Submitted to *Mathematics of Operations Research*
manuscript (Please, provide the manuccript number!)

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

Tractable relaxations of composite functions

Taotao He
Antai College of Economics and Management, Shanghai Jiao Tong University, hetaotao@sjtu.edu.cn

Mohit Tawarmalani
Krannert School of Management, Purdue University, mtawarma@purdue.edu, http://web.ics.purdue.edu/ mtawarma/

In this paper, we introduce new relaxations for the hypograph of composite functions assuming that the outer-function is supermodular and concave-extendable. Relying on the recently introduced relaxation framework of [17], we devise a separation algorithm for the graph of the outer-function over P, where P is a special polytope introduced in [17] to capture the structure of each inner-function using its finitely many bounded estimators. The separation algorithm takes $\mathcal{O}(dn \log d)$ time, where d is the number of inner-functions and n is the number of estimators for each inner-function. Consequently, we derive large classes of inequalities that tighten prevalent factorable programming relaxations. We also generalize a decomposition result of [26, 6] and devise techniques to simultaneously separate hypographs of various supermodular, concave-extendable functions using facet-defining inequalities. Assuming that the outer-function is convex in each argument, we characterize the limiting relaxation obtained with infinitely many estimators as the solution of an optimal transport problem. When the outer-function is also supermodular, we obtain an explicit integral formula for this relaxation.

Key words: Mixed-integer nonlinear programs; Factorable programming; Supermodularity; Staircase triangulation; Convexification via optimal transport
MSC2000 subject classification: Primary: 90C11, 90C26, 90C30; secondary: 60E15
OR/MS subject classification: Primary: programming: integer: nonlinear; secondary: programming: nonlinear: theory; programming: integer: cutting plane-facet generation
History:

---

1. Introduction Mixed-integer nonlinear programs are typically solved using branch-and-bound (B&B); an algorithm that, in the branching step, refines a partition of the variable domains and then, in the bounding step, chooses one partition element to construct a relaxation for the problem. To guarantee that B&B converges, relaxations are constructed so that they approach the original problem as the partition element shrinks. Such relaxations can be constructed using the factorable programming (FP) technique, which is currently used by most state-of-the-art solvers [41, 3, 25, 44]. This technique recursively traverses the expression tree for each nonlinear function and relaxes each operator over a bounding box that covers the ranges for all the operands. More specifically, FP treats each function as a recursive sum and/or product of univariate functions. The technique then relaxes bilinear terms over variable bounds using McCormick envelopes [23] and relaxes each univariate function using its function-specific structure over the range of the independent variable. It is widely accepted that B&B convergence improves if, for any given partition size, a tighter relaxation can be constructed. Although the FP relaxation can be tightened using relaxation hierarchies [36, 2, 28, 20], doing so increases the relaxation size considerably because hierarchies introduce new products and variables to linearize the products.

FP relaxations can be weak because they ignore operand structure while constructing the relaxation for each operator. Instead, [17] recently proposed a new relaxation, referred to hereafter as the composite relaxation, that is tighter than the FP relaxation and uses the same set of variables. The composite relaxation is constructed by relaxing the graph of the outer-function, referred to as the operator in FP parlance, over a polytope P. Instead of just using the bounds for the inner-functions/operands, this polytope allows the relaxation constructor to utilize much more of their structure. This structure is derived from estimators of the inner-functions and their bounds. The main contribution of [17] is that using an oracle that separates the graph of the outer-function over a specific well-structured subset of P, which we refer to as Q, the paper constructs a fast combinatorial algorithm to separate the graph over P. Unfortunately, without any further structure, even maximizing a general outer-function over Q is NP-Hard. Therefore, to efficiently utilize this algorithm, there is a need to identify tractable separation problems over Q.

Here, we identify a class of such instances. In particular, we shall explicitly convexify the outer-function over P and Q, assuming, of course, that the outer-function has some structure. Specifically, we will assume that the function is supermodular and that its concave envelope is determined by its value at the extreme points of Q, a property referred to as its concave-extendability over Q. The polytope Q generalizes the hypercube, a set that arises as the partition element in rectangular B&B and is used as the function domain in many convexification studies [1, 27, 30, 35, 40, 4, 30, 24, 7, 38, 39, 6, 10, 15]. Our use of Q provides a concrete illustration that domains, besides the hypercube, can be used to derive general-purpose cuts for MINLPs.

The sole example of composite relaxations in the literature [17] for which explicit inequalities are available concerns the product of two bounded functions each furnished with an underestimator. We now describe ways in which we extend these results beyond the above example setting. First, as mentioned, we treat a larger class of outer-functions, in particular, those that are supermodular and concave-extendable over Q. Second, we allow arbitrarily many estimators for each function. Formally, for a composite function with d inner-functions, each equipped with n estimators, we devise an algorithm that generates, whenever possible, a facet-defining inequality in $\mathcal{O}(dn \log d)$ time to separate a given point from the hypograph of the outer-function over Q. The number of facet-defining inequalities of this hypograph is $\binom{dn}{n,n,\dots,n}$, which, by Stirling's approximation, grows asymptotically as fast as $\frac{d^{dn+\frac{1}{2}}}{(2n\pi)^{\frac{d-1}{2}}}$, or exponentially with respect to d and n. Though numerous, since these inequalities are generated using a fast combinatorial separation algorithm, they can be derived iteratively, with little computational overhead, to cut off infeasible regions from MINLP relaxations. The geometric structure of these inequalities relates to a certain triangulation of Q, which provides many insights. We show that results in [26] and [6] regarding separability of concave envelopes can be generalized to our setting. We also show that even when additional estimators are available, any inequalities generated using fewer estimators are still facet-defining. We specialize our results to the example setting of [17], proving that the inequalities therein yield the convex hull of the graph of the bilinear product over P; the polytope which was modeled using one estimator for each inner function. Third, we extend our algorithm to allow simultaneous separation of a vector of composite functions, each with an outer-function that is supermodular and concave-extendable over Q. Fourth, we consider infinitely many estimators for each inner-function, assuming additionally that the outer-function is convex when all but one of its arguments are fixed. We show that, in this case, the composite relaxation arises as the solution of an optimal transport problem [45]. The reduction proceeds by expressing each inner-function as the expectation of a random variable that is completely determined by its estimators. When the outer-function is also supermodular, we provide an explicit integral formula, whose evaluation gives a closed-form expression for the composite relaxation.

Notation: Throughout this paper, we shall denote the convex hull of set S by $\mathrm{conv}(S)$, the projection of a set 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 $\mathrm{ri}(S)$. For a function $f:\mathbb{R}^n\to\mathbb{R}$, we denote its convex (resp. concave) envelope over a set S by $\mathrm{conv}_S(f)$ (resp. $\mathrm{conc}_S(f)$), and its graph by $\mathrm{gr}(f)$. 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. More specifically, when we write $f(\cdot)$ we refer to the the function and when we write $f(x)$ we refer to the value of $f(x)$ at x. The vector e_i will denote the i^{th} standard basis vector in $\mathbb{R}^N$, where we will not specify N when it is apparent from the context.

2. Problem setup and geometric structure

Let $\phi\circ f\colon X\subseteq\mathbb{R}^m\to\mathbb{R}$ be a composite function denoted as $(\phi\circ f)(x)=\phi\big(f(x)\big)$, where $f\colon X\to\mathbb{R}^d$ is a vector of functions defined as $f(x):=\big(f_1(x),\dots,f_d(x)\big)$ for $x\in X$, and $\phi:\mathbb{R}^d\to\mathbb{R}$ is a continuous function. We shall refer to $f(\cdot)$ as inner-functions and $\phi(\cdot)$ as the outer-function. Throughout this paper, we assume that for $i\in\{1,\dots,d\}$ the inner-function $f(\cdot)$ is *bounded*, that is, for every $x\in X$, $f_i^L\le f_i(x)\le f_i^U$. It turns out that we can assume, without loss of generality, that, for every $x\in X$, $f(x)\in[0,1]^d$ and $\phi:[0,1]^d\to\mathbb{R}$ since otherwise we can define $\tilde f_i(x)=(f_i(x)-f_i^L)/(f_i^U-f_i^L)$ as the i^{th} inner-function and $\tilde\phi(f)=\phi\big((f_1^U-f_1^L)f_1+f_1^L,\dots,(f_1^U-f_d^L)f_d+f_d^L\big)$ as the outer-function. In [17], the authors proposed a framework to relax the graph of $\phi\circ f$, that is, $\mathrm{gr}(\phi\circ f)=\big\{(x,\phi)\bigm|\phi=\phi\big(f(x)\big),\ x\in X\big\}$. In this section, we review these ideas and relate our setting to some relevant convexification results.

2.1. A relaxation framework for composite functions

Let $n\in\mathbb{Z}$. Let $u\colon\mathbb{R}^m\to\mathbb{R}^{d\times(n+1)}$ be a vector of bounded functions defined as $u(x)=\big(u_1(x),\dots,u_d(x)\big)$ for $x\in\mathbb{R}^m$ and let $a=(a_1,\dots,a_d)$ be a vector in $\mathbb{R}^{d\times(n+1)}$. For all i and $x\in X$, assume that $f_i(x)\in[a_{i0},a_{in}]$ and that u and a satisfy the following inequalities:

$$0\le a_{i0}<\dots<a_{in}\le 1,\quad u_{ij}(x)\le\min\big\{f_i(x),a_{ij}\big\},\quad u_{i0}(x)=a_{i0},\quad u_{in}(x)=f_i(x).\tag{1}$$

In other words, each $u_{ij}(\cdot)$ underestimates the corresponding inner-function $f_i(\cdot)$ and is bounded from above by a constant a_{ij}. In particular, $u_{in}(\cdot)$ (resp. $u_{i0}(\cdot)$) is a special underestimator that equals $f_i(\cdot)$ (resp. a_{i0}). In [17], the polytope $P:=\prod_{i=1}^d P_i$, where

$$P_i=\big\{u_i\in\mathbb{R}^{n+1}\bigm|u_{ij}\le u_{in}\text{ and }a_{i0}\le u_{ij}\le a_{ij}\text{ for all }j\in\{0,\dots,n\}\big\},\tag{2}$$

was introduced as an *abstraction* of underestimators for the inner-functions $f(\cdot)$. We review some basic ideas regarding the structure of P. First, the polytope P_i introduces a variable u_{ij} for each underestimator $u_{ij}(\cdot)$. Thus, $u_{ij}\le u_{in}$ (resp. $u_{ij}\le a_{ij}$) models that $u_{ij}(\cdot)$ underestimates $f_i(\cdot)$ (resp. a_{ij}). Since a_{i0} is a lower bound for $f_i(\cdot)$ on X, Proposition 1 in [17] shows that we are allowed to impose the constraint $a_{i0}\le u_{ij}$. Second, our assumption that each inner function $f_i(\cdot)$ has n underestimators is without loss of generality. Third, if some of the estimators are overestimators, Proposition 2 in [17] gives an affine transformation to reduce the treatment to one involving only underestimators. The following result shows how relaxations for composite functions can be constructed by relaxing the outer-function over P. To this end, we extend the outer-function $\phi:\mathbb{R}^d\to\mathbb{R}$ to define $\bar\phi:\mathbb{R}^{d\times(n+1)}\to\mathbb{R}$ as a function so that

$$\bar\phi(u_1,\dots,u_d)=\phi(u_{1n},\dots,u_{dn})\qquad\text{for each }(u_1,\dots,u_d)\in P.$$

For a subset D of $\mathbb{R}^{d\times n}$, we denote by $\mathrm{conc}_D(\bar\phi)(\cdot)$ the concave envelope of $\bar\phi(\cdot)$ over D. Observe that although the function $\phi(\cdot)$ depends only on variables $u_{\cdot n}:=(u_{1n},\dots,u_{dn})$, the envelope $\mathrm{conc}_P(\bar\phi)(\cdot)$ depends on all the variables $u:=(u_1,\dots,u_n)$.

