transform of a simpler simplex Δ_i defined as:

$$\Delta_i := \big\{z_i \in \mathbb{R}^{n+1} \;\big|\; 0 \le z_{in} \le \cdots \le z_{i1} \le z_{i0} = 1\big\}. \tag{11}$$

Denote by ζ_{ij} the vector $\sum_{j'=0}^{j} e_{j'}$ for all $j = 0, \ldots, n$, where e_j is the j^{th} principal vector in $\mathbb{R}^{n+1}$ with $e_0 = (1, 0, \ldots, 0)^\top$. It follows readily that points $\zeta_{i0}, \ldots, \zeta_{in}$ are affinely independent since the matrix $\zeta_i \in \mathbb{R}^{(n+1)\times(n+1)} := (\zeta_{i0}, \ldots, \zeta_{in})$ is invertible, being the upper triangular matrix of all ones. Moreover, it can be verified that $\operatorname{vert}(\Delta_i) = \{\zeta_{ij}\}_{j=0}^n$. This implies that $\dim(\Delta_i) = n$.

Now, the affine transformation that maps Q_i to Δ_i is $Z_i : s_i \in \mathbb{R}^{n+1} \mapsto z_i \in \mathbb{R}^{n+1}$, where

$$z_{i0} = 1 \qquad \text{and} \qquad z_{ij} = \frac{s_{ij} - s_{ij-1}}{a_{ij} - a_{ij-1}} \quad \text{for } j = 1, \ldots, n. \tag{12}$$

To verify that $\Delta_i = Z_i(Q_i)$ observe that Z_i maps v_{ij} to ζ_{ij}. Besides, the inverse Z_i^{-1} is given by:

$$s_{ij} = a_{i0}z_{i0} + \sum_{k=1}^{j}(a_{ik} - a_{ik-1})z_{ik} \quad \text{for } j = 0, \ldots, n, \tag{13}$$

and $Q_i = Z_i^{-1}(\Delta_i)$. We now characterize the extreme points and facet-defining inequalities of Q_i.

Lemma 2 *Let $a_i := (a_{i0}, \ldots, a_{in})$ be a strictly increasing vector in $\mathbb{R}^{n+1}$ and let $Q_i := \operatorname{conv}\big(\{v_{ij}\}_{j=0}^n\big)$. Then, Q_i is a simplex so that* $\operatorname{vert}(Q_i) = \{v_{ij}\}_{j=0}^n$, *and*

$$Q_i = \left\{(s_{ij})_{j=0}^n \;\middle|\; s_{i0} = a_{i0},\ 0 \le \frac{s_{in} - s_{in-1}}{a_{in} - a_{in-1}} \le \cdots \le \frac{s_{i1} - s_{i0}}{a_{i1} - a_{i0}} \le 1\right\}, \tag{14}$$

where all the inequalities are facet-defining. For j, j', and j'', satisfying $0 \le j < j' < j'' \le n$, each point in Q_i satisfies $0 \le \frac{s_{ij''}-s_{ij'}}{a_{ij''}-a_{ij'}} \le \frac{s_{ij'}-s_{ij}}{a_{ij'}-a_{ij}} \le 1$.

Proof See “Appendix A.4”. □

Next, we lift simplex Q_i into the space of (u_i, s_i) variables by imposing ordering constraints $a_{i0}e \le u_i \le s_i$, $u_{i0} = s_{i0}$, and $u_{in} = s_{in}$. This yields a polytope

$$PQ_i = \big\{(u_i, s_i) \;\big|\; u_{i0} = s_{i0} = a_{i0},\ u_{in} = s_{in},\ a_{i0}e \le u_i \le s_i,\ s_i \in Q_i\big\}, \tag{15}$$

where e is the all-ones vector in $\mathbb{R}^{n+1}$. Let $PQ := \prod_{i=1}^{n} PQ_i$ and

$$\Phi^{PQ} := \big\{(u, s, \phi) \bigm| \phi = \phi(s_{1n}, \ldots, s_{dn}), (u, s) \in PQ\big\}.$$

The next result establishes that Φ^P and Φ^Q are projections of Φ^{PQ}.

Lemma 3 *Let $a_i := (a_{i0}, \ldots, a_{in})$ be a strictly increasing vector in $\mathbb{R}^{n+1}$. Then,* $\text{proj}_{u_i}(PQ_i) = P_i$ *and* $\text{proj}_{s_i}(PQ_i) = Q_i$. *Moreover,* $\Phi^P = \text{proj}_{(u,\phi)}(\Phi^{PQ})$ *and* $\Phi^Q = \text{proj}_{(s,\phi)}(\Phi^{PQ})$.

Proof See "Appendix A.5". □

Remark 1 In the following, it will be useful to interpret the point $(u_i, s_i) \in PQ_i$ in the following way. First, recall that in Sect. 3.1, polytope P_i was introduced as an abstract way of capturing the information about the underestimators of the inner-function $f_i(\cdot)$. The important property of the underestimator $u_{ij}(\cdot)$ exploited in the construction of P_i is its upper bound a_{ij}. Observe now that $u_{ij} \le s_{ij} \le \min\{a_{ij}, u_{in}\}$. This is because $(u_i, s_i) \in PQ_i$ implies that $u_{ij} \le s_{ij}$ and $s_{in} = u_{in}$, and $s_{ij} \le \min\{a_{ij}, s_{in}\}$ is satisfied by all the extreme points of Q_i. Therefore, we may interpret s_{ij} as a variable corresponding to an underestimator of $f_i(\cdot)$ that is tighter than $u_{ij}(\cdot)$ and yet bounded from above by a_{ij}. Our constructions make significant use of this relationship. In short, P_i can be considered as the polytope abstraction of arbitrary underestimators of $f_i(\cdot)$ while Q_i can be regarded as the polytope corresponding to the best underestimators obtained by taking convex combinations of the provided underestimators. □

Observe that Lemma 3 implies that

$$\text{conv}(\Phi^P) = \text{conv}\big(\text{proj}_{(u,\phi)}(\Phi^{PQ})\big) = \text{proj}_{(u,\phi)}\big(\text{conv}(\Phi^{PQ})\big),$$

where the second equality holds because $\text{conv}(AS) = A\,\text{conv}(S)$ for any affine mapping A and a set S. Similarly, $\text{conv}(\Phi^Q) = \text{proj}_{(s,\phi)}\,\text{conv}\big(\Phi^{PQ}\big)$. In the following, we will show that

$$\text{conv}(\Phi^{PQ}) = \big\{(u, s, \phi) \bigm| (s, \phi) \in \text{conv}(\Phi^Q),\ (u, s) \in PQ\big\}. \tag{16}$$

Therefore, the right hand side of (16) yields an extended formulation of $\text{conv}(\Phi^P)$. Next, we establish the equality in (16) in a more general setting, which is also useful when we consider the convex hull of a vector of functions over PQ in Sect. 5. Let $\mathcal{X} := \{(x, \mu) \in \mathbb{R}^{m+k} \mid \mu = f(x), x \in X\}$ be the graph of a vector of functions $f : \mathbb{R}^m \to \mathbb{R}^k$ over a non-empty subset X of $\mathbb{R}^m$. Let $l : \mathbb{R}^m \to \mathbb{R}^n$ and $h : \mathbb{R}^m \to \mathbb{R}^n$ be vectors of affine functions. We consider a set $\mathcal{Z}$ defined as:

$$\mathcal{Z} = \big\{(x, y, \mu) \in \mathbb{R}^{m+n+k} \bigm| l(x) \le y \le h(x),\ (x, \mu) \in \mathcal{X}\big\}.$$

To express Φ^{PQ} in the form of $\mathcal{Z}$, we let $y = u$, $\mathcal{X} = \Phi^Q$, $h(s) = s$, $l_{in}(s) = s_{in}$, and $l_{ij}(s) = a_{i0}$ if $j < n$. Interpreting $\mathcal{Z}$ as such, the following result implies the equality in (16).

Lemma 4 *Assume that $l(x) \le h(x)$ for all $x \in X$. Then, we have* $\mathrm{conv}(\mathcal{Z}) = \big\{(x, y, \mu) \bigm| (x, \mu) \in \mathrm{conv}(\mathcal{X}),\ l(x) \le y \le h(x)\big\}$.

Proof Let $R := \big\{(x, y, \mu) \bigm| (x, \mu) \in \mathrm{conv}(\mathcal{X}),\ l(x) \le y \le h(x)\big\}$. The inclusion $\mathrm{conv}(\mathcal{Z}) \subseteq R$ follows from $\mathcal{Z} \subseteq R$ since R is convex. We now show that $R \subseteq \mathrm{conv}(\mathcal{Z})$. Let $(x', y', \mu') \in R$. Clearly, (x', y', μ') lies in the hypercube, $H := \big\{(x, y, \mu) \bigm| (x, \mu) = (x', \mu'),\ l(x') \le y \le h(x')\big\}$. So, it suffices to show that the vertex set of H belongs to $\mathrm{conv}(\mathcal{Z})$. We only show that vertices of the form $\big(x', l(x'), \mu'\big)$ lie in $\mathrm{conv}(\mathcal{Z})$ since a similar argument applies to the remaining vertices. Let $\mathcal{Z}' := \big\{(x, y, \mu) \bigm| (x, \mu) \in \mathcal{X},\ y = l(x)\big\}$ and express it as the image of $\mathcal{X}$ under the affine transformation $A : (x, \mu) \mapsto \big(x, l(x), \mu\big)$. Therefore, $\mathrm{conv}(\mathcal{Z}') = \mathrm{conv}\big(A(\mathcal{X})\big) = A\big(\mathrm{conv}(\mathcal{X})\big)$ and, so, $\big(x', l(x'), \mu'\big) \in A\big(\mathrm{conv}(\mathcal{X})\big) = \mathrm{conv}(\mathcal{Z}') \subseteq \mathrm{conv}(\mathcal{Z})$, where the last containment follows because $\mathcal{Z}' \subseteq \mathcal{Z}$ is implied by our assumption that $l(x) \le h(x)$ for all $x \in X$. □

4.2 A combinatorial algorithm for polynomial equivalence

Before establishing that the separation problems of $\mathrm{conv}(\Phi^P)$ and $\mathrm{conv}(\Phi^Q)$ are polynomially equivalent, we give an application of this equivalence. For this, we consider the case when the number of inner-functions is fixed and it suffices to restrict $\phi(\cdot)$ to the extreme points of Q in order to construct the convex hull of Φ^Q. We show that the separation problem for $\mathrm{conv}(\Phi^Q)$ has a polynomial-sized LP formulation in this case. Then, by the announced polynomial equivalence of the separation of $\mathrm{conv}(\Phi^P)$ and $\mathrm{conv}(\Phi^Q)$, this yields a polynomial-time separation algorithm for $\mathrm{conv}(\Phi^P)$. Since $\mathrm{vert}(Q) = \prod_{i=1}^d \mathrm{vert}(Q_i)$ and $|\mathrm{vert}(Q_i)| = n + 1$, it follows that $|\mathrm{vert}(Q)| = (n + 1)^d$. In other words, if d is a constant, the number of vertices of Q is polynomial in n. More generally, assuming r and d are constants, the number of r dimensional faces of Q is polynomial in n, being upper-bounded by $(n + 1)^d \binom{nd}{r}$. Since convex/concave-envelopes of many functions depend only on the function value at the vertices or, more generally, low-dimensional faces of the domain over which the function is convex, we can construct polynomial-sized formulations for separation of $\mathrm{conv}(\Phi^Q)$ and, therefore, of $\mathrm{conv}(\Phi^P)$. For simplicity, we only discuss the implication of the equivalence of separating $\mathrm{conc}_Q(\phi)$ and $\mathrm{conc}_P(\phi)$ since a similar discussion directly applies to $\mathrm{conv}_Q(\phi)$ and $\mathrm{conv}_P(\phi)$ by considering $-\phi$ instead.

Definition 1 ([45]) A function $g : D \to \mathbb{R}$, where D is a polytope, is concave-extendable (resp. convex-extendable) from $X \subseteq D$ if the concave (resp. convex) envelope of $g(x)$ is determined by X, that is, the concave envelope of g and $g|_X$ over D are identical, where $g|_X$ is the restriction of g to X:

$$g|_X = \begin{cases} g(x) & x \in X \\ -\infty & \text{otherwise.} \end{cases}$$

□

If $\phi(s_{1n}, \dots, s_{dn})$ is concave-extendable from vertices of Q then, using Theorem 2.4 in [45], we can separate $(\bar{s}, \bar{\phi}) \in \mathbb{R}^{d\times(n+1)} \times \mathbb{R}$ from the hypograph of $\mathrm{conc}_Q(\phi)(s)$

by solving the following linear program:

$$\begin{aligned} \min_{(\alpha,b)} \ & \langle \alpha, \bar{s}\rangle + b \\ \text{s.t.} \ & \langle \alpha, v\rangle + b \geq \phi\big(v_{1n}, \ldots, v_{dn}\big) \ \forall v \in \operatorname{vert}(Q) \\ & \alpha \in \mathbb{R}^{d\times(n+1)}, \ b \in \mathbb{R}, \end{aligned} \tag{17}$$

where a solution (α^*, b^*) that is an extreme point of the feasible region yields a facet-defining inequality of $\operatorname{conc}_Q(\phi)(s)$ tight at $\bar{s}$, *i.e.* $\langle \alpha^*, \bar{s}\rangle + b^* = \operatorname{conc}_Q(\phi)(\bar{s})$. We retain b although it can be absorbed in α_{i0} if $\alpha_{i0} \neq 0$. Since $|\operatorname{vert}(Q)| = (n+1)^d$, it follows that the size of the LP (17) is polynomial in n for a fixed d. As an example of the usefulness of this construction, observe that multilinear functions are convex and concave extendable from the vertices of Q (see [39]). Then, as Theorem 3 shows, the above LP gives a tractable approach to separate $\operatorname{conv}(\Phi^Q)$, when ϕ is multilinear and d is fixed. Techniques in [47] can be used to identify whether a function is concave-extendable (convex-extendable) from vertices of Q.

Theorem 3 *Assume that $\phi(s_{1n}, \ldots, s_{dn})$ is concave-extendable from* $\operatorname{vert}(Q)$ *and d is a fixed constant. For any given $\bar{s} \in \mathbb{R}^{d\times(n+1)}$, there exists a polynomial time procedure to generate a facet-defining inequality of* $\operatorname{conc}_Q(\phi)(s)$ *that is tight at $\bar{s}$.*

Proof Given $\bar{s} \in Q$, the LP (17) can be solved in polynomial time by using an interior point algorithm. Moreover, by Lemma 6.5.1 in [19], an optimal extreme point solution of linear program (17) can be found in polynomial time. Then, the result follows from Theorem 2.4 in [45]. □

Remark 2 Although the separation problem of $\operatorname{conc}_Q(\phi)(s)$ can be directly formulated as a LP of polynomial size, the similar LP formulation for $\operatorname{conc}_P(\phi)(u)$ using the construction of (17) is exponentially-sized in n because the $|\operatorname{vert}(P_i)|$ is exponential in n. To see this, for $i \in \{1, \ldots, d\}$, consider the face of P_i defined as $F_i := P_i \cap \{u_i \mid u_{in} = a_{in}\}$. Since F_i coincides with the hypercube $\{u_i \mid u_{in} = a_{in}, \ a_{i0} \leq u_{ij} \leq a_{ij}, \ \ j = 1, \ldots, n-1\}$ and $a_{i0} < a_{ij}$ for $j = 1, \ldots, n-1$, it follows that $|\operatorname{vert}(F_i)| = 2^{n-1}$. As $\operatorname{vert}(F_i) \subseteq \operatorname{vert}(P_i)$, $|\operatorname{vert}(P_i)| \geq 2^{n-1}$. Therefore, by $\operatorname{vert}(P) = \prod_{i=1}^d |\operatorname{vert}(P_i)|$, we conclude that $|\operatorname{vert}(P)| \geq 2^{d(n-1)}$. □

Remark 3 A convex program, similar to the above LP, can be written to treat the separation problem of $\operatorname{conc}_Q(\phi)(s)$ for more general cases. For example, consider the case when $\operatorname{conc}_Q(\phi)$ is determined by its value over polynomially many faces of Q (for example, faces of dimension r or less, for some constant r) and $\phi(s_{1n}, \ldots, s_{dn})$ is concave over those faces. To treat this case, we replace the constraint in (17) with $b \geq \sup_{x\in F}\big\{\phi(x) - \langle \alpha, x\rangle\big\} = (-\phi)^*_F(-\alpha)$, for each face F of Q which is required in the computation of $\operatorname{conc}_Q(\phi)$. Here, $(-\phi)^*_F$ denotes the Fenchel conjugate of $-\phi(s_{1n}, \ldots, s_{dn})$ with its domain restricted to F. □

In the rest of this subsection, we will focus on proving that separating $\operatorname{conv}(\Phi^P)$ and $\operatorname{conv}(\Phi^Q)$ are polynomially equivalent. In particular, we devise a combinatorial algorithm that solves the separation problem of $\operatorname{conv}(\Phi^P)$ in $\mathcal{O}(dn)$ and a separation oracle call for $\operatorname{conv}(\Phi^Q)$. We start by presenting a brief preview of our construction.