THEOREM 1 **(Theorem 2 in [17])**. *Let $\phi:\mathbb{R}^d \to \mathbb{R}$ be a continuous function and let $f:\mathbb{R}^m \to \mathbb{R}^d$ be a vector of functions, each of which is bounded over $X \subseteq \mathbb{R}^m$. If $\big(a, u(\cdot)\big)$ satisfies (1) then* $\operatorname{hyp}(\phi \circ f) \subseteq \operatorname{proj}_{(x,\phi)}(R)$, *where* $\operatorname{hyp}(\phi\circ f)$ *is the hypograph of $\phi \circ f$ over X and*

$$R := \Big\{(x, u_{\cdot n}, \phi) \;\Big|\; \phi \le \operatorname{conc}_P(\bar\phi)(u),\ u(x) \le u,\ u_{\cdot n} = f(x),\ x \in X\Big\}.$$

If the graph of $f(\cdot)$, expressed using the constraints $u_{\cdot n} = f(x)$ and $x \in X$ in the definition of R, is outer-approximated with a convex set, and, for $j \neq n$, $u_{ij}(\cdot)$ is convex, we obtain a convex relaxation of $\operatorname{hyp}(\phi\circ f)$. *Moreover,* $\operatorname{conc}_P(\bar\phi)(\cdot)$ *is non-increasing in u_{ij}, for all $i \in \{1,\dots,d\}$ and $j \notin \{0,n\}$. Substituting $u_{ij}(\cdot)$ for such u_{ij} variables and dropping $u_{ij} \ge u_{ij}(\cdot)$ projects the u_{ij} variables out of the convex relaxation.* □

The idea behind Theorem 1 is that the constraints in P are satisfied by underestimators and, as such, inequalities valid for the hypograph of $\operatorname{conc}_P(\bar\phi)$ are also valid for the underestimators. Since variables u_{ij}, for $j \neq n$, are eventually replaced with their defining function, $u_{ij}(\cdot)$, the relaxation, just like the factorable one, uses the original variables x and an introduced variable u_{in} for each inner function $f_i(\cdot)$.

In this paper, we will solve the *facet-generation problem* of $\operatorname{conc}_P(\bar\phi)$, assuming that $\bar\phi(\cdot)$ is supermodular and concave-extendable over P. Under these conditions, the hypograph of $\operatorname{conc}_P(\bar\phi)$ is a polyhedron. By the facet-generation problem of a full-dimensional polyhedron S, we mean that, given a vector y, we establish that either $y \in S$ or find a facet-defining inequality of S that is not satisfied by y. The facet-generation problem for $\operatorname{conc}_P(\bar\phi)$ is, in general, NP-Hard because P includes, as a special case, the unit hypercube and $\phi(\cdot)$ can be any bilinear function. Nonetheless, on the positive side, [17] showed that the facet-generation problem for $\operatorname{conc}_P(\bar\phi)$ is tractable if the facet-generation problem for $\operatorname{conc}_Q(\bar\phi)$ and some of its faces is tractable. The polytope Q, which is the domain of the latter function $\operatorname{conc}_Q(\bar\phi)$, is a subset of P that we will describe shortly. Simultaneously, we will also review other results relevant to devising separation algorithms for $\operatorname{conc}_P(\bar\phi)$.

Let $a = (a_1,\dots,a_d)$ be a vector in $\mathbb{R}^{d\times(n+1)}$ so that each subvector a_i is strictly increasing. Then, $Q := \prod_{i=1}^d Q_i$, where Q_i is the simplex in $\mathbb{R}^{n+1}$ with extreme points $\{v_{ij}\}_{j=0}^n$ given as follows:

$$v_{ij} := (a_{i0},\dots,a_{ij},a_{ij},\dots,a_{ij}) \qquad \text{for } j = 0,\dots,n. \tag{3}$$

Given a point $(\bar u, \bar\phi) \in \mathbb{R}^{d\times(n+1)+1}$, the separation algorithm for $\operatorname{conc}_P(\bar\phi)$ constructs another point $(\bar s, \bar\phi) \in \mathbb{R}^{d\times(n+1)+1}$. This point is then separated from $\operatorname{conc}_Q(\bar\phi)$ using the separation oracle for one of its faces. We first describe how $(\bar s, \bar\phi)$ is obtained from $(\bar u, \bar\phi)$. With each point $w \in P_i$, we associate a discrete univariate function $\xi_i(a; u_i)\colon [a_{i0}, a_{in}] \to \mathbb{R}$ defined as follows:

$$\xi_i(a;w) = \begin{cases} w_j & a = a_{ij} \quad \text{for } j \in \{0,\dots,n\} \\ -\infty & \text{otherwise.} \end{cases}$$

Then, $\bar s = (\bar s_1,\dots,\bar s_d)$, where $\bar s_i = \big(\operatorname{conc}(\xi_i)(a_{i0};\bar u_i),\dots,\operatorname{conc}(\xi_i)(a_{in};\bar u_i)\big)$ and $\operatorname{conc}(\xi_i)(\cdot;u_i)$ is the concave envelope of $\xi_i(\cdot;u_i)$ over $[a_{i0}, a_{in}]$. See Figure 1, where we illustrate the discrete univariate function $\xi_i(a;\bar u_i)$ and the corresponding $\bar s_i$ derived using the concave envelope construction. So, $\bar u_i$ (resp. $\bar s_i$) is the vector of values of $\xi_i(\chi;u_i)$ (resp. $\operatorname{conc}(\xi_i)(\chi;u_i)$) generated by sequentially setting χ to the values in $(a_{i0},\dots,a_{in})$. For each $\bar u \in P$, the corresponding $\bar s$ is captured in the set:

$$PQ' := \Big\{(u,s) \;\Big|\; (u_1,\dots,u_d) \in P,\ s_i := \big(\operatorname{conc}(\xi_i)(a_{i0};u_i),\dots,\operatorname{conc}(\xi_i)(a_{in};u_i)\big),\ i = 1,\dots,d\Big\}. \tag{4}$$

Since $\bar u$ and $\bar s$ are related via a concave envelope construction, we can lift $\bar u$ to its unique lifting $(\bar u, \bar s) \in PQ'$ using a two-dimensional convex hull algorithm, such as Graham scan [14], that, for

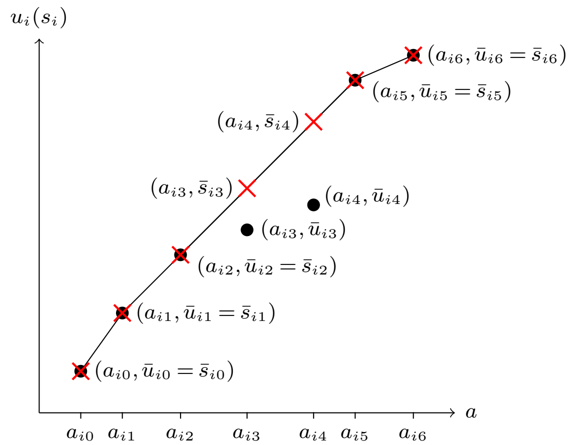


FIGURE 1. Illustration of the map Γ_J.The black dots are the points $(a_{ij}, \bar{u}_{ij})$, and the red crosses are $(a_{ij}, \bar{s}_{ij})$, where $\bar{s}_i = (\Gamma_J)_i(\bar{u}_i)$. In this case, $J_i = \{0,1,2,5,6\}$. Here, $\bar{s}_i \in \operatorname{conv}(v_{i0}, v_{i1}, v_{i2}, v_{i5}, v_{i6})$, where v_{ij} is as defined in (3).

each of the d discrete univariate functions $\xi_i(a; u_i)$ finds its envelope in $\mathcal{O}(n)$ time. Given $\bar{u} \in P$, we will find a face of Q containing $\bar{s}$ and, consequently, identify which face of $\operatorname{conc}_Q(\bar{\phi})$ is to be separated from $(\bar{s}, \bar{\phi})$. Observe that for any $\bar{u}_i \notin Q_i$, $\bar{u}_i$ violates certain facet-defining inequalities of Q_i. As the next result shows, these inequalities define facets, whose intersection yields a face of Q_i containing $\bar{s}_i$. Together, for all i, these faces define the face of interest of Q, which is described using d-tuples of index sets. Consider a collection of d-tuples

$$\mathcal{J} := \big\{ (J_1, \dots, J_d) \bigm| \{0, n\} \subseteq J_i \subseteq \{0, 1, \dots, n\} \; \forall i \in \{1, \dots, d\} \big\}. \tag{5}$$

Then, we associate with a d-tuple, $J = (J_1, \dots, J_d) \in \mathcal{J}$, the face $F_J := F_{1J_1} \times \cdots \times F_{dJ_d}$ of Q, where

$$F_{iJ_i} := \operatorname{conv}\big(\{v_{ij} \mid j \in J_i\}\big),$$

and v_{ij} is defined as in (3). Clearly, F_{iJ_i} is a face of Q_i because Q_i is a simplex whose vertices form a superset of those of F_{iJ_i}. With J, we also associate a linear map, $\Gamma_J \colon \mathbb{R}^{d\times(n+1)} \to \mathbb{R}^{d\times(n+1)}$ that maps a $u \in P$ to $\tilde{u} \in P$ as follows:

$$\tilde{u}_{ij} = u_{ij} \quad \text{for } j \in J_i \qquad \text{and} \qquad \tilde{u}_{ij} = (1-\gamma_{ij}) u_{il(i,j)} + \gamma_{ij} u_{ir(i,j)} \quad \text{for } j \notin J_i, \tag{6}$$

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, for $j \notin J_i$, $\gamma_{ij} = (a_{ij} - a_{il(i,j)})/(a_{ir(i,j)} - a_{il(i,j)})$. In other words, $\tilde{u}_{ij}$ is obtained by restricting the domain of $\xi_i(a; u_i)$, a function of a, to $\{a_{ij}\}_{j \in J_i}$ and then linearly interpolating the function at a_{ij} for $j \notin J_i$ (see Figure 1 for an illustration of the map Γ_J, where we have chosen J_i as the set of indices for which $\bar{u}_{ij} = \bar{s}_{ij}$, consistent with the assumption in Proposition 1 below).

PROPOSITION 1 **(Proposition 5 in [17]).** *Let $(\bar{u}, \bar{s}) \in PQ'$ and $J = (J_1, \dots, J_d)$ be defined so that $J_i := \{j \mid \bar{u}_{ij} = \bar{s}_{ij}\}$. Then, $\bar{s} = \Gamma_J(\bar{u})$. The set Q satisfies the inequalities $s \ge \Gamma_J(s)$. Then F_J is the face of Q defined by the inequalities $s \le \Gamma_J(s)$. Moreover, $\bar{s} \in F_J$.* $\square$

To separate $(\bar{s}, \bar{\phi})$ from $\operatorname{conc}_Q(\bar{\phi})$, we introduce projections of Q defined using index sets (5). As a succinct notation, for any $y \in \mathbb{R}^{d\times(n+1)}$ and for $J \in \mathcal{J}$, we let $y_J := (y_{1J_1}, \dots, y_{dJ_d})$, where y_{iJ_i} consists of coordinates of y_i from the index-set J_i, and let $\bar{J} := (\bar{J}_1, \dots, \bar{J}_d)$, where $\bar{J}_i$ is the complement of J_i, *i.e.*, $\bar{J}_i = \{0, \dots, n\} \setminus J_i$. Using these definitions, we can then write, up to reordering of variables, that $y = (y_J, y_{\bar{J}})$. Now, define

$$Q_J := Q_{1J_1} \times \cdots \times Q_{dJ_d}, \tag{7}$$

where Q_{iJ_i} is the simplex whose extreme points are defined in (3) with the parameter vector $a_{iJ_i} \in \mathbb{R}^{|J_i|}$ and observe that Q_J is a projection of Q to the coordinates contained in the d-tuple J. Given a point $(\bar{s}, \bar{\phi})$ and $J = (J_1, \ldots, J_d)$ defined so that $J_i := \{j \mid \bar{u}_{ij} = \bar{s}_{ij}\}$, we call the facet-generation oracle of $\operatorname{conc}_{Q_J}(\phi)(s_J)$ to generate an inequality $\phi \le \langle \alpha_J, s_J \rangle + b$ of $\operatorname{conc}_{Q_J}(\phi)(s_J)$, which is tight at $\bar{s}_J$, *i.e.*, $\operatorname{conc}_{Q_J}(\phi)(\bar{s}_J) = \langle \alpha_J, \bar{s}_J \rangle + b$. Let $\tilde{\alpha}$ be a vector in $\mathbb{R}^{d\times(n+1)}$ such that $\tilde{\alpha}_J = \alpha_J$ and $\tilde{\alpha}_{\bar{J}} = 0$. By Corollary 2 in [17], the inequality $\phi \le \big\langle (\alpha_J, 0), (s_J, s_{\bar{J}}) \big\rangle + b$ defines a facet of $\operatorname{conc}_P(\bar{\phi})$ that is tight at $\bar{u}$. We summarize the above discussion for later use.