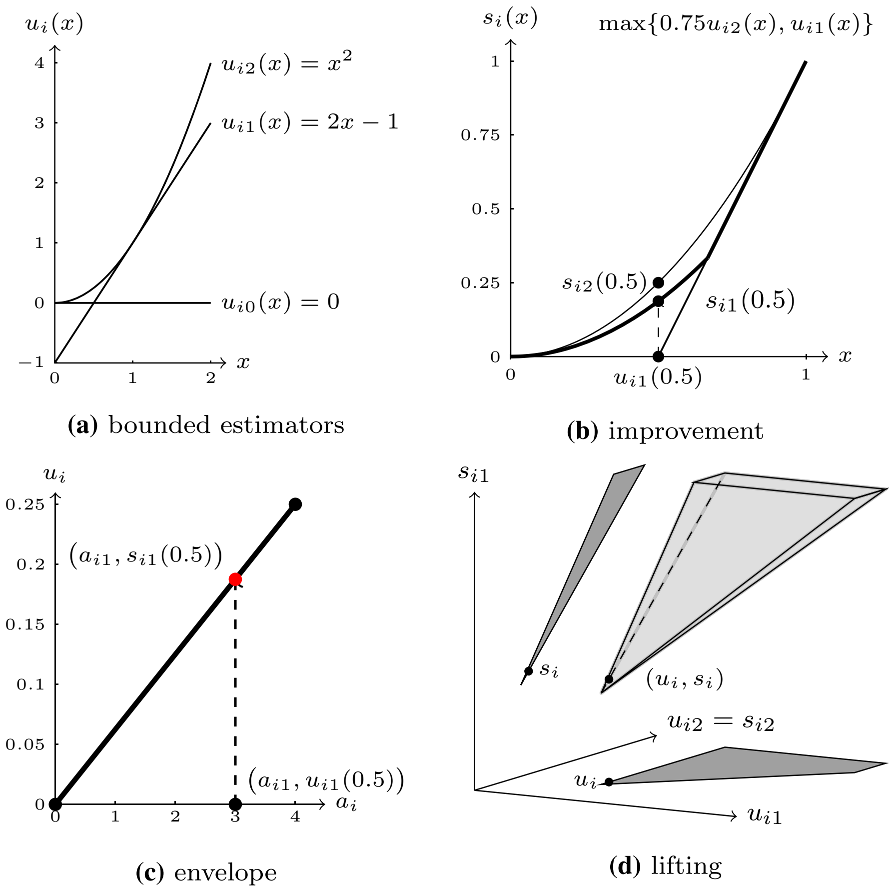

Fig. 1 Illustration of the lifting procedure

Given a point $(\bar{u}, \bar{\phi}) \notin \text{conv}(\Phi^P)$, where $\bar{u} := (\bar{u}_1, \dots, \bar{u}_d) \in P$, we first devise a lifting procedure, Algorithm 1, to lift $\bar{u}_i$ to a particular point $(\bar{u}_i, \bar{s}_i) \in PQ_i$ for all $i = 1, \dots, d$. For this particular pair $(\bar{u}_i, \bar{s}_i)$, we show in Proposition 5 that, for $j \in \{0, \dots, n\}$, $\bar{s}_{ij}$ can be expressed as a convex combination of $\bar{u}_{i0}, \dots, \bar{u}_{in}$. Second, we show that $(\bar{s}, \bar{\phi}) \notin \text{conv}(\Phi^Q)$, where $\bar{s} = (\bar{s}_1, \dots, \bar{s}_d)$. Third, we augment the separation oracle for $\text{conc}_Q(\phi)(s)$ to modify the cut that separates $(\bar{s}, \bar{\phi})$ from $\text{conv}(\Phi^Q)$ into another cut which additionally satisfies a certain sign condition on the coefficients. Last, given a cut for $\text{conc}_Q(\phi)$ which satisfies these sign conditions, we use the relationship between $\bar{s}_i$ and $\bar{u}_i$ to derive an inequality valid for $\text{conv}(\Phi^P)$ that cuts off $(\bar{u}, \bar{\phi})$. The lifting procedure that maps $\bar{u}$ to $(\bar{u}, \bar{s})$ is a cornerstone in our proof architecture. Before presenting the lifting procedure formally, we illustrate, in the next example, the main idea behind the procedure.

Example 4 Consider x^2 where x lies in the interval $[0, 2]$, and define $u_i(x) = \big(u_{i0}(x), u_{i1}(x), u_{i2}(x)\big)$ where $u_{i0}(x) := 0$, $u_{i1}(x) := 2x - 1$, and $u_{i2}(x) := x^2$, which are bounded from above by $a_{i0} = 0$, $a_{i1} = 3$ and $a_{i2} = 4$ respectively (see

Fig. 1a). We derive the largest underestimator $s_{ij}(x)$ bounded from above by a_{ij} by taking convex combinations of the provided underestimators. More specifically, for all j, $s_{ij}(x) := \max\bigl\{\sum_{j'=0}^{2}\lambda'_j u_{ij'}(x) \bigm| \sum_{j'=0}^{2}\lambda'_j a_{ij'} = a_{ij}\bigr\}$. In Fig. 1b, we tighten $u_{i1}(x)$ to $s_{i1}(x) = \max\bigl\{0.75u_{i2}(x), u_{i1}(x)\bigr\}$. This tightening gives an underestimator $s_{i1}(x)$ of x^2 with an upper bound of 3 over [0, 2]. The derivation uses a convex combination of $u_{i2}(x)$ and $u_{10}(x)$, $0.75u_{i2}(x) + 0.25u_{i0}(x)$, to obtain $0.75u_{i2}(x)$ and its upper bound 3. We let $s_{i0}(x) = u_{i0}(x)$ and $s_{i2}(x) = u_{i2}(x)$.

We will find it useful to visualize the evaluation of $s_i(x)$ as depicted in Fig. 1c. Consider $x = 0.5$ and let $u_i = u_i(0.5)$. In order to evaluate $s_{i1}(x)$ at 0.5, we compute $\mathrm{conc}_{[0,2]}(\xi)(a; u_i)$ at $a = a_{i1} = 3$, where $\xi(a; u_i)$ is a univariate discrete function whose graph consists of the points, $\bigl\{(a_{i0}, u_{i0}), (a_{i1}, u_{i1}), (a_{i2}, u_{i2})\bigr\}$, which are depicted as black nodes in Fig. 1c. In this way, the construction of the envelope of this univariate function lifts $u_i = u_i(0.5) = (0, 0, 0.25)$ to $(u_i, s_i) = \bigl(u_i(0.5), s_i(0.5)\bigr) = \bigl((0, 0, 0.25), (0, 0.1875, 0.25)\bigr)$. The obtained pair (u_i, s_i) belongs to $PQ_i = \{(u, s) \mid u_{i0} = 0,\ u_{i1} \le s_{i1},\ u_{i2} = s_{i2},\ s_i \in Q_i\}$, where $Q_i := \bigl\{s \bigm| s_{i0} = 0,\ 0 \le \frac{s_{i2}-s_{i1}}{1} \le \frac{s_{i1}-s_{i0}}{3} \le 1\bigr\}$; see Fig. 1d. Observe that in order to be able to draw a 3-D figure, we do not show axes for the variables u_{i0} and s_{i0} which are fixed to 0 and depict u_{in} and s_{in} on the same axis by exploiting $u_{in} = s_{in}$. □

We now formally introduce the lifting operation illustrated in Example 4. Given a point $\bar{u} = (\bar{u}_1, \ldots, \bar{u}_d) \in P_1 \times \cdots \times P_d$, for each $i \in \{1, \ldots, d\}$, we lift $\bar{u}_i \in P_i$ to a point $(\bar{u}_i, \bar{s}_i)$, where $\bar{s}_i$ is an optimal solution of the following LP:

$$\begin{aligned} \min_{s}\ & \textstyle\sum_{j=0}^{n} s_{ij} \\ \text{s.t.}\ & \bar{u}_i \le s_i \\ & s_i \in Q_i. \end{aligned} \tag{18}$$

By Lemma 3, the feasible region of the linear program (18) is non-empty. We will show that (18) has an unique optimal solution and propose an algorithm, Algorithm 1, to solve (18) in $\mathcal{O}(n)$ operations. This algorithm relies on representing points in P_i as discrete univariate functions in the following way. With a point $u_i \in P_i$, we associate a discrete univariate function $\xi(a; u_i) : [a_{i0}, a_{in}] \to \mathbb{R}$ as follows:

$$\xi(a; u_i) = \begin{cases} u_{ij} & a = a_{ij} \quad \text{for } j \in \{0, \ldots, n\} \\ -\infty & \text{otherwise.} \end{cases} \tag{19}$$

Moreover, let $\hat{\xi}(a; u_i) : [a_{i0}, a_{in}] \to \mathbb{R}$ be the piecewise-linear interpolation of $\xi(a; u_i)$ such that $\hat{\xi}(a; u_i) = \xi(a; u_i)$ for $a \in \{a_{i0}, \ldots, a_{in}\}$ and, for all $j = 1, \ldots, n$, the restriction of $\hat{\xi}(a; u_i)$ to $[a_{ij-1}, a_{ij}]$ is linear. In the next result, we show that this representation leads to a characterization of points in the simplex Q_i as a family of univariate concave functions.

Lemma 5 *Given a point $s_i \in Q_i$, the extension $\hat{\xi}(a; s_i) : [a_{i0}, a_{in}] \to \mathbb{R}$ of $\xi(a; u_i)$ defined as in (19) is a non-decreasing concave function such that $\hat{\xi}(a; s_i) \le a$ for $a \in [a_{i0}, a_{in}]$ and $\hat{\xi}(a_{i0}; s_i) = a_{i0}$. On the other hand, if a concave function ψ :*

$[a_{i0}, a_{in}] \to \mathbb{R}$ *satisfies*

$$\psi(a) \le a \text{ for } a \in [a_{i0}, a_{in}], \quad \psi(a_{i0}) = a_{i0}, \quad \text{and} \quad \psi \text{ is non-decreasing} \tag{20}$$

then $s_i := \big(\psi(a_{i0}), \dots, \psi(a_{in})\big)$ *belongs to* Q_i.

Proof We start by proving the first part. Let $s_i \in Q_i$. By Lemma 2, there exists an unique non-negative vector $\lambda_i \in \mathbb{R}^{n+1}$ such that $s_i = \sum_{j=0}^{n} \lambda_{ij} v_{ij}$, where v_{ij} is defined as in (10). Then, there exists a $j' \in \{1, \dots, n\}$ and $\gamma \in [0, 1]$ such that

$$\begin{aligned}
\hat{\xi}(a; s_i) &= (1-\gamma)\xi(a_{ij'-1}; s_i) + \gamma\xi(a_{ij'}; s_i) \\
&= (1-\gamma)\sum_{j=0}^{n} \lambda_{ij}\xi(a_{ij'-1}; v_{ij}) + \gamma\sum_{j=0}^{n} \lambda_{ij}\xi(a_{ij'}; v_{ij}) \\
&= \sum_{j=0}^{n} \lambda_{ij}\big((1-\gamma)\xi(a_{ij'-1}; v_{ij}) + \gamma\xi(a_{ij'}; v_{ij})\big) = \sum_{j=0}^{n} \lambda_{ij}\hat{\xi}(a; v_{ij}),
\end{aligned}$$

where the first equality is because for $a \in [a_{i0}, a_{in}]$, there exists $j' \in \{1, \dots, n\}$ and $\gamma \in [0, 1]$ such that $a = (1-\gamma)a_{ij'-1} + \gamma a_{ij'}$ and $\hat{\xi}(a; s_i)$ is linear over $[a_{ij'-1}, a_{ij'}]$, the second equality is by linearity of $\xi(a; s_i)$ with respect to s_i, the third equality is by rearrangement of terms, and last equality holds because $\hat{\xi}(a; s_i)$ is linear in a over $[a_{ij'-1}, a_{ij'}]$. Since, for all $j \in \{0, \dots, n\}$, $\hat{\xi}(a; v_{ij}) \le a$ for $a \in [a_{i0}, a_{in}]$, $\hat{\xi}(a_{i0}; v_{ij}) = a_{i0}$, $\hat{\xi}(a; v_{ij})$ is a non-decreasing concave function, and these properties are closed under convex combination, it follows that $\hat{\xi}(a, s_i)$ follows these properties as well.

Assume that $\psi(a)$ is a univariate concave function satisfying (20). Let $s_i = \big(\psi(a_{i0}), \dots, \psi(a_{in})\big)$ and let $z_i = Z_i(s_i)$, where Z_i is as defined in (12). By the discussion preceding Lemma 2, to prove that $s_i \in Q_i$, it suffices to show that $z_i \in \Delta_i$. Clearly, we have $z_{ij} \ge 0$ because ψ is non-decreasing. Moreover, we have $z_{i1} = \frac{s_{i1}-s_{i0}}{a_{i1}-a_{i0}} \le 1$ because $\psi(a_{i0}) = a_{i0}$ and $\psi(a_{i1}) \le a_{i1}$. Finally, we show that the concavity of ψ implies that $z_{in} \le \cdots \le z_{i1}$. Let $j' \in \{1, \dots, n-1\}$, and let $\psi(a) \le L(a)$ be a supergradient inequality of ψ at $a_{ij'}$. It follows readily that

$$\begin{aligned}
z_{ij'+1} &= \frac{\psi(a_{ij'+1}) - \psi(a_{ij'})}{a_{ij'+1} - a_{ij'}} \le \frac{L(a_{ij'+1}) - L(a_{ij'})}{a_{ij'+1} - a_{ij'}} \\
&= \frac{L(a_{ij'}) - L(a_{ij'-1})}{a_{ij'} - a_{ij'-1}} \le \frac{\psi(a_{ij'}) - \psi(a_{ij'-1})}{a_{ij'} - a_{ij'-1}} = z_{ij'},
\end{aligned}$$

where first equality and last equality follow by definition, first inequality holds because $\psi(a_{ij'+1}) \le L(a_{ij'+1})$ and $\psi(a_{ij'}) = L(a_{ij'})$, second equality is the linearity of $L(a)$, second inequality holds because $\psi(a_{ij'}) = L(a_{ij'})$ and $\psi(a_{ij'-1}) \le L(a_{ij'-1})$. $\square$

Now, we present Algorithm 1 to lift a point in P_i to another in PQ_i. To lift, the algorithm constructs the concave envelope of the discrete one-dimensional function defined in (19). Formally, given a point $\bar{u}_i \in P_i$, let $\xi(a; \bar{u}_i)$ be a discrete function associated with the point $\bar{u}_i$. The graph of $\xi(a; \bar{u}_i)$ consists of $(n+1)$ points $(a_{i0}, \bar{u}_{i0}), \ldots, (a_{in}, \bar{u}_{in})$ in $\mathbb{R}^2$, which are sorted using the first coordinate. The concave envelope of $\xi(a; \bar{u}_i)$ over $[a_{i0}, a_{in}]$ can be found using two-dimensional convex hull algorithms (see [10,18]). In particular, since points of the graph of $\xi(a; \bar{u}_i)$ are sorted in terms of the first coordinate, Graham scan [18] takes $\mathcal{O}(n)$ to derive the envelope.

Algorithm 1 Lifting procedure

```
1: procedure LIFTING(ū_i)
2:     construct function ξ(a; ū_i)
3:     apply Graham scan to obtain conc(ξ)(a; ū_i)
4:     (s̄_i0, ..., s̄_in) ← (conc(ξ)(a_i0; ū_i), ..., conc(ξ)(a_in; ū_i))
5:     return s̄_i.
6: end procedure
```

Proposition 4 *Given $\bar{u}_i \in P_i$ Algorithm 1 returns the unique optimal solution $\bar{s}_i$ of (18) in $\mathcal{O}(n)$.*

Proof We first show that $\bar{s}_i$ is a feasible solution to (18). Clearly, $\bar{u}_i \leq \bar{s}_i$ as $\xi(a; \bar{u}_i) \leq \operatorname{conc}(\xi)(a; \bar{u}_i)$. Next, we show that $\operatorname{conc}(\xi)(a; \bar{u}_i)$ satisfies the three conditions in (20), and, therefore, by the second result in Lemma 5, $\bar{s}_i \in Q_i$. First, $\operatorname{conc}(\xi)(a_{i0}; \bar{u}_i) = \xi(a_{i0}; \bar{u}_i) = \bar{u}_{i0} = a_{i0}$. Second, observe that $\xi(a; \bar{u}_i) \leq a$ since, for all i and j, $\bar{u}_{ij} \leq a_{ij}$. This implies that $\operatorname{conc}(\xi)(a; \bar{u}_i) \leq a$ because by definition $\operatorname{conc}(\xi)(a; \bar{u}_i)$ is the smallest concave overestimator of $\xi(a; \bar{u}_i)$ over $[a_{i0}, a_{in}]$. Last, we show the monotonicity of $\operatorname{conc}(\xi)(a; \bar{u}_i)$. Observe that, for every $a \in [a_{i0}, a_{in}]$, $\operatorname{conc}(\xi)(a; \bar{u}_i) \leq \bar{u}_{in} = \xi(a_{in}, \bar{u}_i) = \operatorname{conc}(\xi)(a_{in}; \bar{u}_i)$, where first inequality holds because $\bar{u}_{ij} \leq \bar{u}_{in}$ implies that $\xi(a; \bar{u}_i) \leq \bar{u}_{in}$. Consider two points a', a'' such that $a' < a'' < a_{in}$. Let $\lambda \in [0, 1]$ such that $a'' = (1-\lambda)a' + \lambda a_{in}$. Observe that $\operatorname{conc}(\xi)(a''; \bar{u}_i) \geq (1-\lambda)\operatorname{conc}(\xi)(a'; \bar{u}_i) + \lambda \operatorname{conc}(\xi)(a_{in}; \bar{u}_i) \geq \operatorname{conc}(\xi)(a'; \bar{u}_i)$, where the first inequality is by concavity of $\operatorname{conc}(\xi)$ and the second inequality is because $\operatorname{conc}(\xi)(a'; \bar{u}_i) \leq \operatorname{conc}(\xi)(a_{in}; \bar{u}_i)$.

Next, we prove by contradiction that $\bar{s}_i$ is the optimal solution of (18). Suppose that s'_i is a feasible solution so that $\sum_{j=0}^n s'_{ij} \leq \sum_{j=0}^n \bar{s}_{ij}$ and $s'_i \neq \bar{s}_i$. As $s'_i \in Q_i$ it follows from the first statement of Lemma 5 that $\hat{\xi}(a; s'_i)$ is a concave function over the interval $[a_{i0}, a_{in}]$. Moreover, we have $\xi(a; \bar{u}_i) \leq \hat{\xi}(a; s'_i)$ because $\bar{u}_i \leq s'_i$. In other words, $\hat{\xi}(a; s'_i)$ is a concave overestimator of $\xi(a; \bar{u}_i)$ over $[a_{i0}, a_{in}]$. By the hypothesis, $\sum_{j=0}^n \hat{\xi}(a_{ij}; s'_i) = \sum_{j=0}^n s'_{ij} \leq \sum_{j=0}^n \bar{s}_{ij} = \sum_{j=0}^n \operatorname{conc}(\xi)(a_{ij}; \bar{u}_i)$. Since $s'_i \neq \bar{s}_i$, there exists a $j' \in \{0, \ldots, n\}$ such that $\hat{\xi}(a_{ij'}; s'_i) = s'_{ij'} < \bar{s}_{ij} = \operatorname{conc}(\xi)(a_{ij'}; \bar{u}_i)$, contradicting that $\operatorname{conc}(\xi)(a; \bar{u}_i)$ is the smallest concave overestimator of $\xi(a; \bar{u}_i)$ over $[a_{i0}, a_{in}]$. □

To further understand the output of Algorithm 1, we study the following set

$$PQ' := \bigl\{(u,s) \bigm| s_i = \bigl(\mathrm{conc}(\xi)(a_{i0};u_i),\dots,\mathrm{conc}(\xi)(a_{in};u_i)\bigr),\ i=1,\dots,d\bigr\}. \tag{21}$$

In particular, for a point $(\bar{u},\bar{s}) \in PQ'$, we will recover a mapping from $\bar{u}$ to $\bar{s}$ that is implicit in the construction of $\mathrm{conc}(\xi)$, and, geometrically, we show that $\bar{s}$ lies in a face of Q. In order to do so, we observe that the envelope, $\mathrm{conc}(\xi)(\cdot;\bar{u}_i)$, gives a representation of $(a_{ij},\bar{s}_{ij})$ as a convex combination of points $(a_{i0},\bar{u}_{i0}),\dots,(a_{in},\bar{u}_{in})$. To explicitly characterize such a representation, we consider a collection of d-tuples defined as follows:

$$\mathcal{J} := \bigl\{(J_1,\dots,J_d) \bigm| \{0,n\} \subseteq J_i \subseteq \{0,\dots,n\},\ i=1,\dots,d\bigr\}, \tag{22}$$

and, for each tuple $J \in \mathcal{J}$, we define a linear map, $\Gamma_J : \mathbb{R}^{d\times(n+1)} \to \mathbb{R}^{d\times(n+1)}$, as follows:

$$\begin{aligned} \tilde{u}_{ij} &= u_{ij} && \text{for } i \in \{1,\dots,d\} \quad j \in J_i, \\ \tilde{u}_{ij} &= \gamma_{ij}u_{il(i,j)} + (1-\gamma_{ij})u_{ir(i,j)} && \text{for } i \in \{1,\dots,d\} \quad j \notin J_i, \end{aligned} \tag{23}$$

where $l(i,j) := \max\{j' \in J_i \mid j' \le j\}$, $r(i,j) := \min\{j' \in J_i \mid j' \ge j\}$, and $\gamma_{ij} = (a_{ir(i,j)} - a_{ij})/(a_{ir(i,j)} - a_{il(i,j)})$ for $j \notin J_i$. With each tuple $J \in \mathcal{J}$ we associate a subset $F_J := F_{1J_1} \times \cdots \times F_{dJ_d}$ of Q, where $F_{iJ_i} := \mathrm{conv}\bigl(\{v_{ij} \mid j \in J_i\}\bigr)$ and v_{ij} is defined as in (10). Clearly, F_J is a face of Q since F_{iJ_i} is a face of the simplex Q_i. It is also useful to observe that the face F_{iJ_i} can also be described as the set of points of Q_i that satisfy the following at equality:

$$\frac{s_{ij+1}-s_{ij}}{a_{ij+1}-a_{ij}} \le \frac{s_{ij}-s_{ij-1}}{a_{ij}-a_{ij-1}} \qquad \text{for } j \notin J_i. \tag{24}$$

In the next result, we show that, there exists a $J \in \mathcal{J}$ such that $\bar{s} = \Gamma_J(\bar{u})$ and $\bar{s}$ lies in the face F_J.

Proposition 5 *For each tuple $J \in \mathcal{J}$ and for each $(u,s) \in PQ$, $\bigl(\Gamma_J(u),s\bigr) \in PQ$. Moreover, inequalities $\Gamma_J(s) \le s$ define the face F_J. If $(\bar{u},\bar{s}) \in PQ'$ then $\bar{s} = \Gamma_{J'}(\bar{u})$ and $\bar{s} \in F_{J'}$, where $J' = (J'_1,\dots,J'_d)$ such that $J'_i := \{j \mid \bar{u}_{ij} = \bar{s}_{ij}\}$.*

Proof We start by showing that $\bigl(\Gamma_J(u),s\bigr) \in PQ$ for every $(u,s) \in PQ$. Let $(u,s) \in PQ$ and define $\tilde{u} := \Gamma_J(u)$. Clearly, for $i \in \{1,\dots,d\}$ and $j \in \{0,n\}$, we have $\tilde{u}_{ij} = u_{ij} = s_{ij}$, where first equality holds because $\{0,n\} \subseteq J_i$ and second equality holds because $(u,s) \in PQ$ implies that $u_{i0} = s_{i0}$ and $u_{in} = s_{in}$. Moreover, for all i and j, $a_{i0} \le u_{ij'}$ for every $j' \in \{0,\dots,n\}$ implies that $a_{i0} \le \tilde{u}_{ij}$. Last, for $i \in \{1,\dots,d\}$, $\tilde{u}_i \le s_i$ follows because, for $j \in \{0,\dots,n\}$, we have $\tilde{u}_{ij} \le \mathrm{conc}(\xi)(a_{ij};u_i) \le \hat{\xi}(a_{ij};s_i) = s_{ij}$, where first inequality holds because the point $(a_{ij},\tilde{u}_{ij})$ is expressible as a convex combination of the hypograph of $\xi(a;u_i)$, second inequality holds because, by $(u_i,s_i) \in PQ_i$ and Lemma 5, $\hat{\xi}(a;s_i)$ is a concave overestimator of $\xi(a;u_i)$, thus, of $\mathrm{conc}(\xi)(a;u_i)$, and the equality is by definition.

Next, we show that inequality $s \geq \Gamma_J(s)$ defines the face F_J of Q. Since, for every $s \in Q$, $(s, s) \in PQ$, the validity of $s \geq \Gamma_J(s)$ over Q follows from the first result. We need to show that $s = \Gamma_J(s)$ for $s \in \operatorname{vert}(F_J)$ and $s \neq \Gamma_J(s)$ for $s \in \operatorname{vert}(Q) \setminus \operatorname{vert}(F_J)$. Let $s' = (s'_1, \dots, s'_d) \in \operatorname{vert}(Q)$. Observe that s' satisfies $s \geq \Gamma_J(s)$ at equality if and only if, for $i \in \{1, \dots, d\}$, the associated extension function $\hat{\xi}(a; s'_i)$ is linear over $[a_{i\bar{j}_i}, a_{i\hat{j}_i}]$ for every $0 \leq \bar{j}_i \leq \hat{j}_i \leq n$ such that $\bar{j}_i, \hat{j}_i \in J_i$ and $J_i \cap \{\bar{j}_i + 1, \dots, \hat{j}_i - 1\} = \emptyset$. Observe that $\hat{\xi}(a; v_{ij})$ is linear over intervals that do not contain a_{ij}. If, for each i, there exists $j'_i \in J_i$ such that $s'_i = v_{ij'_i}$ then $\hat{\xi}(a; v_{ij'})$ is linear over $[a_{i\bar{j}_i}, a_{i\hat{j}_i}]$ because j'_i, being in J_i, does not belong to $\{\bar{j}_i + 1, \dots, \hat{j}_i - 1\}$. Therefore, the point s' satisfies $s \geq \Gamma_J(s)$ at equality. On the other hand, assume that for some $i \in \{1, \dots, d\}$, there exists $j'' \notin J_i$, so that $s'_i = v_{ij''}$. Since $\{0, n\} \subseteq J_i$ and $j'' \notin J_i$, there exist $\bar{j}_i, \hat{j}_i \in J_i$ such that $\bar{j}_i < j'' < \hat{j}_i$, $\hat{\xi}(a; v_{ij''_i})$ is not linear over $[a_{i\bar{j}_i}, a_{i\hat{j}_i}]$, and $s' \neq \Gamma_J(s')$.

Last, we prove the third result. Let $(\bar{u}, \bar{s}) \in PQ'$, and define $(\bar{z}_1, \dots, \bar{z}_d) = \big(Z_1(\bar{s}_1), \dots, Z_d(\bar{s}_d)\big)$, where Z_i is defined as in (12). Because $\bar{s}_i \in Q_i$, $0 \leq \bar{z}_{in} \leq \cdots \leq \bar{z}_{i1} \leq \bar{z}_{i0} = 1$. We will show that that $\bar{z}_{ij+1} = \bar{z}_{ij}$ for all i and $j \notin J'_i$. Thus, by (24), $\bar{s} \in F_{J'}$, and, therefore, $\bar{s} = \Gamma_{J'}(\bar{s}) = \Gamma_{J'}(\bar{u})$, where first equality holds by the second result and second equality holds by the definition of $\Gamma_{J'}$ and $\bar{s}_{ij} = \bar{u}_{ij}$ for all i and $j \in J'_i$. Now, we show that $\bar{z}_{ij+1} = \bar{z}_{ij}$ for all $j \notin J'_i$. By definition, there are $\bar{j}$, $\hat{j}$ and $\gamma \geq 0$ such that $\bar{s}_{ij} = \gamma \bar{u}_{i\bar{j}} + (1 - \gamma)\bar{u}_{i\hat{j}}$. Since $j \notin J'_i$, we can assume that $0 \leq \bar{j} < j < \hat{j} \leq n$. If not, assume wlog that $j = \hat{j}$. Then, $0 = a_{i\hat{j}} - a_{ij} = \gamma(a_{i\hat{j}} - a_{i\bar{j}})$ implies that either $\gamma = 0$ or $a_{i\hat{j}} = a_{ij} = a_{i\bar{j}}$. In either case, $\bar{s}_{ij} = \bar{u}_{ij}$, contradicting that $j \notin J'_i$. Then, it follows that $(a_{ij}, \bar{s}_{ij})$, $(a_{i\bar{j}}, \bar{u}_{i\bar{j}})$, and $(a_{i\hat{j}}, \bar{u}_{i\hat{j}})$ are collinear. Let $L(a)$ be the function passing through these points. It follows that:

$$
\begin{aligned}
L(a_{ij}) &\geq \gamma' \operatorname{conc}(\xi)(a_{ij-1}, \bar{u}_i) + (1 - \gamma') \operatorname{conc}(\xi)(a_{ij+1}, \bar{u}_i) \\
&\geq \gamma' L(a_{ij-1}) + (1 - \gamma') L(a_{ij+1}) = L(a_{ij}),
\end{aligned}
$$

where $\gamma' = \frac{a_{ij+1} - a_{ij}}{a_{ij+1} - a_{ij-1}}$. The first inequality is because $L(a_{ij}) = \operatorname{conc}(\xi)(a_{ij}, \bar{u}_i)$ and $\operatorname{conc}(\xi)$ is concave, second inequality is because $\gamma' \in [0, 1]$ and concavity of $\operatorname{conc}(\xi)$ implies that $\operatorname{conc}(\xi)(a_{ij-1}, \bar{u}_i) \geq L(a_{ij-1})$ and $\operatorname{conc}(\xi)(a_{ij+1}, \bar{u}_i) \geq L(a_{ij+1})$, and the equality is because of linearity of L. Therefore, equality holds throughout and $\bar{s}_{ij} = L(a_{ij}) = \gamma' \operatorname{conc}(\xi)(a_{ij-1}, \bar{u}_i) + (1 - \gamma') \operatorname{conc}(\xi)(a_{ij+1}, \bar{u}_i) = \gamma' \bar{s}_{ij-1} + (1 - \gamma')\bar{s}_{ij+1}$, thereby showing that $\bar{z}_{ij+1} = \bar{z}_{ij}$. □

Let $\langle \alpha, s \rangle + \beta\phi + b \geq 0$ be a valid inequality generated for $\operatorname{conv}(\Phi^Q)$. We propose an algorithm, Algorithm 2, to generate another valid inequality $\langle \alpha', s \rangle + \beta\phi + b \geq 0$ for $\operatorname{conv}(\Phi^Q)$ such that $\alpha'_{ij} \leq 0$ for all i and $j \notin \{0, n\}$ and $\langle \alpha, s \rangle \geq \langle \alpha', s \rangle$ for every $s \in Q$. In words, given a valid inequality for $\operatorname{conv}(\Phi^Q)$, the algorithm generates a valid inequality that dominates the original one over Φ^Q and satisfies the sign condition discussed above. The key step in Algorithm 2, the while-loop spanning Steps 13–26, is to modify a positive coefficient to zero in a manner that this positive weight is allocated to its non-zero adjacent coefficients. In the next proposition, we show the correctness of Algorithm 2 and discuss its complexity.