Proposition 2 **(Corollary 2 in [17])**. *Assume that* $\operatorname{conc}_Q(\bar{\phi})(s)$ *is a polyhedral function. Given* $\bar{u} \in P$*, the unique point* $(\bar{u}, \bar{s}) \in PQ'$ *can be found in* $\mathcal{O}(dn)$ *time. Let* $J = (J_1, \ldots, J_d)$*, where* $J_i := \{j \mid \bar{u}_{ij} = \bar{s}_{ij}\}$*. Then,* $\bar{s} \in F_J$ *and* $\bar{s}_J \in Q_J$*. If* $\phi \le \langle \alpha_J, s_J \rangle + b$ *is a facet-defining inequality for* $\operatorname{conc}_{Q_J}(\phi)$ *that is tight at* $\bar{s}_J$ *then the inequality* $\phi \le \big\langle (\alpha_J, 0), (u_J, u_{\bar{J}}) \big\rangle + b$ *defines a facet of* $\operatorname{conc}_P(\bar{\phi})$ *that is tight at* $\bar{u}$*.* $\square$

2.2. Supermodularity and staircase triangulation This paper considers supermodular and concave-extendable functions over a Cartesian product of simplices, Q. Such functions have a concave envelope that is closely related to certain triangulations of Q. In this subsection, we explore these connections. Before we begin, we formally define concave-extendability of functions [40] and triangulations of polyhedral domains [9]; both are prevalent notions in convexification literature.

Definition 1 ([39]). A function $g\colon D \to \mathbb{R}$, where D is a polytope, is said to be concave-extendable from $X \subseteq D$ if the concave envelope of $g(x)$ is determined by X only, that is, the concave envelope of g and $g|_X$ over P are identical, where $g|_X$ is the restriction of g to X that is defined as:

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

A function $g : D \to \mathbb{R}$ is convex extendable from $X \subseteq D$ if $-g$ is concave extendable from X. $\square$

Definition 2 (Triangulation [9]). Let $D \subseteq \mathbb{R}^n$. A set of polyhedra $\mathcal{R} := \{R_1, \ldots, R_r\}$ forms a *polyhedral subdivision* of D if $D = \bigcup_{i=1}^{r} R_i$ and $R_i \cap R_j$ is a (possibly empty) face of both R_i and R_j. Moreover, if each R_i is a simplex, then $\mathcal{R}$ is a *triangulation* of D. $\square$

The non-vertical facets of a polyhedral function, when projected, divide the domain into polyhedral sets, which form a polyhedral subdivision; a subdivision that can be further refined into a triangulation. Thus, if the concave envelope $\operatorname{conc}_Q(\bar{\phi})(\cdot)$ is polyhedral over Q, there is a triangulation $\mathcal{R}$ of Q such that the concave envelope affinely interpolates each simplex R_i of this triangulation (Theorem 2.4 in [39]). By affine interpolation, we mean that the function value at any point $s \in R_i$ is obtained as the affine combination of function values at $\operatorname{vert}(R_i)$. Therefore, $\operatorname{conc}_Q(\bar{\phi})(\cdot)$, when polyhedral, is uniquely described by the triangulation $\mathcal{R}$ of Q.

We will eventually be interested in extending the domain of the concave envelope outside of Q. To do so, we will use the following construction, which extends the domain of $\operatorname{conc}_Q(\bar{\phi})$ to $\operatorname{aff}(Q)$, a set that contains P. We describe this construction for a generic function $\chi : D \to \mathbb{R}$, whose domain, D, is a subset of $\mathbb{R}^n$ and is assumed to be endowed with a triangulation $\mathcal{R}$. Define $\chi^{R_i}(x) : \operatorname{aff}(D) \to \mathbb{R}$ as the unique affine function that satisfies $\chi^{R_i}(x) = \chi(x)$ for all $x \in \operatorname{vert}(R_i)$. Moreover, define $h(x)$ so that, for all $x \in R_i \in \mathcal{R}$, $h(x) = \chi^{R_i}(x)$ and assume that it is concave. Now, to extend $h(x)$ to $\operatorname{aff}(D)$, we consider another function $\chi^{\mathcal{R}}(x)$ defined as $\min_i \chi^{R_i}(x)$ and show that it matches $h(x)$ over D. If not, there exists some $(i, j) \in \{1, \ldots, r\}^2$ and an x such that although $x \in R_i$, $\chi^{R_i}(x) > \chi^{R_j}(x)$. Now, pick $y \in \operatorname{int}(R_j)$ and a sufficiently small $\epsilon > 0$ so that $y + \epsilon(x - y) \in R_j$. Then, as the following argument shows, $h(x)$ cannot be concave, violating our assumption:

$$\begin{aligned} h(y) + \epsilon\big(h(x) - h(y)\big) &= \chi^{R_j}(y) + \epsilon\big(\chi^{R_i}(x) - \chi^{R_j}(y)\big) \\ &> \chi^{R_j}(y) + \epsilon\big(\chi^{R_j}(x) - \chi^{R_j}(y)\big) \\ &= \chi^{R_j}\big(y + \epsilon(x - y)\big) = h\big(y + \epsilon(x - y)\big), \end{aligned} \tag{8}$$

where the first equality is by the definition of h, the first inequality is because $\chi^{R_i}(x) > \chi^{R_j}(x)$ and $\epsilon > 0$, the second equality is because χ^{R_j} is affine, and the last equality is by the definition of h.

We turn our attention now to a specific triangulation of a product of simplices, referred to as the *staircase triangulation*. Let $S_i \subseteq \mathbb{R}^n$ be a simplex so that $\operatorname{vert}(S_i) = \{\nu_{i0}, \dots, \nu_{in}\}$ and let $S := \prod_{i=1}^d S_i$ be the Cartesian product of these simplices. The extreme points of S, which are $\prod_{i=1}^d \{\nu_{i0}, \dots, \nu_{in}\}$, can then be depicted on the grid $\mathcal{G}$ given by $\{0, \dots, n\}^d$. More specifically, the extreme point $(\nu_{ij_i})_{i=1}^d$ will be associated with the grid-point $\{j_i\}_{i=1}^d$. The coordinates of the extreme point are recovered from those of the grid-point using *grid-labels*, markers that label coordinate j along direction i as ν_{ij}. We remark that although grid-labels depend on the specific geometry of S, the grid only depends on the number of simplices and the dimension of each simplex.

A *monotone staircase* is a sequence of $dn+1$ points $p_0, \dots, p_{dn}$, where $p_i \in \mathcal{G}$ for all $i \in \{0, \dots, dn\}$ and the sequence satisfies the following properties, (i) $p_0 = (0, \dots, 0)$ and (ii) for all $i \in \{1, \dots, dn\}$, $p_i - p_{i-1} = e_k$, where $k \in \{1, \dots, d\}$. We refer to movement from p_{i-1} to p_i as the i^{th} move. Since $p_i \in \mathcal{G}$ for all i, by property (ii) there are exactly n moves in each coordinate direction. Moreover, $p_{dn} = (n, \dots, n)$. Thus, the monotone staircase is a lattice path of monotonically increasing points in $\mathbb{Z}^n$ from $(0, \dots, 0)$ to $(n, \dots, n)$, hence resembling a staircase, where each step is of possibly different height. The staircase can be specified succinctly as a vector $\pi = (\pi_1, \dots, \pi_{dn})$, where $\pi_i \in \{1, \dots, d\}$ is the coordinate direction of the i^{th} move. Thus, we will refer to such vector π as *movement* vector in the grid $\mathcal{G}$. Given a vector π, we will often need to track where the k^{th} move leaves us on the grid. This is obtained using the transformation Π, which is defined as $\Pi(\pi, k) := p_0 + \sum_{j=1}^k e_{\pi_j}$. The corresponding staircase can then be recovered as $\big(\Pi(\pi, k)\big)_{k=0}^{dn}$. In Figure 2, we see the set of all monotone staircases on the 4×3 grid.

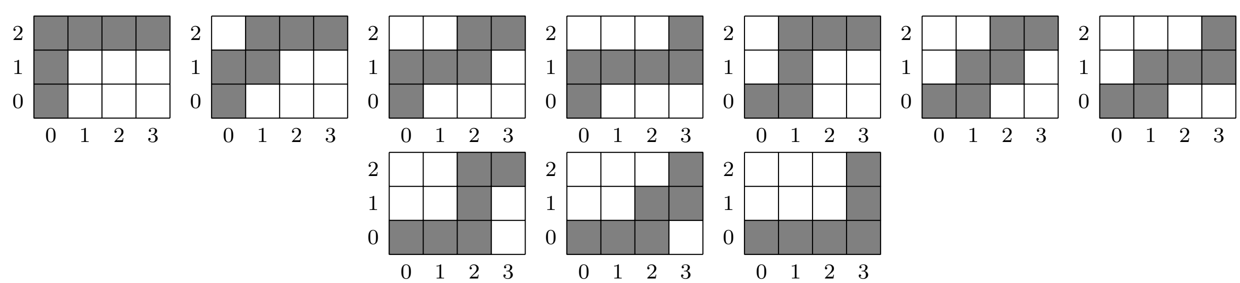


Figure 2. Monotone staircases in the 4×3 grid.

As described before, with each grid point $p_k = (j_i)_{i=1}^d$, we associate the extreme point of S, $(\nu_{ij_1}, \dots, \nu_{dj_d})$, which we denote as $\operatorname{ext}(S, p_k)$. The set of all extreme points associated with the grid-points along a staircase π describe a simplex Ξ_π. In particular, if we let $p_k = \Pi(\pi, k)$ then $\Xi_\pi := \operatorname{conv}\big(\{\operatorname{ext}(S, p_0), \operatorname{ext}(S, p_1), \dots, \operatorname{ext}(S, p_{dn})\}\big)$. That this set defines a simplex follows from the affine independence of $\operatorname{ext}(S, p_k)$ for $k \in \{0, \dots, dn\}$, which in turn follows from linear independence of difference vectors $\operatorname{ext}\big(S, p_{k+1}\big) - \operatorname{ext}\big(S, p_k\big)$ for $k = 0, \dots, dn-1$. In particular, the difference vectors for a move k along i^{th} and another move k' along i' coordinate, where $i' \neq i$, are linearly independent because these vectors are non-zero along different variables. On the other hand, the difference vectors for moves, all of which are along the i^{th} coordinate, are linearly independent because they are of the form $(0_1, \dots, 0_{i-1}, \nu_{ij+1} - \nu_{ij}, 0_{i+1}, \dots, 0_d)$, where $0_{i'}$ is the zero vector in the subspace of variables defining $S_{i'}$ and $\nu_{ij+1} - \nu_{ij}$ are the difference vectors between adjacent extreme points of S_i.