Algorithm 2 Sign Procedure

1: **procedure** SIGN(α)
2: **for** i from 1 to d **do**
3: **for** j from 1 to $n-1$ **do**
4: **if** $\alpha_{ij} > 0$ **then**
5: push(j, J_i^+)
6: **end if**
7: **end for**
8: **for** j from 0 to n **do**
9: $\mathrm{prev}(j) = j-1$
10: $\mathrm{succ}(j) = j+1$
11: $\alpha'_{ij} = \alpha_{ij}$;
12: **end for**
13: **while** $J_i^+ \neq \emptyset$ **do**
14: $j = \mathrm{pop}(J_i^+)$
15: $\vartheta_{ij} = \frac{a_{i\,\mathrm{succ}(j)} - a_{ij}}{a_{i\,\mathrm{succ}(j)} - a_{i\,\mathrm{prev}(j)}}$
16: **for** (j', ϱ) in $\{(\mathrm{prev}(j), \vartheta_{ij}), (\mathrm{succ}(j), 1-\vartheta_{ij})\}$ **do**
17: neg $\leftarrow (\alpha'_{ij'} \leq 0)$;
18: $\alpha'_{ij'} = \alpha'_{ij'} + \varrho\alpha'_{ij}$;
19: **if** neg $=$ true and $\alpha'_{ij'} > 0$ and $0 < j' < n$ **then**
20: push(j', J_i^+)
21: **end if**
22: **end for**
23: $\mathrm{prev}(\mathrm{succ}(j)) = \mathrm{prev}(j)$; $\mathrm{succ}(\mathrm{prev}(j)) = \mathrm{succ}(j)$;
24: $\mathrm{prev}(j) = -1$; $\mathrm{succ}(j) = n+1$;
25: $\alpha'_{ij} = 0$;
26: **end while**
27: **end for**
28: **return** α'.
29: **end procedure**

Proposition 6 *Given a valid inequality $\langle \alpha, s\rangle + \beta\phi + b \geq 0$ for* $\mathrm{conv}(\Phi^Q)$*, Algorithm 2 takes $\mathcal{O}(dn)$ time to generate a valid inequality $\langle \alpha', s\rangle + \beta\phi + b \geq 0$ for* $\mathrm{conv}(\Phi^Q)$ *such that $\alpha'_{ij} \leq 0$ for all i and $j \notin \{0, n\}$ and $\langle \alpha, s\rangle \geq \langle \alpha', s\rangle$ for every $s \in Q$. Moreover, there exists an $\tilde{s} \in Q$ such that, for all i, $\tilde{s}_{in} = s_{in}$ and $\langle \alpha, \tilde{s}\rangle = \langle \alpha', s\rangle$.*

Proof In Step 5, we initialize J_i^+ as a stack of indices j such that $\alpha_{ij} > 0$ and in Steps 9 and 10, we initialize a queue that contains all the indices in $\{0, \dots, n\}$. In Step 14 of Algorithm 2, we pick $j \in J_i^+$ and remove it from the queue in Step 23. Indices are added to J_i^+ only in Step 20 and the added index belongs to $\{1, \dots, n-1\}$. We argue that at the beginning of each iteration of the while loop starting at Step 13, if an index $j \in J_i^+$ then $\alpha'_{ij} > 0$, otherwise $\alpha'_{ij} \leq 0$. Further, no index in J_i^+ is repeated, and if $j \in J_i^+$ then j is also in the queue. This is certainly true at first iteration after initialization ends at Step 12. At Step 20, j' is added only if $\alpha'_{ij'} > 0$ and was negative previously and, therefore, not already in J_i^+. Moreover, j' is in the queue, being either the predecessor or successor of an index j in the queue. At Step 18, it follows that $\alpha'_{ij'}$ can only increase because Step 16 guarantees that $\varrho \in [0, 1]$ since $\vartheta_{ij} \in [0, 1]$ and $\alpha_{ij} > 0$ because j was just removed from J_i^+. Therefore, existing indices in J_i^+ continue to satisfy the invariant regarding positive coefficients. Moreover, if $\alpha_{ij'}$ turns positive at Step 18, it is added to J_i^+ in Step 20. In Step 23, j is removed from the queue, but its only copy in J_i^+ was removed already at Step 14. Since the size of the queue reduces by one in each iteration of the while loop starting at Step 13 and $\{0, n\}$ remain in the queue throughout, the loop executes at most $n - 1$ iterations. Since the outer for loop starting at Step 2 executes d iterations, the complexity of the algorithm is $\mathcal{O}(dn)$. Further, α'_{ij}, for any j not in queue, is set to zero at Step 25 and never updated at Step 18 because j has already been removed from the queue and its only copy was removed from J_i^+ at Step 14. Observe that at termination J_i^+ is empty. Any remaining index j in the queue is such that $\alpha_{ij} \le 0$ and any index j outside the queue is such that $\alpha_{ij} = 0$.

We only need to establish the correctness of the i^{th} iteration of the outer for loop starting at Step 2. Let α^k be the α' at the start of k^{th} iteration of the while-loop spanning Steps 13–26, and, for notational convenience, assume $\alpha^1 = \alpha$. It follows easily that $\langle \alpha^k, s\rangle \le \langle \alpha^{k-1}, s\rangle$, because $s \in Q$ implies, by Lemma 2, that $s_{ij} \ge \vartheta_{ij} s_{i\mathrm{prev}(j)} + (1 - \vartheta_{ij}) s_{i\mathrm{succ}(j)}$. Therefore, $\langle \alpha', s\rangle \le \langle \alpha, s\rangle$. To complete the proof, we need to establish the validity of $\langle \alpha', s\rangle + \beta\phi + b \ge 0$ over Φ^Q. In particular, we will show that for every $(s, \phi) \in \Phi^Q$, there exists $(\tilde{s}, \phi) \in \Phi^Q$ such that, for all i, $\tilde{s}_{in} = s_{in}$ and $\langle \alpha, \tilde{s}\rangle = \langle \alpha', s\rangle$. By assumption, $\langle \alpha, \tilde{s}\rangle + \beta\phi + b \ge 0$, thereby proving that $\langle \alpha', s\rangle + \beta\phi + b \ge 0$. Let $(s, \phi) \in \Phi^Q$, and define $\tilde{s}^0 = s$. During the k^{th} iteration, at Step 14, j is popped from J_i^+, we define $\tilde{s}^k$ such that, for $\mathrm{prev}(j) < j' < \mathrm{succ}(j)$, $\tilde{s}^k_{ij'} = \vartheta_{ij'} \tilde{s}^{k-1}_{i\mathrm{prev}(j)} + (1 - \vartheta_{ij'}) \tilde{s}^{k-1}_{i\mathrm{succ}(j)}$ and $\tilde{s}^k_{ij'} = \tilde{s}^{k-1}_{ij'}$ otherwise. In addition, let H^k be the queue at the beginning of k^{th} iteration. Since $\tilde{s}^k_i$ and $\tilde{s}^{k-1}_i$ differ only at indices not in H^k, it follows easily that $\tilde{s}^k_{ij'} = s_{ij'}$ for $j' \in H^k$. This, together with $\alpha^k_{ij} = 0$ for $j \notin H^k$, implies that $\langle \alpha^k, s\rangle = \langle \alpha^k, \tilde{s}^k\rangle$. Next, we show that $\langle \alpha^{l+1}, \tilde{s}^k\rangle = \langle \alpha^l, \tilde{s}^k\rangle$ for $1 \le l \le k - 1$. In the l^{th} iteration of the while loop, assume that, during Step 14, j' is popped from J_i^+, $\mathrm{prev}(j') = \bar{j}$, and $\mathrm{succ}(j') = \hat{j}$. Let $\vartheta = \frac{a_{i\hat{j}} - a_{ij'}}{a_{i\hat{j}} - a_{i\bar{j}}}$. Since the queue becomes smaller with iterations and $l \le k - 1$, it follows that $j' \notin H^k$. Then, $\langle \alpha^{l+1}, \tilde{s}^k\rangle = \langle \alpha^l, \tilde{s}^k\rangle - \alpha^l_{ij'}\big(\tilde{s}^k_{ij'} - \vartheta \tilde{s}^k_{i\bar{j}} - (1 - \vartheta)\tilde{s}^k_{i\hat{j}}\big)$, where, by the definition of $\tilde{s}^k$, $\{\bar{j} + 1, \dots, \hat{j} - 1\} \notin H^k$, $\hat{\xi}(a; \tilde{s}^k)$ is linear between $[a_{i\bar{j}}, a_{i\hat{j}}]$ and, so, the term in the parenthesis on right-hand-side is zero. Moreover, by Proposition 5, $\tilde{s}^k \in Q$, and by $\{0, n\} \subseteq H^k$, $\tilde{s}^k_{ij} = s_{in}$. It follows by considering the last iteration, say r, of the while-loop that there exists $\tilde{s}^r \in Q$ such that $\tilde{s}^r_{in} = s_{in}$, $\alpha' = \alpha^r$, and $\langle \alpha^r, s\rangle = \langle \alpha^r, \tilde{s}^r\rangle = \langle \alpha, \tilde{s}^r\rangle$. Hence, $(\tilde{s}^r, \phi) \in \Phi^Q$ as $\tilde{s}^r \in Q$ and $\phi = \phi(s_{1n}, \dots, s_{dn}) = \phi(\tilde{s}^r_{1n}, \dots, \tilde{s}^r_{dn})$. □

Given $(\bar{u}, \bar{s}) \in PQ'$ and a valid inequality $\langle \alpha, s\rangle + \beta\phi + b \ge 0$ for $\mathrm{conv}(\Phi^Q)$ satisfying the sign condition guaranteed by Algorithm 2, the following lemma generates an inequality $\langle \alpha', u\rangle + \beta\phi + b \ge 0$ valid for $\mathrm{conv}(\Phi^P)$ so that $\langle \alpha, \bar{s}\rangle = \langle \alpha', \bar{u}\rangle$. In other words, if, for some $\bar{\phi} \in \mathbb{R}$, the inequality $\langle \alpha, s\rangle + \beta\phi + b \geq 0$ separates $(\bar{s}, \bar{\phi})$ from Φ^Q then the modified inequality separates $(\bar{u}, \bar{\phi})$ from Φ^P.

Lemma 6 *Let $\langle \alpha, s\rangle + \beta\phi + b \geq 0$ be an inequality valid for* $\mathrm{conv}(\Phi^Q)$. *Let $(\bar{u}, \bar{s}) \in PQ'$ and define $J = (J_1, \dots, J_d)$ such that $J_i = \{j \mid \bar{u}_{ij} = \bar{s}_{ij}\}$. Consider a linear function $\langle \alpha', \cdot\rangle$ such that $\langle \alpha', u\rangle = \big\langle \alpha, \Gamma_J(u)\big\rangle$, where Γ_J is the linear transformation defined in (23). If, for all $i \in \{1, \dots, d\}$ and $j \notin \{0, n\}$, α_{ij} is non-positive then, for every $(u, s) \in PQ$, $\langle \alpha, s\rangle \leq \langle \alpha', u\rangle$ and equality is attained at $(\bar{u}, \bar{s})$. Moreover, the inequality $\langle \alpha', u\rangle + \beta\phi + b \geq 0$ is valid for* $\mathrm{conv}(\Phi^P)$.

Proof Assume that $\alpha_{ij} \leq 0$ for all i and $j \notin \{0, n\}$. We first observe that, for every $(u, s) \in PQ$, $\langle \alpha, s\rangle \leq \big\langle \alpha, \Gamma_J(u)\big\rangle = \langle \alpha', u\rangle$, where the inequality holds because $\alpha_{ij} \leq 0$ for all i and $j \notin \{0, n\}$ and, by the first result in Proposition 5, $(\Gamma_J(u), s) \in PQ$, and the equality holds by the definition of α' in the statement of the result. In particular, $\langle \alpha, \bar{s}\rangle = \big\langle \alpha, \Gamma_J(\bar{u})\big\rangle = \langle \alpha', \bar{u}\rangle$, where first equality holds because, by the third result in Proposition 5, $\bar{s} = \Gamma_J(\bar{u})$.

Now, we show that $\langle \alpha', u\rangle + \beta\phi + b \geq 0$ is valid for $\mathrm{conv}(\Phi^P)$. Observe that the inequalities, $\langle \alpha', u\rangle \geq \langle \alpha, s\rangle$ and $\langle \alpha, s\rangle + \beta\phi + b \geq 0$, are valid for $\mathrm{conv}(\Phi^{PQ})$, where validity of the latter inequality follows since it is assumed to be valid for $\mathrm{conv}(\Phi^Q)$, which, by Lemma 3, equals $\mathrm{proj}_{(s,\phi)}\big(\mathrm{conv}(\Phi^{PQ})\big)$. This implies that $\langle \alpha', u\rangle + \beta\phi + b \geq 0$ is valid for $\mathrm{conv}(\Phi^{PQ})$ and, hence, for $\mathrm{conv}(\Phi^P)$ since it does not depend on s and, by Lemma 3, $\mathrm{conv}(\Phi^P) = \mathrm{proj}_{(u,\phi)}\big(\mathrm{conv}(\Phi^{PQ})\big)$. □

Now, we are ready to prove that the main result.

Theorem 4 *The separation problem of* $\mathrm{conv}(\Phi^P)$ *can be solved in $\mathcal{O}(dn)$ time besides a call to the separation oracle for* $\mathrm{conv}(\Phi^Q)$.

Proof Let $(\bar{u}, \bar{\phi}) \in \mathbb{R}^{d\times(n+1)+1}$. We assume that $\bar{u} \in P$ because, if not, we can separate $\bar{u}$ from P in $\mathcal{O}(dn)$ time by finding a facet-defining inequality of P that is violated at $\bar{u}$. Let $\bar{s} := (\bar{s}_1, \dots, \bar{s}_d)$, where $\bar{s}_i$ is the point returned by Algorithm 1 when $\bar{u}_i$ is provided as input. Observe that this step takes $\mathcal{O}(dn)$ time. Clearly, we have $(\bar{u}, \bar{s}) \in PQ'$, where PQ' is defined in (21). Now, we call the separation oracle for $\mathrm{conv}(\Phi^Q)$. If the oracle asserts that $(\bar{s}, \bar{\phi}) \in \mathrm{conv}(\Phi^Q)$ then, by equality in (16), $(\bar{u}, \bar{s}, \bar{\phi}) \in \mathrm{conv}(\Phi^{PQ})$, and thus, by Lemma 3, $(\bar{u}, \bar{\phi}) \in \mathrm{conv}(\Phi^P)$.

Now, suppose that $(\bar{s}, \bar{\phi}) \notin \mathrm{conv}(\Phi^Q)$ and the oracle returns a hyperplane $\langle \alpha, s\rangle + \beta\phi + b = 0$ such that, for all $(s, \phi) \in \mathrm{conv}(\Phi^Q)$, $\langle \alpha, s\rangle + \beta\phi + b \geq 0$, whereas $\langle \alpha, \bar{s}\rangle + \beta\bar{\phi} + b < 0$. In this case, we will derive a hyperplane that separates $(\bar{u}, \bar{\phi})$ from $\mathrm{conv}(\Phi^P)$. Using Proposition 6, we can assume that $\alpha_{ij} \leq 0$ for all i and $j \notin \{0, n\}$. Otherwise, we can satisfy this sign requirement by executing Algorithm 2, with complexity $\mathcal{O}(dn)$, using the generated inequality. Now, we utilize Lemma 6 to derive the inequality $\langle \alpha', u\rangle + \beta\phi + b \geq 0$, where α' can easily be computed in $\mathcal{O}(dn)$ time. Given arbitrary $(u, \phi) \in \mathrm{conv}(\Phi^P)$, it follows that $\langle \alpha', \bar{u}\rangle + \beta\bar{\phi} + b = \langle \alpha, \bar{s}\rangle + \beta\bar{\phi} + b < 0 \leq \langle \alpha', u\rangle + \beta\phi + b$, where the first equality is by construction of α' and Lemma 6, the strict inequality is guaranteed by the separation oracle for $\mathrm{conv}(\Phi^Q)$, and the last inequality follows Lemma 6. □

Next, we explore the strength of cuts that are generated by the procedure described in Theorem 4. We will only discuss the strength of valid cuts for the hypograph of

$\operatorname{conc}_P(\phi)(u)$ since a similar argument applies for $\operatorname{conv}_P(\phi)(u)$. Recall that we say an inequality $\phi \le \langle \alpha, u\rangle + b$, valid for $\operatorname{conv}(\Phi^P)$, is *tight* at a given point $\bar{u} \in P$ if $\operatorname{conc}_P(\phi)(\bar{u}) = \langle \alpha, \bar{u}\rangle + b$.