Definition 3 (Staircase Triangulation [9]). The set of all monotone staircases in the grid $\mathcal{G}$ defines a staircase triangulation of $\prod_{i=1}^d S_i$. Each monotone staircase, defined by the movement vector π, yields a simplex, $\Xi(\pi)$, in this triangulation. $\square$

The staircase triangulation, when the simplotope S is a product of standard simplices, is illustrated in [33]. We argue that the staircase triangulation is a triangulation of S. Observe that the $\mathbb{R}^{n\times n}$ matrix, $M_i := (\nu_{i1} - \nu_{i0}, \ldots, \nu_{in} - \nu_{i0})$, is invertible because S_i is a full-dimensional simplex. Then, we consider the affine mapping that maps an $x \in \mathbb{R}^n$ to $UM_i^{-1}(x - \nu_{i0})$, where $U \in \mathbb{R}^{n\times n}$ is an upper-triangular matrix of all ones. Under this transformation, the simplex S_i maps to $\Lambda_i := \{z_i \in \mathbb{R}^n \mid 0 \le z_{in} \le \cdots \le z_{i1} \le 1\}$. Given a point $s = (s_1, \ldots, s_n) \in S$, we obtain $z = (z_1, \ldots, z_n)$, where $z_i = UM_i^{-1}(s_i - \nu_{i0})$. Then, we sort the coordinates of the vector z in a non-increasing order so that if $z_{ij} = z_{ij'}$ for some i and $j' > j$, we place z_{ij} ahead of $z_{ij'}$ in the ordering. Now, for any $k \in \{1, \ldots, dn\}$, if z_{ij} is the k^{th} order-statistic, we define $\Theta(k) = \bigl(\Theta_1(k), \Theta_2(k)\bigr) := (i, j)$. Given a movement vector π, Θ can be recovered using $\Theta_1(k) = \pi_k$, and $\Theta_2(k) = \bigl|\{j \mid \pi(j) = \pi(k), 1 \le j \le k\}\bigr|$. Then, we verify that z (resp. s) belongs to the simplex whose extreme points form the monotone staircase defined by the movement vector $\pi' = \bigl(\Theta_1(k)\bigr)_{k=1}^{dn}$, *i.e.*, $\operatorname{conv}\bigl(\bigl\{\operatorname{ext}\bigl(\Lambda, \Pi(\pi', 0)\bigr), \ldots, \operatorname{ext}\bigl(\Lambda, \Pi(\pi', dn)\bigr)\bigr\}\bigr)$ (resp. $\operatorname{conv}\bigl(\bigl\{\operatorname{ext}\bigl(S, \Pi(\pi', 0)\bigr), \ldots, \operatorname{ext}\bigl(S, \Pi(\pi', dn)\bigr)\bigr\}\bigr)$). We denote $\operatorname{ext}\bigl(\Lambda, \Pi(\pi', k)\bigr)$ as p_k. Note that $p_0 = (0_1, \ldots, 0_d)$, where 0_i is a zero vector in the space of z_i variables and $p_k = p_{k-1} + e_{\Theta(k)}$, where $e_{\Theta(k)}$ is the standard basis vector in the direction of $z_{\Theta(k)}$. Then,

$$z = (1 - z_{\Theta(1)})p_0 + \sum_{k=1}^{dn-1} \bigl(z_{\Theta(k)} - z_{\Theta(k+1)}\bigr)p_k + z_{\Theta(dn)}p_{dn}. \tag{9}$$

Thus, each point z belongs to some simplex in the staircase triangulation. Moreover, points that belong to two simplices have at least two consistent orderings of their coordinates, ensuring some of them are equal, implying that the point belongs to a common face of both simplices. Thus, monotone staircases triangulate S.

If a function $f : [0,1]^n \to \mathbb{R}$ is *supermodular* when restricted to $\{0,1\}^n$ and concave extendable from $\{0,1\}^n$ then $\operatorname{conc}_{[0,1]^n}(f)(x)$ coincides with the *Lovász extension* of f [22, 39]. The Lovász extension interpolates the staircase triangulation of $[0,1]^n$, which can be regarded as a Cartesian product of simplices. In this setting, the corresponding triangulation is referred to as *Kuhn's triangulation*. We relate the concave envelope over a product of simplices to that over certain subsets of $[0,1]^n$, using a result in [39].

Definition 4 ([43]). A function $\eta(x) \colon S \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be supermodular if $\eta(x' \vee x'') + \eta(x' \wedge x'') \ge \eta(x') + \eta(x'')$ for all x', $x'' \in S$. Here, $x' \vee x''$ denotes the component-wise maximum and $x' \wedge x''$ denotes the component-wise minimum of x' and x'' and S is assumed to be a lattice, that is, $x' \vee x''$ and $x' \wedge x''$ belong to S whenever x' and x'' belong to S. $\square$

Proposition 3. *Let* $\Delta_i = \{z_i \in \mathbb{R}^{n+1} \mid 0 \le z_{in} \le \cdots \le z_{i1} \le z_{i0} = 1\}$ *and* $\Delta = \prod_{i=1}^d \Delta_i$. *Then,* $\operatorname{vert}(\Delta)$ *defines a lattice. Consider a function* $\eta : \Delta \to \mathbb{R}$ *that is supermodular when restricted to* $\operatorname{vert}(\Delta)$ *and is concave-extendable from* $\operatorname{vert}(\Delta)$. *Then, the concave envelope of* $\eta(\cdot)$ *over* Δ *is given by the staircase triangulation of* Δ, *i.e.* $\operatorname{conc}_\Delta(\eta)(z) = \eta^{\mathcal{S}}(z)$ *for every* $z \in \Delta$, *where* $\mathcal{S}$ *is the staircase triangulation of* Δ.

Proof. See Appendix A.

Although detecting whether a function is supermodular is NP-Hard [8], there are important special cases where this property can be readily detected [43]. A particularly useful result establishes supermodularity of a composition of functions.

Lemma 1 **(Lemma 2.6.4 in [43])**. *Consider a lattice* X *and let* $K = \{1, \ldots, k\}$. *For* $i \in K$, *let* $f_i(x)$ *be increasing supermodular (resp. submodular) functions on* X *and* Z_i *be convex subsets of* $\mathbb{R}$. *Assume* $Z_i \supseteq \{f_i(x) \mid x \in X\}$. *Let* $g(z_1, \ldots, z_k, x)$ *be supermodular in* $(z_1, \ldots, z_k, x)$ *on* $\prod_{i=1}^k Z_i \times X$. *If for* $i \in K$, $\bar{z}_{i'} \in Z_{i'}$ *for* $i' \in K \setminus \{i\}$, *and* $\bar{x} \in X$, $g(\bar{z}_1, \ldots, \bar{z}_{i-1}, z_i, \bar{z}_{i+1}, \ldots, \bar{z}_k, \bar{x})$ *is increasing (resp. decreasing) and convex in* z_i *on* Z_i *then* $g\bigl(f_1(x), \ldots, f_k(x), x\bigr)$ *is supermodular on* X. $\square$

By choosing $g(z_1,\dots,z_k,x)$ appropriately as $z_1z_2\cdots z_k$ or $-z_1z_2\cdots z_k$, it follows that a product of nonnegative, increasing (resp. decreasing) supermodular functions is also nonnegative increasing (resp. decreasing) and supermodular; see Corollary 2.6.3 in [43]. Also, it follows trivially that a conic combination of supermodular functions is supermodular.

3. On finitely many estimators for inner functions

In Section 3.1, we devise a facet generation algorithm, Algorithm 1, that separates $\text{conc}_P(\bar{\phi})$ assuming the outer-function $\phi(\cdot)$ is concave-extendable and supermodular. As a result, we discover various interesting properties of the envelope $\text{conc}_P(\bar{\phi})$. In Section 3.2, we develop a decomposition result that applies to a class of bilinear functions and show that, for a function $\phi(\cdot)$ from this class, $\text{conc}_Q(\bar{\phi})$ is obtained by summing the concave envelopes of each bilinear term. Last, in Section 3.3, we simultaneously convexify the hypographs of multiple functions whose concave envelopes share the same triangulation. In particular, we use Algorithm 1 to solve the facet generation problem of a vector of concave-extendable supermodular functions.

3.1. Tractable concave envelopes

In this subsection, we will, under previously stated technical conditions on the outer-function $\phi(\cdot)$, characterize, in closed form, $\text{conc}_Q(\bar{\phi})$ and solve the facet generation problem of $\text{conc}_P(\bar{\phi})$ in $\mathcal{O}(dn\log d)$ time. More specifically, given a $\bar{u}$, we will show that the following algorithm generates a facet-defining inequality for $\text{conc}_P(\bar{\phi})$ that is tight at $\bar{u}$:

- Given $\bar{u}\in P$, find the $\bar{s}$ so that $(\bar{u},\bar{s})\in PQ'$, and use the staircase triangulation of Q to generate a facet-defining inequality $\phi\le\langle\alpha,s\rangle+b$ of $\text{conc}_Q(\bar{\phi})$ which is tight at $\bar{s}$.
- Let $J=(J_1,\dots,J_d)$, where $J_i=\{j\mid \bar{u}_{ij}=\bar{s}_{ij}\}$. Then, we show that the inequality $\phi\le\big\langle\alpha,\Gamma_J(u)\big\rangle+b$ defines a facet of $\text{conc}_{Q_J}(\bar{\phi})$ and thus, leveraging Proposition 2, a facet of $\text{conc}_P(\bar{\phi})$ which is tight at $\bar{u}$.

Moreover, specializing to a multilinear function $\phi(\cdot)$, we find that $\text{conc}_P(\bar{\phi})$, the envelope over P, is obtained simply by extending the domain, from Q to P, of certain inequalities that describe $\text{conc}_Q(\bar{\phi})$, the envelope over Q. As an illustration, we specialize our result to a bilinear term $\phi(\cdot)$ with only one non-trivial underestimator for each inner function, and show that the inequalities describe both $\text{conc}_P(\bar{\phi})$ and $\text{conv}_P(\bar{\phi})$ in this setting.

We start by presenting an affine transformation between Q and Δ and its properties. Lemma 2 in [17] shows that for each $i=\{1,\dots,d\}$, the simplex Q_i can be expressed as:

$$0\le\frac{s_{in}-s_{in-1}}{a_{in}-a_{in-1}}\le\cdots\le\frac{s_{i1}-s_{i0}}{a_{i1}-a_{i0}}\le 1,\quad s_{i0}=a_{i0}. \tag{10}$$

Observe that, because of the last equality, Q_i is not full-dimensional. Nevertheless, we can mimic the construction following Definition 3 using the above representation of Q_i. Here, for an $s_i\in Q_i$ we define $z_i\in\Delta_i:=\{z\mid 0\le z_{in}\le\cdots\le z_{i1}\le z_{i0}=1\}$ as follows:

$$z_{i0}=1\qquad\text{and}\qquad z_{ij}=\frac{s_{ij}-s_{ij-1}}{a_{ij}-a_{ij-1}}\quad\text{for } j=1,\dots,n. \tag{11}$$

We refer to the above transformation as Z_i so that $z_i=Z_i(s_i)$. The inverse of Z_i is then defined as:

$$s_{ij}=a_{i0}z_{i0}+\sum_{k=1}^{j}(a_{ik}-a_{ik-1})z_{ik}\quad\text{for } j=0,\dots,n. \tag{12}$$

The transformation Z_i has a few properties that will be useful in our development. First, being an affine transformation, it maps vertices of Q_i to those of Δ_i. Recall that $\text{vert}(Q_i)=\{v_{i0},\dots,v_{in}\}$, where $v_{ij}=(a_{i0},\dots,a_{ij-1},a_{ij},\dots,a_{in})$. After the transformation, the vertex v_{ij} maps to $\zeta_{ij}=\sum_{j'=0}^{j}e_{ij'}$. Clearly, $\text{vert}(\Delta_i)$ form a chain, where $\zeta_{i0}\le\cdots\le\zeta_{in}$. Consequently, $\text{vert}(Q_i)$ also form a chain, where the vertices are ordered as $v_{i0} \le \cdots \le v_{in}$. This is because (12) is an increasing mapping. More specifically, if $z_i, z'_i \in \Delta_i$ such that $z_i \ge z'_i$, then $z_{i0} = 1$ and $a_{ik} - a_{ik-1} > 0$ imply that $s_i = Z_i^{-1}(z_i) \ge Z_i^{-1}(z'_i) = s'_i$. Then, the definition of $\vee$ and $\wedge$ as coordinate-wise maximum and minimum and the observation that $\mathrm{vert}(Q_i)$ form a chain together imply

$$Z_i^{-1}(\zeta_{ij} \vee \zeta_{ij'}) = Z_i^{-1}(\zeta_{ij}) \vee Z_i^{-1}(\zeta_{ij'}) \qquad Z_i^{-1}(\zeta_{ij} \wedge \zeta_{ij'}) = Z_i^{-1}(\zeta_{ij}) \wedge Z_i^{-1}(\zeta_{ij'}) \tag{13}$$

$$Z_i(v_{ij} \vee v_{ij'}) = Z_i(v_{ij}) \vee Z_i v_{ij'} \qquad Z_i(v_{ij} \wedge v_{ij'}) = Z_i(v_{ij}) \wedge Z_i v_{ij'}, \tag{14}$$

In other words, Z_i^{-1} (resp. Z_i) distributes over $\vee$ and $\wedge$ as long as the arguments are vertices of Δ_i (resp. Q_i). The vertex $(v_{ij_i})_{i=1}^d$ maps to $(\zeta_{ij_i})_{i=1}^d$ under Z_i and, graphically, will be represented by the same grid-point $(j_i)_{i=1}^d$ on $\mathcal{G}$. Consider two sets defined as the Cartesian product of simplices, $Q := \prod_{i=1}^d Q_i$ and $\Delta := \prod_{i=1}^d \Delta_i$. It follows that we may define a lattices over $\mathrm{vert}(Q)$ and $\mathrm{vert}(\Delta)$, *i.e.*, Cartesian products of d chains each with $n+1$ elements. Moreover, observe that the affine transformation $Z(s) = \bigl(Z_1(s_1), \dots, Z_n(s_n)\bigr)$, which maps from Q to Δ, is a *lattice isomorphism* from $\mathrm{vert}(Q)$ to $\mathrm{vert}(\Delta)$. The lattice isomorphism plays an important role in relating the concave envelope of the supermodular function, $\phi(\cdot)$, over Q to the concave envelope of a function over Δ. The latter function is also supermodular because $\mathrm{vert}(\Delta)$ is obtained from $\mathrm{vert}(Q)$ via a lattice isomorphism. The envelopes are transformed from one set to another because convexification commutes with affine transformations. What remains is the characterization of the concave envelope of the latter function, which is given by the staircase triangulation of Δ as described in Proposition 3. We summarize this construction in the next lemma, which will apply to our context simply by replacing R_i with Q_i.

Lemma 2. *Let $R = \prod_{i=1}^d R_i \subseteq \mathbb{R}^{d\times(n+1)}$, where R_i is a simplex in $\mathbb{R}^{n+1}$ whose vertices $\{\bar{v}_{i0}, \dots, \bar{v}_{in}\}$ form a chain, i.e., $\bar{v}_{i0} \le \bar{v}_{i1} \le \cdots \le \bar{v}_{in}$ and $\bar{v}_{ij0} = a_{i0}$ for all $j \in \{0, \dots, n\}$. Let $\psi : R \to \mathbb{R}$ be a function which is concave-extendable from* $\mathrm{vert}(R)$ *and is supermodular over the lattice defined over R. Then, $\mathrm{conc}_R(\psi)(r) = \psi^{\mathcal{S}}(r)$ for every $r \in \mathbb{R}$, where $\mathcal{S}$ is the staircase triangulation of R.*

Proof. By translating R if necessary, we may assume without loss of generality that $a_{i0} \ne 0$. Since $\{\bar{v}_{i0}, \dots, \bar{v}_{in}\}$ are the vertices of a simplex, they are affinely independent and $V_i = \bigl[\bar{v}_{i0} \, \dots \, \bar{v}_{in}\bigr]$, is an invertible $(n+1)\times(n+1)$ matrix. Let $U \in \mathbb{R}^{(n+1)\times(n+1)}$ be the upper triangular matrix of all ones. It follows that $A_i(r_i) := (U \circ V_i^{-1})(r_{ij})$ maps $\bar{v}_{ij}$ to ζ_{ij}, and its inverse $A_i^{-1}(z_i) := (V_i \circ U^{-1})(z_i)$ maps ζ_{ij} to $\bar{v}_{ij}$, for every $j \in \{0, \dots, n\}$. Therefore, the affine transformation $A(r_1, \dots, r_d) = \bigl(A_i(r_i)\bigr)_{i=1}^d$ induces a lattice isomorphism from $\mathrm{vert}(R)$ to $\mathrm{vert}(\Delta)$, that is, (13) and (14) are satisfied if Z replaced by A and v_{ij} with $\bar{v}_{ij}$. The lattice isomorphism implies that the function $\eta(z) := \psi\bigl(A^{-1}(z)\bigr)$ is supermodular over the induced lattice on $\mathrm{vert}(\Delta)$. Since A is affine, $\eta(\cdot)$ is concave extendable over $\mathrm{vert}(\Delta)$ and $\mathrm{conc}_R(\psi)(r) = \mathrm{conv}_\Delta(\eta)(A(r))$. Since, by Proposition 3, the concave envelope of $\eta(\cdot)$ over Δ is determined by the staircase triangulation, the concave envelope of ψ is given by a triangulation of R, obtained by affinely transforming each simplex in the staircase triangulation of Δ using A^{-1}. This yields the staircase triangulation of R, where ζ_{ij} is mapped to the vertex $\bar{v}_{ij}$. $\square$

In particular, we will use the following form of Lemma 2.

Proposition 4. *Consider an outer-function $\phi : \mathbb{R}^d \to \mathbb{R}$, and let $\bar{\phi} : \mathbb{R}^{d\times(n+1)} \to \mathbb{R}$ be its extension defined as $\bar{\phi}(s) = \phi(s_{1n}, \dots, s_{dn})$ for $s \in \mathbb{R}^{d\times(n+1)}$. Assume that $\bar{\phi}(\cdot)$ is concave-extendable from* $\mathrm{vert}(Q)$ *and is supermodular when restricted to the lattice set* $\mathrm{vert}(Q)$. *Let $\mathcal{S}$ be the staircase triangulation of Q. Then, for each $s \in Q$, $\mathrm{conc}_Q(\bar{\phi})(s) = \bar{\phi}^{\mathcal{S}}(s)$.* $\square$

Remark 1. Suppose that $\phi(\cdot)$ is shown to be supermodular over $\prod_{i=1}^d [a_{i0}, a_{in}]$, we show that $\bar{\phi}(\cdot)$ is supermodular over $\mathrm{vert}(Q)$. Since $\prod_{i=1}^d \{a_{i0}, \dots, a_{in}\} \subseteq \prod_{i=1}^d [a_{i0}, a_{in}]$, it follows readily that

$\phi(\cdot)$ is supermodular over $\prod_{i=1}^d \{a_{i0},\dots,a_{in}\}$. Then, the supermodularity of $\bar{\phi}(\cdot)$ over the lattice set $\operatorname{vert}(Q)$ follows since for two extreme points $v' := (v_{1j_1'},\dots,v_{dj_d'})$ and $v'' := (v_{1j_1''},\dots,v_{dj_d''})$ of Q

$$\begin{aligned}\bar{\phi}(v'\wedge v'')+\bar{\phi}(v'\vee v'') &= \phi(a_{1j_1'}\wedge a_{1j_1''},\dots,a_{dj_d'}\wedge a_{dj_d''})+\phi(a_{1j_1'}\vee a_{1j_1''},\dots,a_{dj_d'}\vee a_{dj_d''})\\ &\ge \phi(a_{1j_1'},\dots,a_{dj_d'})+\phi(a_{1j_1''},\dots,a_{dj_d''})\\ &= \bar{\phi}(v')+\bar{\phi}(v''),\end{aligned}$$

where the inequality follows from the supermodularity of $\phi(\cdot)$ over $\prod_{i=1}^d \{a_{i0},\dots,a_{in}\}$. $\square$

Next, we explicitly derive the inequalities that interpolate the function $\bar{\phi}(\cdot)$ over the extreme points of each simplex in the staircase triangulation. We first recall the connections between monotone staircases in the grid and simplices of the triangulation as it applies in this context. Let Ω be the set of movement vectors that define monotone staircases over the grid $\mathcal{G}$ given by $\{0,1,\dots,n\}^d$. For $\pi\in\Omega$, the k^{th} extreme point $\operatorname{ext}(Q,\Pi(\pi,k))$ will be denoted as $\mathcal{V}(\pi,k)$. The corresponding simplex $\operatorname{conv}\bigl(\mathcal{V}(\pi,0),\dots,\mathcal{V}(\pi,dn)\bigr)$ will be denoted as Υ_π and the triangulation $\{\Upsilon_\pi\}_{\pi\in\Omega}$ as Υ. In addition, we define $m(i,j)=k$ if $\pi(k)=i$ and $j=\sum_{k'\le k}\mathbb{1}\bigl(\pi(k')=\pi(k)\bigr)$, *i.e.*, for a pair (i,j), $m(i,j)$ returns k if the k^{th} movement is the j^{th} step in coordinate direction i. Observe that $\mathcal{V}\bigl(\pi,m(i,j)\bigr)-\mathcal{V}\bigl(\pi,m(i,j)-1\bigr)=(0_1,\dots,0_{i-1},v_{ij}-v_{ij-1},0_{i+1},\dots,0_d)$, where 0_k is the zero vector in the space of s_k variables. If $\mathcal{V}(\pi,k)=(s_1,\dots,s_d)$, we denote $(s_{1n},\dots,s_{dn})$ by $\mathcal{V}_{\cdot n}(\pi,k)$. Let $\langle\alpha^\pi,s\rangle+\beta^\pi$ be the unique affine function so that, for all i, $\alpha^\pi_{i0}=0$ and $\bar{\phi}(s)=\langle\alpha^\pi,s\rangle+\beta^\pi$ for $s\in\operatorname{vert}(\Upsilon_\pi)$. Then, to derive expressions for α^π,β^π observe that:

$$\begin{aligned}\phi\Bigl(\mathcal{V}_{\cdot n}\bigl(\pi,m(i,j)\bigr)\Bigr)-\phi\Bigl(\mathcal{V}_{\cdot n}\bigl(\pi,m(i,j)-1\bigr)\Bigr) &= \Bigl\langle\alpha^\pi,\mathcal{V}_{\cdot n}\bigl(\pi,m(i,j)\bigr)-\mathcal{V}_{\cdot n}\bigl(\pi,m(i,j)-1\bigr)\Bigr\rangle\\ &= \langle\alpha^\pi_i,v_{ij}-v_{ij-1}\rangle=(a_{ij}-a_{ij-1})\sum_{j'=j}^{n}\alpha^\pi_{ij'}.\end{aligned}$$

Let $\vartheta(\pi,i,j)=\frac{\phi(\mathcal{V}_{\cdot n}(\pi,m(i,j)))-\phi(\mathcal{V}_{\cdot n}(\pi,m(i,j)-1))}{a_{ij}-a_{ij-1}}$. Then, by differencing the equations and fitting the equation at $\mathcal{V}(\pi,0)$, we obtain the following explicit formulae:

$$\begin{aligned}\alpha^\pi_{ij} &= \begin{cases}0 & j=0\\ \vartheta(\pi,i,j)-\vartheta(\pi,i,j+1) & 1\le j<n\\ \vartheta(\pi,i,n) & j=n\end{cases}\\ b^\pi &= \phi\bigl(\mathcal{V}_{\cdot n}(\pi,0)\bigr)-\bigl\langle\alpha^\pi,\mathcal{V}(\pi,0)\bigr\rangle.\end{aligned}\tag{15}$$

In the next example, we illustrate how Proposition 4 and (15) yield tighter convex relaxations than ones obtained using techniques in [23, 39].

Example 1. Consider a nonlinear function $\sqrt{x_1+x_2^2}$ over $[0,5]\times[0,2]$. Let $s_1(x):=(0,x_1)$, and $s_2(x):=\bigl(0,\ \max\{\frac{3}{4}x_2^2,2x_2-1\},\ x_2^2\bigr)$. It turns out $a_1=(0,5)$ and $a_2=(a_{20},a_{21},a_{22})=(0,3,4)$ is a vector of upper bounds for $s_1(x)$ and $s_2(x)$ over $[0,5]\times[0,2]$, respectively. Moreover, it can be verified that, for each $x\in[0,5]\times[0,2]$, the point $s_i(x)$ satisfies (10), that is $s_i(x)\in Q_i$, where Q_i is the simplex whose extreme points are defined as in (3). Clearly, $\sqrt{s_{11}+s_{22}}$ is submodular over $[0,5]\times[0,4]$ and is convex-extendable from extreme points of $Q_1\times Q_2$. By Proposition 4 and (15) and after substitution, we obtain that $\max\bigl\{\psi_1(x),\psi_2(x),\psi_3(x)\bigr\}$ underestimates $\sqrt{x_1+x_2^2}$ for $x\in[0,5]\times[0,2]$, where

$$\begin{aligned}\psi_1(x) &:= \frac{\sqrt{5}}{5}x_1+\left(\frac{\sqrt{8}-\sqrt{5}}{3}-\frac{\sqrt{9}-\sqrt{8}}{1}\right)\max\left\{\frac{3}{4}x_2^2,2x_2-1\right\}+\frac{\sqrt{9}-\sqrt{8}}{1}x_2^2\\ \psi_2(x) &:= \frac{\sqrt{8}-\sqrt{3}}{5}x_1+\left(\frac{\sqrt{3}}{3}-\frac{\sqrt{9}-\sqrt{8}}{1}\right)\max\left\{\frac{3}{4}x_2^2,2x_2-1\right\}+\frac{\sqrt{9}-\sqrt{8}}{1}x_2^2\\ \psi_3(x) &:= \frac{\sqrt{9}-\sqrt{4}}{5}x_1+\left(\frac{\sqrt{3}}{3}-\frac{\sqrt{4}-\sqrt{3}}{1}\right)\max\left\{\frac{3}{4}x_2^2,2x_2-1\right\}+\frac{\sqrt{4}-\sqrt{3}}{1}x_2^2,\end{aligned}$$

whose convexity can be verified easily.