Proposition 7 *Assume that the concave envelope,* $\operatorname{conc}_P(\phi)(u)$*, is closed. Given a polynomial time separation oracle for* $\operatorname{conc}_Q(\phi)(s)$ *which yields tight cuts, there exists a polynomial time separation algorithm for* $\operatorname{conc}_P(\phi)(u)$ *which generates tight cuts.*

Proof Let $\bar{u} \in P$ and define $\bar{s} = (\bar{s}_1, \ldots, \bar{s}_d)$, where $\bar{s}_i$ is the point returned by Algorithm 1 when $\bar{u}_i$ is provided as the input. Suppose that the separation oracle generates a valid inequality $\phi \le \langle \alpha, s\rangle + b$ of $\operatorname{conc}_Q(\phi)(s)$, which is tight at $\bar{s}$. We assume without loss of generality that $\alpha_{ij} \le 0$ for all i and $j \notin \{0, n\}$ since otherwise we apply Algorithm 2 to generate a new inequality which is tight at $\bar{s}$ and satisfies the sign requirement. Let $\phi \le \langle \alpha', u\rangle + b$ be the inequality obtained using Lemma 6. Then,

$$\begin{aligned}\operatorname{conc}_Q(\phi)(\bar{s}) &= \langle \alpha, \bar{s}\rangle + b = \langle \alpha', \bar{u}\rangle + b \\ &\ge \operatorname{conc}_P(\phi)(\bar{u}) \ge \operatorname{conc}_P(\phi)(\bar{s}) \ge \operatorname{conc}_Q(\phi)(\bar{s}),\end{aligned}$$

where the first equality is because $\phi \le \langle \alpha, s\rangle + b$ is tight at $\bar{s}$, second equality holds because, by third result in Proposition 5, $\bar{s} = \Gamma_J(\bar{u})$ and, by Lemma 6, $\langle \alpha', \bar{u}\rangle = \big\langle \alpha, \Gamma_J(\bar{u})\big\rangle$, the first inequality holds because, by Lemma 6, $\phi \le \langle \alpha', u\rangle + b$ is valid for $\operatorname{conc}_P(\phi)(u)$, second inequality holds because, by Lemma 1, the closedness of $\operatorname{conc}_P(\phi)(u)$ implies that it is non-increasing in u_{ij} for all i and $j \notin \{0, n\}$ and $(\bar{u}, \bar{s}) \in PQ$ implies that $\bar{u} \le \bar{s}$ and $\bar{u}_{ij} = \bar{s}_{ij}$ for all i and $j \in \{0, n\}$, and the last inequality follows because $Q \subseteq P$ and thus $\operatorname{conc}_P(\phi)(s) \ge \operatorname{conc}_Q(\phi)(s)$ for every $s \in Q$. Therefore, equalities hold throughout. In particular, we obtain that $\langle \alpha', \bar{u}\rangle + b = \operatorname{conc}_P(\phi)(\bar{u}) = \operatorname{conc}_P(\phi)(\bar{s}) = \operatorname{conc}_Q(\phi)(\bar{s})$. ◻

We argued in the proof of Proposition 7 that, given a pair $(\bar{u}, \bar{s}) \in PQ'$, we have $\operatorname{conc}_P(\phi)(\bar{u}) = \operatorname{conc}_Q(\phi)(\bar{s})$. We denote by $\mathcal{L}(u_i)$ the feasible region of (18) with u_i as the given input. It is easy to see that, for every $u \in T(\bar{u}, \bar{s}) := \{u \mid \bar{u} \le u \le \bar{s}\}$, we have $(u, \bar{s}) \in PQ'$ because, for $i \in \{1, \ldots, d\}$, $\mathcal{L}(u_i) \subseteq \mathcal{L}(\bar{u}_i)$ and $\bar{s}_i \in \mathcal{L}(u_i)$. Therefore, the following result follows from the proof of Proposition 7.

Corollary 1 *Assume that the concave envelope* $\operatorname{conc}_P(\phi)(u)$ *is closed. Then, given* $(\bar{u}, \bar{s}) \in PQ'$*, we have* $\operatorname{conc}_P(\phi)(u) = \operatorname{conc}_Q(\phi)(\bar{s})$ *for every* $u \in T(\bar{u}, \bar{s})$*, where* $T(\bar{u}, \bar{s}) = \{u \mid \bar{u} \le u \le \bar{s}\}$.

4.3 Application in factorable programming

In this subsection, we show that the factorable programming scheme can be improved by considering a special case of the problem treated in Sects. 3 and 4. We consider the case when $\phi(\cdot)$ is a bilinear term and P is defined with $d = n = 2$.

Theorem 5 *Let $f_1^L \leq a_1 \leq f_1^U$ and $f_2^L \leq a_2 \leq f_2^U$. Then, consider the set:*

$$P = \big\{(u_1, f_1, u_2, f_2) \;\big|\; f_1^L \leq u_1 \leq \min\{f_1, a_1\},\, f_1 \leq f_1^U,\\ f_2^L \leq u_2 \leq \min\{f_2, a_2\},\, f_2 \leq f_2^U\big\}.$$

The following overestimators inequalities are valid for the epigraph of $f_1 f_2$ over P:

$$f_1 f_2 \leq \min \left\{ \begin{array}{l} r_1 := f_2^L f_1 + f_1^U f_2 - f_1^U f_2^L \\ r_2 := (f_2^L - a_2)u_1 + (a_1 - f_1^U)u_2 + a_2 f_1 + f_1^U f_2 - a_1 f_2^L \\ r_3 := (f_2^L - f_2^U)u_1 + a_1 f_2 + f_2^U f_1 - a_1 f_2^L \\ r_4 := (f_1^L - f_1^U)u_2 + a_2 f_1 + f_1^U f_2 - f_1^L a_2 \\ r_5 := (a_2 - f_2^U)u_1 + (f_1^L - a_1)u_2 + f_2^U f_1 + a_1 f_2 - f_1^L a_2 \\ r_6 := f_1 f_2^U + f_1^L f_2 - f_1^L f_2^U \end{array} \right\}.$$

Proof To show that r_2 is a valid overestimator, observe that $f_1 f_2 = r_2 - (a_1 - u_1)(u_2 - f_2^L) - (f_1 - u_1)(a_2 - u_2) - (f_2 - u_2)(f_1^U - f_1) \leq r_2$. Similarly, it follows that r_3, r_4, and r_5 are overestimators because $f_1 f_2 = r_3 - (a_1 - u_1)(f_2 - f_2^L) - (f_1 - u_1)(f_2 - f_2^U) \leq r_3$, $f_1 f_2 = r_4 - (f_1 - f_1^L)(a_2 - u_2) - (f_1^U - f_1)(f_2 - u_2) \leq r_4$, and $f_1 f_2 = r_5 - (u_1 - f_1^L)(a_2 - u_2) - (a_1 - u_1)(f_2 - u_2) - (f_1 - u_1)(f_2^U - f_2) \leq r_5$. □

Observe that Theorems 1 and 5 use a slightly different notation to denote P than the rest of Sects. 3–5. In particular, we hide the subscript j of u_{ij}, as it is unnecessary when there is a single underestimator. We also drop the subscript j of s_{ij} in the foregoing discussion.

The proof of Theorems 1 and 5 provided a direct verification of the validity of these inequalities and do not provide an intuition into how these inequalities were derived in the first place. We use the results in Sect. 4 to show a constructive derivation of one of the inequalities in Theorem 5. Incidentally, this construction also shows that the inequality is a facet of $\text{conv}(\Phi^P)$. A similar argument can be constructed for each of the inequalities showing that all the inequalities in Theorems 1 and 5 are facet-defining. Consider, for example, the inequality $f_1 f_2 \leq r_3$ and a point $p := (\bar{u}_1, \bar{f}_1, \bar{u}_2, \bar{f}_2) = (a_1, 0.5a_1 + 0.5 f_1^U, 0.5 f_2^L + 0.5a_2, 0.25 f_2^L + 0.75 f_2^U)$ that belongs to P. Algorithm 1 maps this point to $q := (\bar{s}_1, \bar{f}_1, \bar{s}_2, \bar{f}_2)$, a point in Q, with $\bar{s}_1 := \bar{u}_1$ and $\bar{s}_2 := \frac{f_2^U - a_2}{f_2^U - f_2^L} f_2^L + \frac{a_2 - f_2^L}{f_2^U - f_2^L}(0.25 f_2^L + 0.75 f_2^U) = 0.25 f_2^L + 0.75a_2$. Let $g_3 := (f_2^L - f_2^U)s_1 + a_1 f_2 + f_2^U f_1 - a_1 f_2^L$. The inequality $f_1 f_2 \leq g_3$ is tight at the extreme points of Q given by $(f_1^L, f_1^L, f_2^L, f_2^L)$, (a_1, a_1, f_2^L, f_2^L), (a_1, a_1, a_2, a_2), (a_1, a_1, a_2, f_2^U), and (a_1, f_1^U, a_2, f_2^U), where for each point, the variables are ordered as in (s_1, f_1, s_2, f_2). Then, we obtain the inequality $f_1 f_2 \leq g_3$ by interpolating $f_1 f_2$ at the listed vertices. The validity of this inequality over Q uses the same argument as the proof of validity of $f_1 f_2 \leq r_3$ in Theorem 5. Moreover, observe that q can be written as a convex combination of the above extreme points with multipliers 0, 0.25, 0, 0.25, and 0.5 respectively. Since the coefficient of s_1 in g_3 is already non-positive, we do not need to invoke Algorithm 2 to satisfy this sign-condition. Then, as in Lemma 6, we replace each s_i with its defining expression in terms of u_i to obtain an inequality that is valid for $\mathrm{conc}_P(\phi)$ and defines one of its facets. We provide more detail and illustrate the ideas involved. Normally, we would replace s_1 (resp. s_2) with u_1 (resp. $\frac{f_2^U-a_2}{f_2^U-f_2^L}f_2^L + \frac{a_2-f_2^L}{f_2^U-f_2^L}f_2$). However, since s_2 does not appear in the expression defining g_3, we obtain r_3 simply by substituting u_1 for s_1 in g_3, and thus derive the inequality, $f_1 f_2 \le r_3$, that is valid for $\mathrm{conc}_P(\phi)$. Using the points tight in Q, it follows easily that this inequality also defines a facet for $\mathrm{conc}_P(\phi)$. We next show that it is also tight at p, the point that was initially chosen for its derivation, thereby, demonstrating the more general fact shown in the proof of Proposition 7. Consider the point $r := (\bar{s}_1, \bar{f}_1, f_2^L, \bar{f}_2)$ that is a convex combination of tight points, as seen from the expression, $r = 0.25(a_1, a_1, f_2^L, f_2^L) + 0.25(a_1, a_1, f_2^L, f_2^U) + 0.5(a_1, f_1^U, f_2^L, f_2^U)$. Here, the points in the right hand side belong to P, are tight, and were obtained from the tight points in Q by substituting the s_2 coordinate with f_2^L. Then, since $p = \frac{1}{3}r + \frac{2}{3}q$, we have expressed p as a convex combination of tight points and shown that $f_1 f_2 \le r_3$ defines a facet for $\mathrm{conc}_P(\phi)$ and is tight at p.

Remark 4 Consider the bilinear product $f_1(x) f_2(x)$ and assume $f_1(x) \le o_1(x)$ yields an overestimator, $o_1(x)$. Then, Theorems 1 and 5 can be used by replacing f_1 with $-f_1$. In this case, the more involved transformation discussed in (7) is not necessary. Nevertheless, (7) is useful if besides the overestimator, $o_1(x)$, we also have an underestimator, $u_1(x)$, available for $f_1(x)$ and wish to exploit both estimators in the construction of cuts.

In the next example, we present a preliminary computational result that demonstrates the computational benefits of using inequalities from Theorems 1 and 5 on randomly generated problems.

Example 5 We consider a class of optimization problems of the form:

$$\begin{aligned} \min_{x,Y}\ & \langle c, x\rangle + Q \circ Y \\ \text{s.t.}\ & x \in [1,2]^n, \\ & y = \left(x_1^2, x_1^3, x_1^4, x_2^2, x_2^3, x_2^4, \ldots, x_n^2, x_n^3, x_n^4\right), \\ & Y = y^\top y, \end{aligned}$$

where $\circ$ denote the component-wise product of two matrices, and Q is a strictly upper triangular matrix. Clearly, the number of nonlinear monomials in the problem is determined by the number of variables n and the density ν of Q, *i.e.* the number of nonzero elements divided by the number of strictly upper triangular entries, that is $\left((3n)^2 - 3n\right)/2$. For a given pair (n, ν), we generated 50 problem instances, where the problem data were randomly generated as follows. The coefficients $(c_i)_{i=1}^n$ were uniformly generated from the interval $[-512, -2]$, while the strictly upper triangular entries of Q were zero with a probability of $1-\nu$ and uniformly generated from the interval $[1, 2]$ with a probability of ν.

For each problem instance, we constructed two linear programming (LP) relaxations. For the first relaxation, we used factorable programming. Here, each monomial $x_i^a x_j^b$ was relaxed using McCormick inequalities and each univariate function x_i^a was

Table 1 Factorable relaxation gap closed by Theorems 1 and 5. Each gap closed is an average over 50 problem instances

$n = 5$		$n = 10$		$n = 20$	
ν	Gap Closed (%)	ν	Gap Closed (%)	ν	Gap Closed (%)
0.1	67	0.05	65	0.025	61
0.2	59	0.1	53	0.05	49
0.3	45	0.15	44	0.075	40

outer-approximated using subgradient inequalities at 11 points $\{1, 1.1, \dots, 1.9, 2\}$. In the second relaxation, for each monomial term $x_i^a x_j^b$, we used Theorems 1 and 5 to generate inequalities for each pair of subgradient inequalities. Clearly, the second relaxation is tighter than the first. Our computational results are presented in Table 1. The table reports the percentage of factorable relaxation gap that is closed by using the second relaxation, that is, we compute $\frac{L_2 - L_1}{U - L_1}$, where L_1 and L_2 is the lower bound obtained by using the first and second relaxation, respectively, while U is the best upper bound obtained by `SCIP` [16] within a time limit of 500s. Observe that not all problems were solved to global optimality in 500s by `SCIP`; therefore, the percentage of gap closed reported is an underestimate of the actual gap closed. □

5 Extensions: simultaneous hull and facet-generation

In this section, we consider a *vector* of functions $\theta : \mathbb{R}^d \to \mathbb{R}^k$ over the polytope P and their graph

$$\Theta^P := \big\{(u, \theta) \;\big|\; \theta = \theta(u_{1n}, \dots, u_{dn}),\; u \in P\big\}.$$

We refer to $\mathrm{conv}(\Theta^P)$ as the *simultaneous hull* of Θ^P and are interested in solving the separation problem associated with this set. We, similarly, define $\Theta^Q := \big\{(s, \theta) \;\big|\; \theta = \theta(s_{1n}, \dots, s_{dn}),\; s \in Q\big\}$ and generalize the results from Sect. 4.2 to this setting. More specifically, we will show that, given a polynomial time separation oracle for $\mathrm{conv}(\Theta^Q)$, the separation problem for $\mathrm{conv}(\Theta^P)$ can also be solved in polynomial time. We will also prove a sharper result when $\mathrm{conv}(\Theta^Q)$ is a polytope, such as is the case when $\theta(\cdot)$ is a vector of multilinear functions (see Corollary 2.7 in [43]). For this setting, we will assume that we have access to a polynomial time oracle that generates facet-defining cuts for a family of lower-dimensional polytopes of $\mathrm{conv}(\Theta^Q)$, using which we will generate facet-defining cuts for $\mathrm{conv}(\Theta^P)$ in polynomial time.