The standard factorable relaxation introduces $f = x_2^2$, $g = x_1 + f$, and $h = \sqrt{g}$ to represent $\sqrt{x_1 + x_2^2}$. Then, $\sqrt{g}$ is relaxed to $\frac{1}{3}g$ over $[0,9]$, yielding a convex underestimator $\frac{1}{3}(x_1 + x_2^2)$ for $\sqrt{x_1 + x_2^2}$ over $[0,5] \times [0,2]$. The technique in [39] relaxes $\sqrt{x_1 + f}$, convex-extendable and submodular over $\{0,5\} \times \{0,4\}$, to $\max\bigl\{\frac{x_1}{5} + \frac{f}{2},\ \frac{\sqrt{5}}{5}x_1 + \frac{3-\sqrt{5}}{4}f\bigr\}$, yielding a convex underestimator $\max\bigl\{\frac{x_1}{5} + \frac{x_2^2}{2},\ \frac{\sqrt{5}}{5}x_1 + \frac{3-\sqrt{5}}{4}x_2^2\bigr\}$. It turns out that $\max\bigl\{\frac{x_1}{5} + \frac{x_2^2}{2},\ \frac{\sqrt{5}}{5}x_1 + \frac{3-\sqrt{5}}{4}x_2^2\bigr\} \geq \frac{1}{2}\bigl(\frac{x_1}{5} + \frac{x_2^2}{2}\bigr) + \frac{1}{2}\bigl(\frac{\sqrt{5}}{5}x_1 + \frac{3-\sqrt{5}}{4}x_2^2\bigr) \geq \frac{1}{3}(x_1 + x_2^2)$. Moreover, $\psi_1(x) \geq \frac{\sqrt{5}}{5}x_1 + \frac{3-\sqrt{5}}{4}x_2^2$ and $\psi_3(x) \geq \frac{x_1}{5} + \frac{1}{2}x_2^2$. $\square$

By construction, the inequality $\phi \leq \langle \alpha^\pi, s\rangle + b^\pi$ is tight over $\Upsilon_\pi := \operatorname{conv}\bigl(\mathcal{V}(\pi,0),\dots,\mathcal{V}(\pi,dn)\bigr)$, where, by tight, we mean that, for $s \in \Upsilon_\pi$, $\operatorname{conc}_Q(\bar{\phi})(s) = \langle \alpha^\pi, s\rangle + b^\pi$. More generally, the tight set for a valid inequality $f \leq \langle \alpha, x\rangle + b$ of a function $f : X \to \mathbb{R}$ will represent the set $\{x \in X \mid \operatorname{conc}_X(f)(x) = \langle \alpha, x\rangle + b\}$, and will be denoted as $T_f^{(\alpha,b)}(X)$. So, we can succinctly express our conclusion regarding the tight set of $\phi \leq \langle \alpha^\pi, s\rangle + b^\pi$ as $\Upsilon_\pi \subseteq T_{\bar{\phi}}^{(\alpha^\pi,b^\pi)}(Q)$. Although (15) describes the coefficients of the interpolating inequalities in the general case, we remark the following special case, where the coefficient α_{ij}^π becomes zero.

Remark 2. Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a multilinear function and consider a movement vector π such that the j^{th} move along coordinate i is adjacent to the $j+1^{\text{th}}$ move along this coordinate, *i.e.*, $m(i,j+1) = m(i,j) + 1$. Then, because the function $\phi(s_{1n},\dots,s_{dn})$ is affine when all but s_{in} is fixed, it follows that $\vartheta(\pi,i,j) = \vartheta(\pi,i,j+1)$, which implies by (15) that $\alpha_{ij}^\pi = 0$. $\square$

Algorithm 1 Facet-Generation over Q

```
procedure Facet-Generation(s̄)
    z̄ ← Z(s̄);
    BeginSort
        sort z̄ to find a movement vector π so that s̄ ∈ Υ_π;
        z̄_{i0} are sorted before z̄_{ij} for j ≠ 0;
        if, for any j ≥ 1, z_{ij} = z_{ij+1} then they are adjacent in the sorted order;
    EndSort
    compute an affine function φ^{Υ_π}(s) = ⟨α^π, s⟩ + b^π by using equation (15);
    return (α^π, b^π).
end procedure
```

Now, we use Proposition 4 to compute, at a given point $\bar{s} \in Q$, a non-vertical facet-defining inequality of the hypograph of $\operatorname{conc}_Q(\bar{\phi})(s)$. To this end, it suffices to find a movement vector π so that $\bar{s}$ belongs to the corresponding simplex Υ_π and then to compute the function $\phi^{\Upsilon_\pi}(s)$ using (15), where $\phi^{\Upsilon_\pi}(s)$ is the affine interpolating function tight at $\mathcal{V}(\pi,0),\dots,\mathcal{V}(\pi,dn)$. As shown, in our discussion following Definition 3, that a simple sorting of the coordinates of $\bar{z} := Z(\bar{s})$ reveals this staircase. In our context, $\bar{z}_{i0} = 1$ for all i and recall that to derive π we ignore the ordering of these coordinates assuming they are placed first in the sorted order. Then, if the $d+k$ largest coordinate of $\bar{z}$ is $\bar{z}_{ij}$, we let $\Theta(k) = (i,j)$ and define $\pi = \bigl(\Theta_1(1),\dots,\Theta_1(dn)\bigr)$. Slightly adjusting (9), we can express $\bar{s}$ as a convex combination of $\mathcal{V}(\pi,0),\dots,\mathcal{V}(\pi,dn)$ as follows:

$$\bar{s} = \bigl(1 - z_{\Theta(1)}\bigr)\mathcal{V}(\pi,0) + \sum_{k=1}^{dn-1}\bigl(z_{\Theta(k)} - z_{\Theta(k+1)}\bigr)\mathcal{V}(\pi,k) + z_{\Theta(dn)}\mathcal{V}(\pi,dn).$$

Corollary 1. *Assume that $\bar{\phi}(\cdot)$ is concave-extendable from $\operatorname{vert}(Q)$ and is supermodular when restricted to the vertices of Q. Given a point $\bar{s} \in Q$, Algorithm 1 takes $\mathcal{O}(dn\log d)$ operations to find a non-vertical facet-defining inequality of $\operatorname{conc}_Q(\bar{\phi})(\cdot)$ which is tight at $\bar{s}$.*

Proof. The correctness of Algorithm 1 is due to Proposition 4 and because Algorithm 1 identifies a π such that $\bar{s} \in \Upsilon_\pi$. The time complexity is $\mathcal{O}(dn \log d)$ because the computation of Z takes $\mathcal{O}(dn)$ time and d sorted lists, each of size n, can be merged in $\mathcal{O}(dn \log d)$ time using the d-way merge sort algorithm (see 5.4.1 in [19]). □

The inequality obtained using Algorithm 1 is facet-defining for $\operatorname{conc}_Q(\bar{\phi})$ since it interpolates the extension $\bar{\phi}(\cdot)$ over the extreme points of a simplex Υ_π. Moreover, when $\bar{s}$ belongs to a face of Q, this inequality describes a facet for the concave envelope of $\bar{\phi}(\cdot)$ over this face, a property that can be exploited, as shown in [17], to develop facet-defining inequalities over P. Recall that for $J = (J_1, \dots, J_d) \in \mathcal{J}$, where $\mathcal{J}$ is a collection of d-tuples defined as in (5), we defined $F_J := \prod_{i=1}^d F_{iJ_i}$ as a face of Q, where F_{iJ_i} is defined as the convex hull of $\{v_{ij} \mid j \in J_i\}$. We can also describe the face F_{iJ_i} as the set of points of Q_i which satisfy the following facet-defining constraints of Q_i at equality:

$$\frac{s_{ij+1} - s_{ij}}{a_{ij+1} - a_{ij}} \le \frac{s_{ij} - s_{ij-1}}{a_{ij} - a_{ij-1}} \qquad \text{for } j \notin J_i. \tag{16}$$

The face F_J can also be visualized as $Z^{-1}(F'_J)$, where F'_J is a face of Δ defined as:

$$F'_J = \big\{ z \in \Delta \mid z_{ij} = z_{i,j-1} \text{ for } i \in \{1, 2, \dots, d\},\ j \in \{0, 1, \dots, n\} \setminus J_i \big\}.$$

Corollary 2. *Assume $\bar{s} \in F_J$, and, when $\bar{s}$ is input, let (α^π, b^π) be the pair generated by Algorithm 1. If $\bar{\phi}(\cdot)$ is supermodular when restricted to the vertices of Q and concave-extendable from* $\operatorname{vert}(Q)$ *then (α^π, b^π) defines a non-vertical facet of* $\operatorname{conc}_{F_J}(\phi)(s)$ *and the corresponding inequality is tight at $\bar{s}$. Moreover, if $\phi(\cdot)$ is a multilinear function then, for all $j \notin J_i$, $\alpha^\pi_{ij} = 0$.*

Proof. As $\bar{s} \in F_J$ it follows from (16) that $\bar{z}_{ij+1} = \frac{\bar{s}_{ij+1} - \bar{s}_{ij}}{a_{ij+1} - a_{ij}} = \frac{\bar{s}_{ij} - \bar{s}_{ij-1}}{a_{ij} - a_{ij-1}} = \bar{z}_{ij}$ for all i and $j \notin J_i$. Therefore, the sorting in Algorithm 1 guarantees that the movement vector π is such that for all i and $j \notin J_i$, the $j+1^{\text{st}}$ move along coordinate i follows immediately after the j^{th} move. This implies that for $i \in \{1, \dots, d\}$ and $j \notin J_i$

$$m(i,j) + 1 = m(i, j+1). \tag{17}$$

Therefore, when $\phi(\cdot)$ is multilinear, the last statement in the result follows from Remark 2.

Under the assumption on $\bar{\phi}(\cdot)$, it follows from Corollary 1 that the inequality $\phi \le \langle \alpha^\pi, s\rangle + b^\pi$ is valid for $\operatorname{conc}_Q(\bar{\phi})$, and, thus, also valid for $\operatorname{conc}_{F_J}(\phi)$. We will show that $\dim\big(T_{\bar{\phi}}^{(\alpha^\pi, b^\pi)}(F_J)\big) = \dim(F_J)$. Clearly, we have $\dim\big(T_{\bar{\phi}}^{(\alpha^\pi, b^\pi)}(F_J)\big) \le \dim(F_J)$. Now, consider the simplex Υ_π defined by the movement vector π. It follows readily that $\Upsilon_\pi \cap F_J \subseteq T_{\bar{\phi}}^{(\alpha^\pi, b^\pi)}(F_J)$ since $\operatorname{conc}_Q(\bar{\phi})(s) = \operatorname{conc}_{F_J}(\bar{\phi})(s)$ for every $s \in F_J$. Thus, the proof is complete if we can show that $\dim(F_J) \le \dim\big(\operatorname{vert}(\Upsilon_\pi) \cap F_J\big)$, where $\operatorname{vert}(\Upsilon_\pi) = \{\mathcal{V}(\pi, 0), \dots, \mathcal{V}(\pi, dn)\}$. Since $\mathcal{V}(\pi, k-1) \le \mathcal{V}(\pi, k)$ for all $k = 1, \dots, dn$, it follows from (17) that for all i and $j \notin J_i$, the only grid point where the grid-label along i^{th} coordinate is v_{ij} is $\Pi\big(\pi, m(i,j)\big)$ with the corresponding point $\mathcal{V}(\pi, m(i,j))$. In other words, for all i and $j \notin J_i$, $\operatorname{vert}(\Upsilon_\pi) \cap \{s \mid s_i = v_{ij}\} = \mathcal{V}\big(\pi, m(i,j)\big)$ and $\mathcal{V}\big(\pi, m(i,j)\big) \in F_J$ if $j \in J_i$ because, for $i' \ne i$, $\Pi(\pi, m(i,j))_{i'} \in J_{i'}$. This implies that $\big|\operatorname{vert}(\Upsilon_\pi) \cap F_J\big| \ge dn + 1 - \sum_{i=1}^d \big|\bar{J}_i\big| = \dim(F_J) + 1$, where $\bar{J}_i = \{0, \dots, n\} \setminus J_i$. Since points in $\operatorname{vert}(\Upsilon_\pi)$ are affinely independent, we conclude that $\dim(F_J) \le \dim\big(\operatorname{vert}(\Upsilon_\pi) \cap F_J\big)$. □

This additional property of facet-defining inequalities for $\operatorname{conc}_Q(\bar{\phi})$ allows us to show that the facet generation problem of $\operatorname{conc}_P(\bar{\phi})$ can be solved in $\mathcal{O}(dn \log d)$. To prove this result, we need the following result shown in [17] which relates the concave envelope over the face F_J to the envelope over the projection Q_J, where Q_J is defined in (7). Recall that two sets $C \subseteq \mathbb{R}^c$ and $D \subseteq \mathbb{R}^d$ are *affinely isomorphic* if there is an affine map $f\colon \mathbb{R}^c \to \mathbb{R}^d$ that is a bijection between the points of the two sets. Consider an affine map $A\colon s_J \to \tilde{s}$ defined as

$$\tilde{s}_{ij} = s_{ij} \quad \text{for } j \in J_i \quad \text{and} \quad \tilde{s}_{ij} = (1 - \gamma_{ij}) s_{il(i,j)} + \gamma_{ij} s_{ir(i,j)} \quad \text{for } j \notin J_i, \tag{18}$$

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 $\gamma_{ij}=\frac{a_{ij}-a_{il(i,j)}}{a_{ir(i,j)}-a_{il(i,j)}}$. The inverse of A is defined as a map which transforms s to s_J.

LEMMA 3 **(Lemma 8 in [17])**. *Assume that* $\operatorname{conc}_Q(\bar\phi)$ *is a polyhedral function. Let* $J=(J_1,\dots,J_d)\in\mathcal{J}$. *Then,* $\operatorname{conc}_{F_J}(\bar\phi)(s)=\operatorname{conc}_{Q_J}(\bar\phi)(s_J)$ *for every* $s\in F_J$. *Let* $\phi\le\langle\alpha,s\rangle+b$ *be a valid inequality of* $\operatorname{conc}_Q(\bar\phi)(\cdot)$ *so that* $\alpha_{\bar J}=0$. *Then, the two tight sets,* $T_{\bar\phi}^{(\alpha,b)}(F_J)$ *and* $T_{\bar\phi}^{(\alpha_J,b)}(Q_J)$ *are affinely isomorphic under the affine map* A *defined in (18).* □

THEOREM 2. *Assume that* $\bar\phi(\cdot)$ *is concave-extendable from* $\operatorname{vert}(Q)$ *and is supermodular when restricted to the vertices of* Q. *Let* $\bar u\in P$. *Then, a facet-defining inequality of* $\operatorname{conc}_P(\bar\phi)$ *which is tight at* $\bar u$ *can be found in* $\mathcal{O}(dn\log d)$ *operations.*

Proof. Let $\bar u\in P$ and let $\bar s$ be the unique point so that $(\bar u,\bar s)\in PQ'$, which can be found in $\mathcal{O}(nd)$ and where PQ' is defined as in (4). Define $J=(J_1,\dots,J_d)$, where $J_i:=\{j\mid \bar u_{ij}=\bar s_{ij}\}$. It follows from Proposition 1 that $\bar s\in F_J$. Given $\bar s$ as input, let (α^π,b^π) denote the pair generated by Algorithm 1. Now, we derive an inequality $\phi\le\langle\alpha',s\rangle+b'$ defined so that $\langle\alpha',s\rangle+b'=\langle\alpha^\pi,\Gamma_J(s)\rangle+b^\pi$, where $\Gamma_J\colon s\in\mathbb{R}^{d(n+1)}\to\tilde s\in\mathbb{R}^{d(n+1)}$ is a linear map defined in (6). We will show that $\phi\le\langle\alpha',s\rangle+b'$ defines a facet of $\operatorname{conc}_{F_J}(\bar\phi)(s)$ which is tight at $\bar s$. Then, since $\alpha'_{\bar J}=0$, it follows from Lemma 3 that (α'_J,b') defines a non-vertical facet of $\operatorname{conc}_{Q_J}(\bar\phi)(\cdot)$ which is tight at $\bar s_J$. Therefore, by Proposition 2, (α',b') defines a non-vertical facet of $\operatorname{conc}_P(\bar\phi)(\cdot)$ which is tight at $\bar u$.

We now show that $\phi\le\langle\alpha',s\rangle+b'$ defines a facet of $\operatorname{conc}_{F_J}(\bar\phi)(\cdot)$ tight at $\bar s$. The validity of the inequality for $\operatorname{conc}_{F_J}(\bar\phi)(\cdot)$ follows because for every $s\in F_J$

$$\operatorname{conc}_{F_J}(\bar\phi)(s)\le\langle\alpha^\pi,s\rangle+b^\pi=\langle\alpha^\pi,\Gamma_J(s)\rangle+b^\pi=\langle\alpha',s\rangle+b',\tag{19}$$

where the first inequality holds by the validity of $\phi\le\langle\alpha^\pi,s\rangle+b^\pi$ for $\operatorname{conc}_{F_J}(\bar\phi)(\cdot)$, first equality holds because, by Proposition 1, $s\in F_J$ implies $s=\Gamma_J(s)$, and the second equality is by the definition of (α',b'). Now, the proof is complete because, by Corollary 2, the first inequality in (19) is satisfied at equality for $\dim(F_J)+1$ affinely independent points in F_J, and, in particular, for the point $\bar s$. □

We briefly summarize the algorithmic construction of the facet-defining inequalities for the concave envelope derived in Theorem 2. The construction uses three sets. The first set, $\Delta=\prod_{i=1}^d\Delta_i$, is defined in Proposition 3 and used to construct the concave envelope of concave-extendable supermodular functions. The second set, $P=\prod_{i=1}^d P_i$, is defined in (2) and abstracts the composite function structure. Inequalities obtained over P, upon substitution of underestimators, allow derivation of inequalities, in the space of original problem variables, that are valid for the hypograph of the original composite function. The third set is $Q=\prod_{i=1}^d Q_i$, whose vertex representation is given in (3) and its hyperplane representation is given in (10), and this set serves as a bridge for connecting the results on Δ with those for P. In particular, given a $\bar u\in P$, we obtain $\bar s\in Q$ using the concave envelope construction given in (4). Then, we transform $\bar s$ to $\bar z\in\Delta$ using the affine isomorphism Z defined in (11). The construction of the inequality proceeds in the opposite order. The inequality constructed for $\bar z\in\Delta$ is transformed into one for $\bar s\in Q$ simply using the inverse affine isomorphism, Z^{-1}. To derive the inequality for P, we revert the concave envelope construction, using the fact that, for $j\notin J_i$, each $\bar s_{ij}$ is obtained as a convex combination of $\bar u_{il(i,j)}$ and $\bar u_{ir(i,j)}$, described in (6). The set $J=(J_1,\dots,J_d)$ is derived during the concave envelope construction, where $J_i=\{j\mid\bar u_{ij}=\bar s_{ij}\}$, as defined in the proof of Theorem 2. This definition guarantees, by Proposition 1, that $\bar s$ belongs to the face F_J of Q as given in (16) and this face is used to describe, in the proof of Proposition 2 (see [17]), the set of points tight on the generated inequality. This tight set consists of $\dim(F_J)+1$ vertices of F_J and the points $\{u\mid\bar u_{ij}\le u_{ij}\le\bar s_{ij}\forall(i,j)\}$.

Next, we specialize our study to the case when the outer-function is multilinear. Let $\mathcal{S}$ be the staircase triangulation of Q. Recall that $\bar\phi^{\mathcal{S}}\colon \operatorname{aff}(Q)\to\mathbb{R}$ is obtained by extending the affine interpolation function of $\bar\phi(\cdot)$ over the affine hull of Q. In Proposition 4, we argued that, under the assumed conditions on $\bar{\phi}(\cdot)$, $\text{conc}_Q(\bar{\phi})(s) = \phi^{\mathcal{S}}(s)$ for every $s \in Q$. Next, we show that if a multilinear function $\bar{\phi}(\cdot)$ is supermodular over $\text{vert}(Q)$ then, for every $u \in P$, $\bar{\phi}^{\mathcal{S}}(u) = \text{conc}_P(\bar{\phi})(u)$. The point to note here is that, for the multilinear case, the concave envelope over P requires no other non-vertical inequalities beyond those needed to describe the concave envelope over Q, a result that does not hold in general for concave-extendable, supermodular functions.

COROLLARY 3. *Assume that the outer-function $\phi(\cdot)$ is multilinear and the extension $\bar{\phi}(\cdot)$ is supermodular when restricted to vertices of Q. Let $\mathcal{S}$ be the staircase triangulation of Q. Then, for every $u \in P$, $\text{conc}_P(\bar{\phi})(u) = \bar{\phi}^{\mathcal{S}}(u)$.*

Proof. Let $\bar{u} \in P$. Then, as in Proposition 2, compute $(\bar{u}, \bar{s}) \in PQ'$ and define $J = (J_1, \dots, J_d)$ so that $J_i = \{j \mid \bar{u}_{ij} = \bar{s}_{ij}\}$. Since $\phi(\cdot)$ is multilinear, its extension $\bar{\phi}(\cdot)$ is concave-extendable from $\text{vert}(Q)$. Moreover, $\bar{\phi}(\cdot)$ is supermodular when restricted to $\text{vert}(Q)$. Therefore, we may construct a facet-defining inequality using Algorithm 1, whose output will be denoted as the pair (α^π, b^π). Then, by Proposition 2, $\bar{s} \in F_J$ and, by Corollary 2, the inequality $\phi \le \langle \alpha^\pi, s \rangle + b^\pi$ defines a facet of $\text{conc}_{F_J}(\bar{\phi})$ such that for all i and $j \notin J_i$, $\alpha^\pi_{ij} = 0$. Moreover, by Corollary 1, this inequality is tight at $\bar{s}$. Then, it follows from Lemma 3 that $\phi \le \langle \alpha^\pi_J, s_J \rangle + b^\pi$ is a facet-defining inequality of $\text{conc}_{Q_J}(\bar{\phi})$ that is tight at $\bar{s}_J$. Thus, by Proposition 2, $\phi \le \langle \alpha^\pi, u \rangle + b^\pi$ is a facet-defining inequality of $\text{conc}_P(\bar{\phi})$ that is tight at $\bar{u}$. $\square$