5.1 Polynomial time equivalence of separations for simultaneous hulls

In this subsection, we show that proof techniques for Theorem 4 can be generalized to establish polynomial time equivalence of separations between $\mathrm{conv}(\Theta^P)$ and $\mathrm{conv}(\Theta^Q)$. First, Lemma 3 and the equality in (16) can be easily generalized to the context of $\mathrm{conv}(\Theta^P)$ using a similar argument. Let Θ^{PQ} be the graph of θ over polytope PQ, that is, $\Theta^{PQ} := \{(u, s, \theta) \mid \theta = \theta(s_{1n}, \ldots, s_{dn}),\ (u, s) \in PQ\}$.

Lemma 7 $\mathrm{proj}_{(u,\theta)}(\Theta^{PQ}) = \Theta^P$ *and* $\mathrm{proj}_{(s,\theta)}(\Theta^{PQ}) = \Theta^Q$. *Moreover, it holds that* $\mathrm{conv}(\Theta^P) = \{(u, \theta) \mid (s, \theta) \in \mathrm{conv}(\Theta^Q), (u, s) \in PQ\}$ □

Now, observe that Algorithm 2 can be applied to modify coefficients of a valid inequality for $\mathrm{conv}(\Theta^Q)$.

Proposition 8 *Given a valid inequality* $\langle \alpha, s\rangle + \langle \beta, \theta\rangle + b \geq 0$ *for* $\mathrm{conv}(\Theta^Q)$*, Algorithm 2 generates a valid inequality* $\langle \alpha', s\rangle + \langle \beta, \theta\rangle + b \geq 0$ *for* $\mathrm{conv}(\Theta^Q)$ *so that* $\alpha'_{ij} \leq 0$ *for all* i *and* $j \notin \{0, n\}$ *and* $\langle \alpha', s\rangle \leq \langle \alpha, s\rangle$ *for every* $s \in Q$.

Proof Let α' be the vector returned by Algorithm 2 when α is given as input. Then, for any $(s, \theta) \in \Theta^Q$, there exists $\tilde{s} \in Q$ such that $\langle \alpha, s\rangle \geq \langle \alpha', s\rangle = \langle \alpha, \tilde{s}\rangle \geq -\langle \beta, \theta\rangle - b$, where the first inequality, first equality, and the existence of $\tilde{s}$ follow from Proposition 6. The second inequality follows because $(\tilde{s}, \theta) \in \Theta^Q$ and $\langle \alpha, s\rangle + \langle \beta, \theta\rangle + b \geq 0$ is assumed to be valid for $\mathrm{conv}(\Theta^Q)$. □

Theorem 6 *The separation problem of* $\mathrm{conv}(\Theta^P)$ *can be solved in polynomial time given a polynomial separation oracle for* $\mathrm{conv}(\Theta^Q)$.

Proof The proof is similar to that of Theorem 4. We construct $\bar{s}$ using Algorithm 1 with $\bar{u}$ as input. If $(\bar{s}, \bar{\theta}) \in \mathrm{conv}(\Theta^Q)$ then, since $(\bar{u}, \bar{s}) \in PQ$, Lemma 7 shows that $(\bar{u}, \bar{\theta}) \in \mathrm{conv}(\Theta^P)$. If $(\bar{s}, \bar{\theta}) \notin \mathrm{conv}(\Theta^Q)$, we use the separation oracle of $\mathrm{conv}(\Theta^Q)$ and Algorithm 2 and Proposition 8 to obtain an inequality $\langle \alpha, s\rangle + \langle \beta, \theta\rangle + b \geq 0$ valid for $\mathrm{conv}(\Theta^Q)$ that separates $(\bar{s}, \bar{\theta})$ from $\mathrm{conv}(\Theta^Q)$. Then, we use the transformation of Lemma 6 to obtain α' and observe that, for all $(u, \theta) \in \mathrm{conv}(\Theta^P)$ and $(u, s) \in PQ$, $\langle \alpha', u\rangle + \langle \beta, \theta\rangle + b \geq \langle \alpha, s\rangle + \langle \beta, \theta\rangle + b \geq 0$. Since, by Lemma 6, $\langle \alpha', \bar{u}\rangle = \langle \alpha, \bar{s}\rangle < -\langle \beta, \bar{\theta}\rangle - b$, the inequality is not satisfied at $(\bar{u}, \bar{\theta})$ and, thus, separates $(\bar{u}, \bar{\theta})$ from $\mathrm{conv}(\Phi^P)$. □

5.2 Polynomial time equivalence of facet generations for simultaneous hulls

In the following, we will show that the *facet generation problem* of $\mathrm{conv}(\Theta^P)$ can be solved in polynomial time, given a facet generation oracle for $\mathrm{conv}(\Theta^{Q_J})$ for $J \in \mathcal{J}$. By a facet generation problem for a polyhedron S, we mean that in addition to the separation problem of S, we return a hyperplane that contains S and does not contain x if $x \notin \mathrm{aff}(S)$, or return a facet-defining inequality of S that is not satisfied at x.

We start by formally defining a family of polytopes of the form Q that result when subsets of the outer-approximators of $f_i(x)$ are considered. Consider $J = (J_1, \ldots, J_d) \in \mathcal{J}$, where $\mathcal{J}$ is defined as in (22). For any $y \in \mathbb{R}^{d\times(n+1)}$, let $y_J := (y_{1J_1}, \ldots, y_{dJ_d})$, where y_{iJ_i} are the components of y_i corresponding to the index J_i. Let $\bar{J} := (\bar{J}_1, \ldots, \bar{J}_d)$, where $\bar{J}_i$ is the complement of J_i, *i.e.*, $\bar{J}_i = \{0, \ldots, n\}\backslash J_i$. Using these definitions, we can now write, up to reordering of variables, that $y = (y_J, y_{\bar{J}})$. Let $a = (a_1, \ldots, a_d)$ be a vector in $\mathbb{R}^{d\times(n+1)}$ so that a_i is strictly increasing for every $i \in \{1, \ldots, d\}$. For any $J = (J_1, \ldots, J_d) \in \mathcal{J}$, we define $Q_J := Q_{1J_1} \times \cdots \times Q_{dJ_d}$, where Q_{iJ_i} is the simplex defined in (10) with parameter $a_{iJ_i} \in \mathbb{R}^{|J_i|}$, and consider the graph of $\theta(\cdot)$ over Q_J defined as $\Theta^{Q_J} := \big\{(s_J, \theta) \bigm| \theta = \theta(s_{1n}, \dots, s_{dn}),\ s_J \in Q_J\big\}$. With the collection of d-tuples $\mathcal{J}$, we also associate a family of faces of Q, that is, for each $J \in \mathcal{J}$, $F_J := F_{1J_1} \times \cdots \times F_{dJ_d}$, where $F_{iJ_i} := \operatorname{conv}\big(\{v_{ij} \mid j \in J_i\}\big)$ and $v_{ij} = (a_{i0}, \dots, a_{ij-1}, a_{ij}, \dots, a_{ij})$ for all i and j. Similarly, we consider the graph of $\theta(\cdot)$ over F_J defined as $\Theta^{F_J} := \big\{(s, \theta) \bigm| \theta = \theta(s_{1n}, \dots, s_{dn}),\ s \in F_J\big\}$. In the next result, we will provide the invertible affine isomorphism relating points in polytopes $\operatorname{conv}(\Theta^{Q_J})$ with those in $\operatorname{conv}(\Theta^{F_J})$. Recall that two polytopes $X \subseteq \mathbb{R}^m$ and $Y \subseteq \mathbb{R}^n$ are *affinely isomorphic* if there is an affine map $f : \mathbb{R}^m \to \mathbb{R}^n$ that is a bijection between the points of the two sets. Let $f : \mathbb{R}^m \to \mathbb{R}^k$ and X be a convex subset of $\mathbb{R}^m$. For any valid inequality $\langle \alpha, x\rangle + \langle \beta, \mu\rangle + b \geq 0$ of the convex hull of $\big\{(x, \mu) \mid \mu = f(x),\ x \in X\big\}$, we shall denote by $T_f^{(\alpha,\beta,b)}(X)$ the face of $\operatorname{conv}\big(\big\{(x, \mu) \bigm| \mu = f(x),\ x \in X\big\}\big)$ defined by the valid inequality.

Lemma 8 *Assume that* $\operatorname{conv}(\Theta^Q)$ *is a polytope. Let* $J = (J_1, \dots, J_d) \in \mathcal{J}$ *and let* $\langle \alpha, s\rangle + \langle \beta, \theta\rangle + b \geq 0$ *be a valid inequality of* $\operatorname{conv}(\Theta^Q)$ *so that* $\alpha_{\bar{J}} = 0$. *Then, the face* $T_\theta^{(\alpha,\beta,b)}(F_J)$ *of* $\operatorname{conv}(\Theta^{F_J})$ *is affinely isomorphic to the face* $T_\theta^{(\alpha_J,\beta,b)}(Q_J)$ *of* $\operatorname{conv}(\Theta^{Q_J})$.

Proof Let $J = (J_1, \dots, J_d) \in \mathcal{J}$. Since F_J is a face of Q and $\operatorname{conv}(\Theta^Q)$ is a polytope, it follows readily that $\operatorname{conv}(\Theta^{F_J})$ is a polytope. We first show that $\operatorname{conv}(\Theta^{F_J})$ is affinely isomorphic to $\operatorname{conv}(\Theta^{Q_J})$. Consider an affine map $A : s_J \to t$ such that $t_{ij} = s_{ij}$ for $j \in J_i$ and $t_{ij} = (1 - \gamma_{ij})s_{il(i,j)} + \gamma_{ij}s_{ir(i,j)}$ for $j \notin J_i$, where $l(i, j) = \max\{j' \in J_i \mid j' \leq j\}$, $r(i, j) = \min\{j' \in J_i \mid j' \geq j\}$, and $\gamma_{ij} = (a_{ij} - a_{il(i,j)})/(a_{ir(i,j)} - a_{il(i,j)})$. It follows from second result in Proposition 5 that A maps the polytope Q_J into the face F_J. The inverse of A is defined as $s \to s_J$ and maps the face F_J into the simplex Q_J. This is because, for any $s_i \in \operatorname{vert}(F_{iJ_i})$, there exists a $k \in J_i$ such that $s_{ij} = \min\{a_{ij}, a_{ik}\}$ for all $j \in J_i$. Thus, $s_{iJ_i} \in \operatorname{vert}\big(Q_{iJ_i}\big)$. Consider the affine transformation, Π, defined as $(s_J, \theta) \to \big(A(s_J), \theta\big)$ and its inverse, Π^{-1}, $(s, \theta) \to (s_J, \theta)$. Note that, in calling the projection operation as an inverse of Π, we are interpreting Π as a transformation into the affine hull of Θ^{F_J} rather than into Θ^Q. In other words, $\Pi\Theta^{Q_J} = \Theta^{F_J}$ and $\Pi^{-1}\Theta^{F_J} = \Theta^{Q_J}$. Therefore, $\operatorname{conv}(\Theta^{F_J}) = \operatorname{conv}(\Pi\Theta^{Q_J}) = \Pi\operatorname{conv}(\Theta^{Q_J})$ and, similarly, $\operatorname{conv}(\Theta^{Q_J}) = \Pi^{-1}\operatorname{conv}(\Theta^{F_J})$. It follows that $\operatorname{conv}(\Theta^{F_J})$ is affinely isomorphic to $\operatorname{conv}(\Theta^{Q_J})$.

Now, let $\langle \alpha, s\rangle + \langle \beta, \theta\rangle + b \geq 0$ be a valid inequality for $\operatorname{conv}(\Theta^Q)$ so that $\alpha_{\bar{J}} = 0$. Then, the validity of the inequality $\langle \alpha_J, s_J\rangle + \langle \beta, \theta\rangle + b \geq 0$ for $\operatorname{conv}(\Theta^{Q_J})$ follows since $\alpha_{\bar{J}} = 0$. Clearly, the corresponding faces, $T_\theta^{(\alpha,\beta,b)}(F_J)$ and $T_\theta^{(\alpha_J,\beta,b)}(Q_J)$, are affinely isomorphic under the mapping Π. □

As a consequence of Proposition 8, we obtain the following monotonic property for a facet-defining inequality of $\operatorname{conv}(\Theta^Q)$.

Lemma 9 *Assume that* $\operatorname{conv}(\Theta^Q)$ *is a polytope. Let* $\langle \alpha, s\rangle + \langle \beta, \theta\rangle + b \geq 0$ *be a face-defining inequality of* $\operatorname{conv}(\Theta^Q)$. *Then,* $\alpha_{ij} \leq 0$ *for all* i *and* $j \notin \{0, n\}$. □

Assume that $\operatorname{conv}(\Theta^Q)$ is a polytope and observe that $\operatorname{conv}(\Theta^{PQ})$ is a polytope since, by the second part of Lemma 7, $\operatorname{conv}(\Theta^{PQ}) = \big\{(u, s, \theta) \bigm| (s, \theta) \in$

$\text{conv}(\Theta^Q),\ (u,s)\in PQ\big\}$. Consequently, by the first part of Lemma 7, $\text{conv}(\Theta^P)$ is a polytope.

Theorem 7 *Assume that* $\text{conv}(\Theta^Q)$ *is a polytope. The facet generation problem of* $\text{conv}(\Theta^P)$ *can be solved in polynomial time if there exists a polynomial time facet generation oracle of* $\text{conv}(\Theta^{Q_J})$ *for every* $J\in\mathcal{J}$.

Proof Let $(\bar{u},\bar{\theta})\in\mathbb{R}^{d\times(n+1)+k}$ and assume that $\bar{u}\in P$. Let $\bar{s}$ be the point returned by Algorithm 1 with $\bar{u}$ as input, and define $J=\{J_1,\dots,J_d\}$, where $J_i:=\{j\mid \bar{u}_{ij}=\bar{s}_{ij}\}$. We assume that $(\bar{s},\bar{\theta})\notin\Theta^Q$; otherwise $(\bar{u},\bar{\theta})\in\Theta^P$ as in the proof of Theorem 6. We only consider the case when the facet-generation oracle produces a non-vertical facet-defining inequality, $\langle\alpha_J,s_J\rangle+\langle\beta,\theta\rangle+b\ge 0$, of $\text{conv}(\Theta^{Q_J})$ with $\langle\alpha_J,\bar{s}_J\rangle+\langle\beta,\bar{\theta}\rangle+b<0$. Instead, if the oracle returns a hyperplane which contains Θ^{Q_J}, a similar proof without the need for the point $(\tilde{s}_J,\tilde{\theta})$, which we define later, can be constructed easily. By Lemma 9, $\alpha_{ij}\le 0$ for all i and $j\notin J_i\backslash\{0,n\}$. Let $\tilde{\alpha}_J=\alpha_J$ and $\tilde{\alpha}_{\bar{J}}=0$. Then, it suffices to show that $\langle\tilde{\alpha},u\rangle+\langle\beta,\theta\rangle+b\ge 0$ is facet-defining for $\text{conv}(\Theta^P)$ since $\langle\tilde{\alpha},\bar{u}\rangle+\langle\beta,\bar{\theta}\rangle+b<0$.

To prove the validity of the inequality, we consider a point $(u,s,\theta)\in\Theta^{PQ}$ and observe that $\langle\tilde{\alpha},u\rangle\ge\langle\tilde{\alpha},s\rangle=\langle\tilde{\alpha}_J,s_J\rangle\ge-\langle\beta,\theta\rangle-b$, where the first inequality holds because, for all i and $j\notin\{0,n\}$, $\tilde{\alpha}_{ij}\le 0$ and $s_{ij}\ge u_{ij}$, and $s_{in}=u_{in}$, the first equality is because $\tilde{\alpha}_{\bar{J}}=0$, and the second inequality is because of the validity of $\langle\tilde{\alpha}_J,s_J\rangle+\langle\beta,\theta\rangle+b\ge 0$ for Θ^{Q_J}. Therefore, the inequality is valid for Θ^{PQ}, and hence, by Lemma 7, for Θ^P.

For simplicity of notation, for any set S we abbreviate $T_\theta^{(\tilde{\alpha},\beta,b)}(S)$ as $T(S)$. We will show that $\dim\big(T(P)\big)=\dim\big(\Theta^P\big)-1$, and thus, conclude that $(\tilde{\alpha},\beta,b)$ defines a facet of $\text{conv}(\Theta^P)$.

First, we construct a subset H of $T(P)$ so that, for $i'\in\{1,\dots,d\}$ and $j'\notin J_i$, there exist two points, $(\hat{s},\hat{\theta})$ and $(\check{s},\check{\theta})$, in H which differ only in coordinate corresponding to $s_{i'j'}$, that is, $(\hat{s},\hat{\theta})-(\check{s},\check{\theta})=(\delta e_{i'j'},0)$ where $\delta\ne 0$ and $e_{i'j'}$ is the j'^{th} principal vector in the i'^{th} subspace. Consider $(\hat{s}_J,\hat{\theta})\in\text{ri}\big(T(Q_J)\big)$ and observe that $\hat{s}_J\in\text{ri}(Q_J)$. If not, one of the inequalities defining Q_J is tight at all points in $T(Q_J)$ contradicting that (α_J,β,b) defines a non-vertical facet of $\text{conv}(\Theta^{Q_J})$. Now, we extend the point $\hat{s}_J$ to the point $\hat{s}$ of F_J using the transformation A defined as in the proof of Lemma 8. By Lemma 8, $\hat{s}\in\text{ri}(F_J)$ and $(\hat{s},\hat{\theta})\in\text{ri}\big(T(F_J)\big)$. It follows that $\hat{s}_{ij}>a_{i0}$ for all i and $j\ne 0$. Moreover, there exist a set of points $\big\{(s^k,\theta^k)\big\}_{k\in K}\subseteq T(F_J)\cap\Theta^{F_J}$ and convex multipliers $\lambda\in\mathbb{R}^{|K|}$ such that $(\hat{s},\hat{\theta})=\sum_{k\in K}\lambda_k(s^k,\theta^k)$. Now, let $\check{s}^k$ be a point so that $\check{s}^k_{ij}=s^k_{ij}$ for $(i,j)\ne(i',j')$ and $\check{s}^k_{ij}=a_{i0}$ otherwise. Since $\tilde{\alpha}_{i'j'}=0$ and $j'\ne n$, it follows that, for all $k\in K$, $(\check{s}^k,\theta^k)\in T(P)\cap\Theta^P$. Then, we define $(\check{s},\check{\theta}):=\sum_k\lambda_k(\check{s}^k,\theta^k)$. Therefore, by construction, $(\check{s},\check{\theta})\in T(P)$ and $(\hat{s},\hat{\theta})\in T(F_J)\subseteq T(P)$, and $(\hat{s},\hat{\theta})-(\check{s},\check{\theta})=\big((\hat{s}_{i'j'}-a_{i'0})e_{i'j'},0\big)\ne 0$.

Second, we argue that whenever $(s,\theta)\in\Theta^Q$ and $\delta_{ij}\in\mathbb{R}$, $i\in\{1,\dots,d\}$ and $j\notin\{0,n\}$, the point $\big(s+\sum_{i=1}^d\sum_{j\notin\{0,n\}}e_{ij}\delta_{ij},\theta\big)$ belongs to $\text{aff}\big(T(P)\cup(A(\tilde{s}_J),\tilde{\theta})\big)$ for some $(\tilde{s}_J,\tilde{\theta})\in\Theta^{Q_J}$. We start with the case when $\delta_{ij}=0$ for all i and $j\ne\{0,n\}$. It follows from Lemma 2 that the point $(s_J,\theta)\in\Theta^{Q_J}$. Thus, there exists a point $(\tilde{s}_J,\tilde{\theta})\in\Theta^{Q_J}$, not dependent on (s_J,θ), such that (s_J,θ) can be expressed as an affine combination of points in $T(Q_J) \cup (\tilde{s}_J, \tilde{\theta})$. Let Π be an affine mapping so that $\Pi(s_J, \theta) = \big(A(s_J), \theta\big)$. Then,

$$\begin{aligned}\Pi(s_J, \theta) \in \Pi\big(\operatorname{aff}\big(T(Q_J) \cup (\tilde{s}_J, \tilde{\theta})\big)\big) &= \operatorname{aff}\big(T(F_J) \cup (A(\tilde{s}_J), \tilde{\theta})\big)\\ &\subseteq \operatorname{aff}\big(T(P) \cup (A(\tilde{s}_J), \tilde{\theta})\big),\end{aligned}$$

where first inclusion holds because $(s_J, \theta) \in \operatorname{aff}\big(T(Q_J) \cup (\tilde{s}_J, \tilde{\theta})\big)$, the equality holds by Lemma 8 and by the fact that A, as an affine transformation, commutes with affine combinations, and the containment holds due to $T(F_J) \subseteq T(P)$. Because the point (s, θ) differs from $\big(A(s_J), \theta\big)$ only in coordinates corresponding to variables of the type s_{ij} for $i \in \{1, \dots, d\}$ and $j \notin J_i$, it follows from the existence of H that (s, θ) can be expressed as an affine combination of points in $T(P) \cup \big(A(\tilde{s}_J), \tilde{\theta}\big)$. Next, we prove the general case. Let $i'' \in \{1, \dots, d\}$ and $j'' \notin \{0, n\}$. Then, consider the point (s'', θ''), where $s''_i = v_{i0}$ for $i \neq i''$, $s''_{i''} = v_{i''j''}$ (see Lemma 2 for the definition of v_{ij}), and $\theta'' = \theta(s''_{1n}, \dots, s''_{dn})$. Clearly, there exists $\epsilon > 0$ so that two distinct points, (s'', θ'') and $(s'' - \epsilon e_{i''j''}, \theta'')$, belong to Θ^Q and are, as shown above, expressible as affine combinations of points in $T(P)$ and $\big(A(\tilde{s}_J), \tilde{\theta}\big)$. Since these points differ only in the variable $s_{i''j''}$, it follows that we can change the variable $s_{i''j''}$ for any $(s, \theta) \in \Theta^Q$ arbitrarily while remaining in the affine hull of $T(P) \cup \big(A(\tilde{s}_J), \tilde{\theta}\big)$.

Last, let $(\dot{u}, \dot{\theta}) \in \Theta^P$ and define $(\dot{s}, \dot{\theta})$ as a point so that $\dot{s}_i$ is the point returned by Algorithm 1 when $\dot{u}_i$ is provided as input. Then, it follows that $(\dot{s}, \dot{\theta}) \in \Theta^Q$ Since $\dot{u}_{ij} = \dot{s}_{ij}$ for all $i \in \{1, \dots, d\}$ and $j \in \{0, n\}$, it follows that there exists $(\tilde{s}_J, \tilde{\theta})$ so that $(\dot{u}, \dot{\theta})$ is expressible as an affine combination of $T(P) \cup \big(A(\tilde{s}_J), \tilde{\theta}\big)$. This shows that $\dim\big(T(P)\big) = \dim(\Theta^P) - 1$. □

Corollary 2 *Assume that* $\operatorname{conc}_Q(\phi)(s)$ *is a polyhedral function. Let* $\bar{u} \in P$ *and let* $\bar{s} = (\bar{s}_1, \dots, \bar{s}_d)$, *where* $\bar{s}_i$ *is returned by Algorithm 1 when it is provided with* $\bar{u}_i$ *as input. Define* $J = (J_1, \dots, J_d)$, *where* $J_i = \{j \mid \bar{u}_{ij} = \bar{s}_{ij}\}$. *Assume that there is an oracle that, given* $\bar{s}_J$, *generates a facet-defining inequality,* $\phi \leq \langle \alpha_J, s_J \rangle + b$ *of* $\operatorname{conc}_{Q_J}(\phi)$ *tight at* $\bar{s}_J$. *Define* $\tilde{\alpha} \in \mathbb{R}^{d(n+1)}$ *so that* $\tilde{\alpha}_J = \alpha_J$ *and* $\tilde{\alpha}_{\bar{J}} = 0$. *Then,* $\phi \leq \langle \tilde{\alpha}, u \rangle + b$ *is a facet-defining inequality for* $\operatorname{conc}_P(\phi)$ *that is tight at* $\bar{u}$. *Besides the call to the oracle, this inequality is generated in* $\mathcal{O}(dn)$ *time.*

Proof By Theorem 7, $\phi \leq \langle \tilde{\alpha}, s \rangle + b$ is facet-defining inequality for $\operatorname{conc}_P(\phi)(u)$. The proof is complete by observing that

$$\operatorname{conc}_P(\phi)(\bar{u}) = \operatorname{conc}_Q(\phi)(\bar{s}) = \operatorname{conc}_{F_J}(\phi)(\bar{s}) = \langle \tilde{\alpha}, \bar{s} \rangle + b = \langle \tilde{\alpha}, \bar{u} \rangle + b,$$

where first equality holds by Corollary 1, second equality follows from $\bar{s} \in F_J$, third equality holds by the assumed tightness property of the oracle and the affine isomorphism of Θ^{Q_J} and Θ^{F_J} shown in Lemma 8, and the last equality follows by $\tilde{\alpha}_{\bar{J}} = 0$ and $\bar{u}_J = \bar{s}_J$. □

6 Conclusions

In this paper, we tightened the factorable relaxation by proposing a new relaxation framework for mixed-integer nonlinear programs. The framework gives the first structured approach to relax composite functions using the inner-function structure. This is achieved by relaxing the outer-function over a polytope P that encapsulates information about the inner-functions implicit in their estimators. The structure of P is relatively complex in that its extreme points grow exponentially with the number of estimators, even when the number of inner-functions is fixed. Instead, we devised a fast combinatorial algorithm to solve the separation problem over P using an oracle to separate over Q, a much simpler subset of P. For vertex-generated outer-functions, with a fixed number of inner-functions, we gave a tractable polyhedral representation for the convex hull over P. When the outer-function is a bilinear term, and each inner-function has one estimator, we developed closed-form expressions for new valid inequalities, generalizing the factorable programming scheme. More specifically, if the inner-functions have n_1 and n_2 estimators, we derived $4n_1n_2+2n_1+2n_2$ inequalities besides the four McCormick inequalities. The new relaxations do not introduce variables beyond those used in the factorable scheme. In [22], we consider specially structured outer-functions for which convexification of the graph over Q is tractable, and using our results here, devise tractable algorithms for convexification over P. If the convex hull of the graph of outer-function over Q is polyhedral and there is a facet-generation oracle for this graph and some of its subsets, we constructed a facet-generation separation algorithm for this graph over P. Finally, we generalized our results to the setting involving a vector of outer-functions.

A Appendix

A.1 Proof of Theorem 2

Proof First, we show that $R = \mathrm{proj}_{(x,u_{\cdot n},\phi)}(\tilde{R})$, where

$$\tilde{R} := \left\{ (x,u,\phi) \;\middle|\; \begin{array}{l} \mathrm{conv}_P(\phi)(u) \le \phi \le \mathrm{conc}_P(\phi)(u),\; x \in X,\; u_{\cdot n} = f(x),\; u_{i0} = a_{i0} \\ u_{ij}(x) \le u_{ij}\; \forall j \in A_{ij}\backslash\{n\},\; u_{ij} \le u_{ij}(x)\; \forall j \in B_{ij},\; u \in P \end{array} \right\}.$$

Obviously, $R \subseteq \mathrm{proj}_{(x,u_{\cdot n},\phi)}(\tilde{R})$. To show that $\mathrm{proj}_{(x,u_{\cdot n},\phi)}(\tilde{R}) \subseteq R$, we consider a point $(x,u,\phi) \in \tilde{R}$. It follows readily that $u_{\cdot n} = f(x)$ and $x \in X$. Moreover, we have $\phi \le \mathrm{conc}_P(\phi)(u_1,\ldots,u_d) \le \mathrm{conc}_P(\phi)\big(\tilde{u}_1(x,u_{1n}),\ldots,\tilde{u}_d(x,u_{dn})\big)$, where the second inequality holds due to the monotonicity of $\mathrm{conc}_P(\phi)(u)$ shown in Lemma 1, and $u_{ij}(x) \le u_{ij}$ for $j \in A_{ij}\backslash\{n\}$, $u_{ij} \le u_{ij}(x)$ for $j \in B_{ij}$ and $\tilde{u}_{in}(x,u_{in}) = u_{in}$. A similar argument shows that $\mathrm{conv}_P(\phi)\big(\tilde{u}_1(x,u_{1n}),\ldots,\tilde{u}_d(x,u_{dn})\big) \le \phi$. Then, by (4), $\big(\tilde{u}_1(x,u_{1n}),\ldots,\tilde{u}_d(x,u_{dn}),\phi\big) \in \mathrm{conv}(\Phi^P)$. Hence, $(x,u_{\cdot n},\phi) \in R$.

Next, we show that $\mathrm{gr}(\phi\circ f) \subseteq \mathrm{proj}_{(x,\phi)}(\tilde{R})$, thus proving that $\mathrm{gr}(\phi\circ f) \subseteq \mathrm{proj}_{(x,\phi)}(R)$. Let $(x,\phi) \in \mathrm{gr}(\phi\circ f)$ and define $u = u(x)$, where, in particular, $u_{\cdot n} = f(x)$. It follows readily that $u \in P$ because, by construction, we have

$\{u \mid u = u(x), x \in X\} \subseteq P$. Moreover, since $(x, \phi) \in \mathrm{gr}(\phi \circ f)$, we have $\phi = \phi\big(f(x)\big) = \phi(u_{\cdot n})$, which implies that $(u, \phi) \in \Phi^P$ and, thus, $(u, \phi) \in \mathrm{conv}(\Phi^P)$. In other words, $\mathrm{conv}_P(\phi)(u) \le \phi \le \mathrm{conc}_P(\phi)(u)$. Therefore, $(x, u, \phi) \in \tilde{R}$ and, thus, $\mathrm{gr}(\phi \circ f) \subseteq \mathrm{proj}_{(x,\phi)}(\tilde{R})$.

Last, assume that $\{(x, u_{\cdot n}) \mid u_{\cdot n} = f(x),\ x \in X\}$ is outer-approximated by a convex set, and $u_{ij}(x)$ is either a convex underestimator or concave overestimator for all i and $j < n$. Then, $\tilde{R}$ is convex, and therefore, R is convex being a projection of $\tilde{R}$. □

A.2 Proof of Proposition 2

Proof Since $\Phi^{\hat{P}} \subseteq G$ and G is convex, it follows that $\mathrm{conv}(\Phi^{\hat{P}}) \subseteq G$. Now, we show that $\mathrm{conv}(\Phi^{\hat{P}}) \supseteq G$. Let $(t, \phi) \in G$. It follows that $t \in \hat{P}$ and there exist convex multipliers λ_k and $(t^k, \phi^k) \in \Phi^{\bar{P}}$ such that $(t, \phi) = \sum_k \lambda_k (t^k, \phi^k)$. Let $\bar{t}^k_{ij} = \max\big\{t^k_{ij}, u^L_i, t^k_{in} + a_{ij} - u^U_i\big\}$. We show that $(\bar{t}^k_{ij})^n_{j=0} \in \hat{P}_i$. This is easily verified since $\bar{t}^k_{in} = t^k_{in}$, $\bar{t}^k_{i0} = t^k_{i0}$, and $\max\big\{t^k_{ij}, u^L_i, t^k_{in} + a_{ij} - u^U_i\big\} \le \min\big\{t^k_{in}, a_{ij}\big\}$. Since $t^k_{in} = \bar{t}^k_{in}$, it follows that $(\bar{t}^k, \phi^k) \in \mathrm{conv}(\Phi^{\hat{P}})$. Then, $(\bar{t}, \phi) := \sum_k \lambda_k (\bar{t}^k, \phi^k) \in \mathrm{conv}(\Phi^{\hat{P}})$. However, $\bar{t} \in \hat{P}$ and $\bar{t} \ge t$. Therefore, $\mathrm{conv}(\Phi^{\hat{P}}) \supseteq G$ because $t \in \hat{P}$ and

$$\mathrm{conv}_{\hat{P}}(\phi)(t) \le \mathrm{conv}_{\hat{P}}(\phi)(\bar{t}) \le \phi \le \mathrm{conc}_{\hat{P}}(\phi)(\bar{t}) \le \mathrm{conc}_{\hat{P}}(\phi)(t).$$

The second and third inequality follow from $(\bar{t}, \phi) \in \mathrm{conv}(\Phi^{\hat{P}})$. We will now show the first inequality, and the argument for the last inequality is similar. To see that $\mathrm{conv}_{\hat{P}}(\phi)(t) \le \mathrm{conv}_{\hat{P}}(\phi)(\bar{t})$, let $\bar{u} = T^{-1}(\bar{t})$ and $u = T^{-1}(t)$ and observe that

$$\mathrm{conv}_{\hat{P}}(\phi)(t) = \mathrm{conv}_P(\phi)(u) \le \mathrm{conv}_P(\phi)(\bar{u}) = \mathrm{conv}_{\hat{P}}(\phi)(\bar{t}),$$

where the first and last equality is because convex envelopes do not change when the domain and argument undergo the same invertible affine transformation and the inequality follows from Lemma 1 for $-\phi$ because, for $j \in A_i \backslash \{0, n\}$ (resp. $j \in B_i$), $u_{ij} \le \bar{u}_{ij}$ (resp. $u_{ij} \ge \bar{u}_{ij}$), $u_{i0} = \bar{u}_{i0}$, and $u_{in} = \bar{u}_{in}$. The last statement in the result follows since T is an affine transformation and $T(u) \in \hat{P}$ if and only if $u \in P$. □

A.3 Proof of Proposition 3

Proof Let $(x, u_{\cdot n}, \phi) \in R\big(u(x), a\big)$. Define $u = u(x)$ and thus $(u, \phi) \in \mathrm{conv}\big(\Phi^{P(a)}\big)$. Let $u' = u'(x)$. By the definition of $u'(x)$, $u'_i = u_i(x)\Lambda_i$. To show that $R\big(u(x), a\big) \subseteq R\big(u'(x), a'\big)$, we only need to show that $(u', \phi) \in \mathrm{conv}\big(\Phi^{P(a')}\big)$. Let $\mathcal{A}$ be the affine transform used to obtain (u', ϕ) from (u, ϕ). Then, the result follows because:

$$(u', \phi) = \mathcal{A}(u, \phi) \in \mathcal{A}\,\mathrm{conv}\big(\Phi^{P(a)}\big) = \mathrm{conv}\big(\mathcal{A}\Phi^{P(a)}\big) \subseteq \mathrm{conv}\big(\Phi^{P(a')}\big),$$

where the first equality is by definition, the second equality is because $\mathcal{A}$, being an affine transform, commutes with convexification. All that remains to show is the final containment. This follows if we show that, for each $(\bar{u}, \bar{\phi}) \in \Phi^{P(a)}$, the point $(\bar{u}', \bar{\phi}) := \mathcal{A}(\bar{u}, \bar{\phi}) \in \Phi^{P(a')}$, where we recall that $\mathcal{A}$ is such that $\bar{u}'_i = \bar{u}_i \Lambda_i$. Observe that $a_{i0}e^{n+1}\Lambda_i \le \bar{u}_i\Lambda_i \le a_i\Lambda_i$ shows that $a_{i0}e^{n'+1} \le \bar{u}'_i \le a'_i$ and $\bar{u}_i\Lambda_i \le \bar{u}_{in+1}e^{n+1}\Lambda_i$ shows that $\bar{u}'_i \le \bar{u}_{in+1}e^{n'+1} = \bar{u}'_{in+1}e^{n'+1}$. Finally, since $\bar{u}_{in} = \bar{u}'_{in}$, it follows that $\bar{\phi} = \phi(\bar{u}_{1n}, \ldots, \bar{u}_{dn}) = \phi(\bar{u}'_{1n}, \ldots, \bar{u}'_{dn})$. Therefore, $(\bar{u}', \bar{\phi}) \in \Phi^{P(a')}$ and the result follows. $\square$

A.4 Proof of Lemma 2

Proof The inequality description follows from the preceding discussion since Q_i and Δ_i are related by an invertible affine transform, which maps v_{ij} to ζ_{ij} and the constraints in (14) transform those in (11). Now, we show the last statement in the result. For any $0 \le j < j' \le 1$, it follows that $\frac{s_{ij'}-s_{ij}}{a_{ij'}-a_{ij}} = \sum_{k=j+1}^{j'} \frac{s_{ik}-s_{ik-1}}{a_{ik}-a_{ik-1}} \frac{a_{ik}-a_{ik-1}}{a_{ij'}-a_{ij}}$, which is a convex combination of $\frac{s_{ik}-s_{ik-1}}{a_{ik}-a_{ik-1}}$ for $k = j+1, \ldots, j'$. Therefore, it follows from (14) that $0 \le \frac{s_{ij''}-s_{ij'}}{a_{ij''}-a_{ij'}} \le \frac{s_{ij'}-s_{ij}}{a_{ij'}-a_{ij}} \le 1$. $\square$

A.5 Proof of Lemma 3

Proof We first show that $\mathrm{proj}_{s_i}(PQ_i) = Q_i$. Clearly, $\mathrm{proj}_{s_i}(PQ_i) \subseteq Q_i$ since for every $(u_i, s_i) \in PQ_i$ we have $s_i \in Q_i$. To show $Q_i \subseteq \mathrm{proj}_{s_i}(PQ_i)$, we consider a point $s_i \in Q_i$ and observe $(s_i, s_i) \in PQ_i$. Second, we argue that $\mathrm{proj}_{u_i}(PQ_i) = P_i$. To show $P_i \subseteq \mathrm{proj}_{u_i}(PQ_i)$, let $u_i \in P_i$ and define $s_{ij} = \min\{a_{ij}, u_{in}\}$ for all j. We will show $(u_i, s_i) \in PQ_i$. It follows readily that, for $j \in \{0, n\}$, $u_{ij} = s_{ij}$ and $u_i \le s_i$. In addition, observe that there exists a j' such that $a_{ij'-1} \le u_{in} \le a_{ij'}$. By direct computation, $s_i = \lambda v_{ij'-1} + (1-\lambda)v_{ij'}$, where $\lambda = \frac{a_{ij'}-u_{in}}{a_{ij'}-a_{ij'-1}}$. In other words, $s_i \in Q_i$. To prove $\mathrm{proj}_{u_i}(PQ_i) \subseteq P_i$, we consider a point $(u_i, s_i) \in PQ_i$ and show $u_i \in P_i$. Clearly, $u_{i0} = a_{i0}$. Also, for $j = 1, \ldots, n$, $a_{i0} \le u_{ij} \le s_{ij} \le \min\{u_{in}, a_{ij}\}$, where the first two inequalities hold by the definition of PQ_i and the last inequality holds because $u_{in} = s_{in}$ and the inequality, $s_{ij} \le \min\{a_{ij}, s_{in}\}$, is satisfied by all extreme points of Q_i. The last two statements follow similarly because Φ^P, Φ^Q, Φ^{PQ} are obtained from P, Q, and PQ respectively, by adding a coordinate ϕ which depends only on coordinates shared by P, Q, and PQ, namely $(u_{1n}, \ldots, u_{dn}) = (s_{1n}, \ldots, s_{dn})$. $\square$

References

1. Al-Khayyal, F.A., Falk, J.E.: Jointly constrained biconvex programming. Math. Oper. Res. **8**(2), 273–286 (1983)
2. Anstreicher, K.M., Burer, S.: Computable representations for convex hulls of low-dimensional quadratic forms. Math. Program. **124**(1–2), 33–43 (2010)
3. Bao, X., Khajavirad, A., Sahinidis, N.V., Tawarmalani, M.: Global optimization of nonconvex problems with multilinear intermediates. Math. Program. Comput. **7**(1), 1–37 (2015)

4. Belotti, P., Lee, J., Liberti, L., Margot, F., Wächter, A.: Branching and bounds tightening techniques for non-convex MINLP. Optim. Methods Softw. **24**(4–5), 597–634 (2009)
5. Benson, H.P.: Concave envelopes of monomial functions over rectangles. Naval Res. Logist. **51**(4), 467–476 (2004)
6. Bienstock, D., Michalka, A.: Cutting-planes for optimization of convex functions over nonconvex sets. SIAM J. Optim. **24**(2), 643–677 (2014)
7. Boland, N., Dey, S.S., Kalinowski, T., Molinaro, M., Rigterink, F.: Bounding the gap between the McCormick relaxation and the convex hull for bilinear functions. Math. Program. **162**(1–2), 523–535 (2017)
8. Burer, S., Kılınç-Karzan, F.: How to convexify the intersection of a second order cone and a nonconvex quadratic. Math. Program. **162**(1–2), 393–429 (2017)
9. Cafieri, S., Lee, J., Liberti, L.: On convex relaxations of quadrilinear terms. J. Global Optim. **47**(4), 661–685 (2010)
10. Chan, T.M.: Optimal output-sensitive convex hull algorithms in two and three dimensions. Discrete Comput. Geom. **16**(4), 361–368 (1996)
11. CMU-IBM Cyber-Infrastructure for MINLP collaborative site (2019). http://www.minlp.org. Accessed 10 Oct 2019
12. Crama, Y., Rodríguez-Heck, E.: A class of valid inequalities for multilinear 0–1 optimization problems. Discrete Optim. **25**, 28–47 (2017)
13. Del Pia, A., Khajavirad, A.: A polyhedral study of binary polynomial programs. Math. Oper. Res. **42**(2), 389–410 (2016)
14. Del Pia, A., Khajavirad, A.: The multilinear polytope for acyclic hypergraphs. SIAM J. Optim. **28**(2), 1049–1076 (2018)
15. Del Pia, A., Khajavirad, A., Sahinidis, N.V.: On the impact of running intersection inequalities for globally solving polynomial optimization problems. Math. Program. Comput. **12**(2), 165–191 (2020)
16. Gleixner, A., Bastubbe, M., Eifler, L., Gally, T., Gamrath, G., Gottwald, R.L., Hendel, G., Hojny, C., Koch, T., Lübbecke, M.E., Maher, S.J., Miltenberger, M., Müller, B., Pfetsch, M.E., Puchert, C., Rehfeldt, D., Schlösser, F., Schubert, C., Serrano, F., Shinano, Y., Viernickel, J.M., Walter, M., Wegscheider, F., Witt, J.T., Witzig, J.: The SCIP Optimization Suite 6.0. Technical report, Optimization Online (2018)
17. Gleixner, A.M., Berthold, T., Müller, B., Weltge, S.: Three enhancements for optimization-based bound tightening. J. Global Optim. **67**, 731–757 (2017)
18. Graham, R.L.: An efficient algorithm for determining the convex hull of a finite planar set. Inf. Process. Lett. **1**(4), 132–133 (1972)
19. Grötschel, M., Lovász, L., Schrijver, A.: Geometric Algorithms and Combinatorial Optimization, vol. 2. Springer, Berlin (2012)
20. Günlük, O., Linderoth, J.: Perspective reformulations of mixed integer nonlinear programs with indicator variables. Math. Program. **124**(1–2), 183–205 (2010)
21. Gupte, A., Kalinowski, T., Rigterink, F., Waterer, H.: Extended formulations for convex hulls of some bilinear functions. Discrete Optim. **36**, 100569 (2020)
22. He, T., Tawarmalani, M.: Tractable relaxations of composite functions. Working paper (2018)
23. Hiriart-Urruty, J.B., Lemaréchal, C.: Fundamentals of Convex Analysis. Springer, Berlin (2012)
24. Horst, R., Tuy, H.: Global Optimization—Deterministic Approaches. Springer, Berlin (1996)
25. Jach, M., Michaels, D., Weismantel, R.: The convex envelope of (n-1)-convex functions. SIAM J. Optim. **19**(3), 1451–1466 (2008)
26. Khajavirad, A., Sahinidis, N.V.: Convex envelopes of products of convex and component-wise concave functions. J. Global Optim. **52**(3), 391–409 (2012)
27. Khajavirad, A., Sahinidis, N.V.: Convex envelopes generated from finitely many compact convex sets. Math. Program. **137**(1–2), 371–408 (2013)
28. Kılınç-Karzan, F., Yıldız, S.: Two-term disjunctions on the second-order cone. Math. Program. **154**(1–2), 463–491 (2015)
29. Luedtke, J., Namazifar, M., Linderoth, J.: Some results on the strength of relaxations of multilinear functions. Math. Program. **136**(2), 325–351 (2012)
30. McCormick, G.P.: Computability of global solutions to factorable nonconvex programs: part i Convex underestimating problems. Math. Program. **10**(1), 147–175 (1976)
31. Meyer, C.A., Floudas, C.A.: Trilinear monomials with mixed sign domains: facets of the convex and concave envelopes. J. Global Optim. **29**(2), 125–155 (2004)

32. Misener, R., Floudas, C.A.: ANTIGONE: Algorithms for continuous/integer global optimization of nonlinear equations. J. Global Optim. **59**(2–3), 503–526 (2014)
33. Misener, R., Smadbeck, J.B., Floudas, C.A.: Dynamically generated cutting planes for mixed-integer quadratically constrained quadratic programs and their incorporation into GloMIQO 2. Optim. Methods Softw. **30**(1), 215–249 (2015)
34. Modaresi, S., Vielma, J.P.: Convex hull of two quadratic or a conic quadratic and a quadratic inequality. Math. Program. **164**(1–2), 383–409 (2017)
35. Muller, B., Serrano, F., Gleixner, A.: Using two-dimensional projections for stronger separation and propagation of bilinear terms. SIAM J. Optim. **30**(2), 1339–1365 (2020)
36. Najman, J., Mitsos, A.: Tighter McCormick relaxations through subgradient propagation. J. Global Optim. **75**(3), 565–593 (2019)
37. Nguyen, T.T., Richard, J.P.P., Tawarmalani, M.: Deriving convex hulls through lifting and projection. Math. Program. **169**(2), 377–415 (2018)
38. Padberg, M.: The boolean quadric polytope: some characteristics, facets and relatives. Math. Program. **45**(1), 139–172 (1989)
39. Rikun, A.D.: A convex envelope formula for multilinear functions. J. Global Optim. **10**(4), 425–437 (1997)
40. Sherali, H.D.: Convex envelopes of multilinear functions over a unit hypercube and over special discrete sets. Acta mathematica vietnamica **22**(1), 245–270 (1997)
41. Sherali, H.D., Adams, W.P.: A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems, vol. 31. Springer, Berlin (2013)
42. Speakman, E., Lee, J.: Quantifying double McCormick. Math Oper Res **42**(4), 1230–1253 (2017)
43. Tawarmalani, M.: Inclusion certificates and simultaneous convexification of functions. Working paper (2010)
44. Tawarmalani, M., Richard, J.P.P., Chung, K.: Strong valid inequalities for orthogonal disjunctions and bilinear covering sets. Math. Program. **124**(1–2), 481–512 (2010)
45. Tawarmalani, M., Richard, J.P.P., Xiong, C.: Explicit convex and concave envelopes through polyhedral subdivisions. Math. Program. **138**(1–2), 531–577 (2013)
46. Tawarmalani, M., Sahinidis, N.V.: Semidefinite relaxations of fractional programs via novel convexification techniques. J. Global Optim. **20**(2), 133–154 (2001)
47. Tawarmalani, M., Sahinidis, N.V.: Convex extensions and envelopes of lower semi-continuous functions. Math. Program. **93**(2), 247–263 (2002)
48. Tawarmalani, M., Sahinidis, N.V.: Global optimization of mixed-integer nonlinear programs: a theoretical and computational study. Math. Program. **99**(3), 563–591 (2004)
49. Tawarmalani, M., Sahinidis, N.V.: A polyhedral branch-and-cut approach to global optimization. Math. Program. **103**(2), 225–249 (2005)
50. Vigerske, S.: MINLPLIB 2. In: Proceedings of the XII Global Optimization Workshop MAGO 2014, pp. 137–140 (2014)