We have described a way to develop inequalities for composite functions as long as the outer-function is supermodular and concave-extendable. To extend the applicability of this result, we now turn our attention to a particular linear transformation that can be used to convert some functions that are not ordinarily supermodular into supermodular functions. This transformation is well-studied when the domain of the function is $\{0,1\}^d$, a special case of $\text{vert}(Q)$. In this case, the transformation, often referred to as switching, chooses a set $D \subseteq \{1, \dots, d\}$ and considers a new function $\phi'(x_1, \dots, x_d)$ defined as $\phi(y_1, \dots, y_d)$, where $y_i = (1 - x_i)$ if $i \in D$ and $y_i = x_i$ otherwise. We will now generalize this switching operation to Q. To do so, we will need permutations σ_i of $\{0, \dots, n\}$ for each $i \in \{1, \dots, d\}$. We use the permutation σ_i to define an affine transformation that maps v_{ij} to $v_{i\sigma_i(j)}$. Let P^{σ_i} be a permutation matrix in $\mathbb{R}^{(n+1)\times(n+1)}$ such that, for all (i', j'), $P^{\sigma_i}_{i'j'} = 1$ when $i' = \sigma(j')$ and zero otherwise. Then, the affine transformation associated with σ_i is given by $A^{\sigma_i} = Z_i^{-1} \circ UP^{\sigma_i}U^{-1} \circ Z_i$, where $\circ$ denotes the composition operator and U is an upper triangular matrix of all ones. We let $A^\sigma(s) := \big(A^{\sigma_1}(s_1), \dots, A^{\sigma_d}(s_d)\big)$. We will particularly be interested in the case where $\sigma_i = (n, \dots, 0)$ for $i \in T$ and $\sigma_i = (0, \dots, n)$ otherwise. In this case, we denote $A^\sigma(s)$ by $s(T)$. Clearly, for $i \notin T$, $s(T)_i = s_i$. To compute $s(T)_i$ where $i \in T$, we use the following expression

$$s(T)_{ij} = a_{i0} + \sum_{k=1}^{j} (a_{ik} - a_{ik-1})(1 - z_{in+1-k}) \quad \text{for } j = 0, \dots, n, \tag{20}$$

where z denotes $Z(s)$. Then, we associate the outer-function $\phi(\cdot)$ with a function $\phi(T): \mathbb{R}^{d\times(n+1)} \to \mathbb{R}$ defined as $\phi(T)(s_1, \dots, s_d) = \phi\big(s(T)_{1n}, \dots, s(T)_{dn}\big)$ and we say that $\phi(T)$ is obtained from $\phi(\cdot)$ by switching T. It follows easily that $\text{conc}_Q(\bar{\phi})(s) = \text{conc}_Q\big(\phi(T)\big)\big(s(T)\big)$ (Similar conclusions can be easily drawn for switching with arbitrary permutations σ, where $\phi(\sigma)(s_1, \dots, s_d)$ is defined as $\phi\big(A^\sigma(s)_{1n}, \dots, A^\sigma(s)_{dn}\big)$. In this case, $\text{conc}_Q(\bar{\phi})(s) = \text{conc}_Q\big(\phi(\sigma)\big)\big((A^\sigma)^{-1}(s)\big)$.). More specifically, if the switched function $\phi(T)$ is supermodular when restricted to the vertices of Q then $\text{conc}_Q(\bar{\phi})$ is determined by the *switched* staircase triangulation specified by T, whose grid-labels are obtained by labelling coordinates directions, for $i \in T$, as they were, and, for $i \notin T$, in a reversed order $v_{in}, \dots, v_{i0}$. Then, for any movement vector π in the grid given by $\{0, \dots, n\}^d$, the corresponding simplex is defined as the convex hull of $\text{ext}\big(Q(T), \Pi(\pi, 0)\big), \dots, \text{ext}\big(Q(T), \Pi(\pi, dn)\big)$, where $Q(T) := \{s(T) \mid s \in Q\}$. The following result records the above construction for later use.

COROLLARY 4. *Assume function $\bar{\phi}(\cdot)$ is concave-extendable from* $\mathrm{vert}(Q)$. *Let T be a subset of $\{1,\ldots,d\}$. If $\phi(T)(\cdot)$ is supermodular when restricted to* $\mathrm{vert}(Q)$ *then* $\mathrm{conc}_Q(\bar{\phi})(\cdot)$ *is determined by the switched staircase triangulation specified by T.* $\square$

To illustrate inequalities in Corollary 4, we now consider a special case that was studied in [17] and used to improve factorable programming. The case setting requires that the outer-function $\phi(\cdot)$ is a bilinear term and each inner function has only one non-trivial underestimator. Note that the validity of the following inequalities was established in [17]. Here, we apply Corollary 4 to additionally show that these inequalities are facet-defining and that they describe the convex hull of the outer-function over P. This result also serves as an example of showing how (switched) staircase triangulation yields the convex hull over Q, and therefore, by Corollary 3, generates the convex hull over P.

COROLLARY 5. *Let $a_{i0} \le a_{i1} \le a_{i2}$ for $i=1,2$, and define $P := \big\{(u,f) \mid a_{i0} \le u_i \le \min\{f_i, a_{i1}\},\ f_i \le a_{i2},\ i=1,2\big\}$. Then, non-vertical facet-defining inequalities of the convex hull of $\big\{(u,f) \mid \phi = f_1 f_2,\ (u,f)\in P\big\}$ are given as follows:*

$$\begin{aligned}
&\phi \ge e_1 := a_{22}f_1 + a_{12}f_2 - a_{12}a_{22},\\
&\phi \ge e_2 := (a_{22}-a_{21})u_1 + (a_{12}-a_{11})u_2 + a_{21}f_1 + a_{11}f_2 + a_{11}a_{21} - a_{11}a_{22} - a_{12}a_{21},\\
&\phi \ge e_3 := (a_{22}-a_{20})u_1 + a_{20}f_1 + a_{11}f_2 - a_{11}a_{22},\\
&\phi \ge e_4 := (a_{12}-a_{10})u_2 + a_{21}f_1 + a_{10}f_2 - a_{12}a_{21},\\
&\phi \ge e_5 := (a_{21}-a_{20})u_1 + (a_{11}-a_{10})u_2 + a_{20}f_1 + a_{10}f_2 - a_{11}a_{21},\\
&\phi \ge e_6 := a_{10}f_2 + a_{20}f_1 - a_{10}a_{20},\\
&\phi \le r_1 := a_{20}f_1 + a_{12}f_2 - a_{12}a_{20},\\
&\phi \le r_2 := (a_{20}-a_{21})u_1 + (a_{11}-a_{12})u_2 + a_{21}f_1 + a_{12}f_2 - a_{11}a_{20},\\
&\phi \le r_3 := (a_{20}-a_{22})u_1 + a_{11}f_2 + a_{22}f_1 - a_{11}a_{20},\\
&\phi \le r_4 := (a_{10}-a_{12})u_2 + a_{21}f_1 + a_{12}f_2 - a_{10}a_{21},\\
&\phi \le r_5 := (a_{21}-a_{22})u_1 + (a_{10}-a_{11})u_2 + a_{22}f_1 + a_{11}f_2 - a_{10}a_{21},\\
&\phi \le r_6 := f_1a_{22} + a_{10}f_2 - a_{10}a_{22}.
\end{aligned}$$

Proof. Let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be the bilinear function $\phi(f_1,f_2) = f_1f_2$. We verify that the set of inequalities, $\phi \ge e_i$, $i=1,\ldots,6$, defines the set of non-vertical facets of the epigraph of $\mathrm{conv}_P(\phi)\big((u_1,f_1),(u_2,f_2)\big)$. Let $v_{i0} = (a_{i0},a_{i0})$, $v_{i1} = (a_{i1},a_{i1})$, $v_{i2} = (a_{i1},a_{i2})$, and define $Q_i := \mathrm{conv}(\{v_{i0},v_{i1},v_{i2}\})$ for $i=1,2$. For $T=\{2\}$, we have

$$\phi(T)\big((u_1,f_1),(u_2,f_2)\big) := f_1\left(a_{20} + (a_{21}-a_{20})\Big(1-\frac{f_2-u_2}{a_{22}-a_{21}}\Big) + (a_{22}-a_{21})\Big(1-\frac{u_2-a_{20}}{a_{21}-a_{20}}\Big)\right),$$

where we have used (20) to derive the term after f_1 on the right hand side of the above expression. Using the above expression, $\phi(T)(v_{1j_1},v_{2j_2}) = a_{1j_1}a_{2(2-j_2)}$ for $j_1,j_2\in\{0,1,2\}$. It follows that $\phi(T)$ is submodular when restricted to vertices of Q since $a_{i0}\le a_{i1}\le a_{i2}$ for $i=1,2$. Let $\{\pi^1,\ldots,\pi^6\}$ be the set of movement sequences in $\mathbb{Z}^2$ from $(0,0)$ to $(2,2)$, where $\pi^1 := (1,1,2,2)$, $\pi^2 := (1,2,1,2)$, $\pi^3 := (1,2,2,1)$, $\pi^4 := (2,1,1,2)$, $\pi^5 := (2,1,2,1)$, and $\pi^6 := (2,2,1,1)$. The set of movement sequences defines the switched staircase triangulation $\big\{\Upsilon_{\pi^1}(T),\ldots,\Upsilon_{\pi^6}(T)\big\}$, where

$$\Upsilon_{\pi^i}(T) = \mathrm{conv}\left(\left\{(v_{1j_1},v_{2(2-j_2)}) \;\middle|\; (j_1,j_2) = (0,0) + \sum_{p=1}^{k} e_{\pi^i_p},\ k=0,\ldots,4\right\}\right) \qquad \text{for } i=1,\ldots,6,$$

(see Figure 3 for the grid representation of the triangulation). Since the bilinear term is obviously convex-extendable from $\mathrm{vert}(Q)$, it follows from Corollary 4 that the convex envelope of ϕ over Q is determined by the switched staircase triangulation $\big\{\Upsilon_{\pi^1}(T),\ldots,\Upsilon_{\pi^6}(T)\big\}$. Moreover, each function

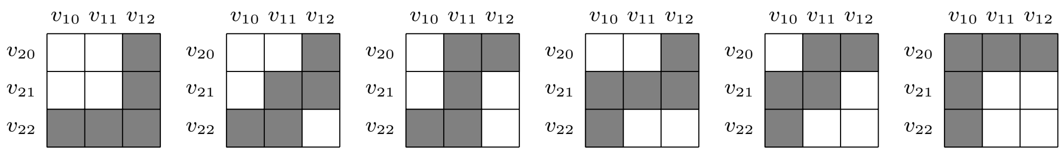


FIGURE 3. grid representation of switched staircase triangulation $\{\Upsilon_{\pi^1}(T),\dots,\Upsilon_{\pi^6}(T)\}$

e_i affinely interpolates f_1f_2 over the extreme points of simplex $\Upsilon_{\pi^i}(T)\big)$. As such, each inequality $\phi \ge e_i$, where $i\in\{1,\dots,6\}$, describes a non-vertical facet of the epigraph of $\text{conv}_Q(\phi)(s)$, and the result follows from Corollary 3.

A similar argument can be used to show that, for each $i\in\{1,\dots,6\}$, $\phi\le r_i$ defines a non-vertical facet-defining of the hypograph of $\text{conc}_P(\bar\phi)(u)$. This case does not require switching since the bilinear term is already supermodular. □

In the next example, we illustrate how Corollaries 3 and 4 can be used to derive tighter convex relaxations than the one in Corollary 5 by using additional underestimators.

EXAMPLE 2. Consider the function $x_1^2x_2^2$ over $[0,2]^2$. Here, we treat the bilinear term $\phi(f_1,f_2)=f_1f_2$ as the outer-function and x_1^2 and x_2^2 as inner-functions. For $i=1,2$, let $u_i(x)=\big(0,\ 2(2-\sqrt{3})x_i-(2-\sqrt{3})^2,\ 2x_i-1, x_i^2\big)$ and $a_i=(0,1,3,4)$. Notice that, using one single estimator $2x_i-1$ of x_i^2, [17] derived the following convex underestimator for $x_1^2x_2^2$

$$\max\left\{\begin{aligned}4x_1^2+4x_2^2-16,\ 2x_1+2x_2+3x_1^2+3x_2^2-17,\ 8x_1+3x_2^2-16\\ 3x_1^2+8x_2-16,\ 6x_1+6x_2-15,\ 0\end{aligned}\right\}, \tag{21}$$

which is obtained by substituting $2x_i-1$ and x_i^2 with their defining relation in inequalities from Corollary 5. Next, we illustrate how estimators $u_i(\cdot)$ and their bounds a_i are simultaneously exploited to relax $x_1^2x_2^2$ over $[0,2]^2$. Let $T=\{2\}$. By Corollaries 3 and 4, the supermodularity

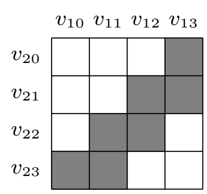


FIGURE 4. the switched simplex associated with $(1,2,1,2,1,2)$

of $\phi(T)(\cdot)$ over vertices of Q implies that interpolating $\bar\phi(\cdot)$ over the *switched* simplex given by $(1,2,1,2,1,2)$ (see Figure 4) yields $\bar\phi(u)\ge w_{14}=u_{11}+2u_{12}+u_{13}+u_{21}+2u_{22}+u_{23}-11$ for $u\in P$. After substitution, we obtain that $x_1^2x_2^2\ge x_1^2+(8-2\sqrt{3})x_1+x_2^2+(8-2\sqrt{3})x_2-15-2(2-\sqrt{3})^2$ for $x\in[0,2]^2$, which utilizes all available underestimators $u_i(\cdot)$ of inner-functions and their upper bounds on $[0,2]^2$. The underestimator w_{14} is not dominated by (21) because it evaluates to a higher value at $(x_1,x_2)=(1.63,1.4)$ than any underestimator in (21). For completeness, we include the full description of the convex and concave envelope of $\bar\phi(\cdot)$ over P in Appendix B. □

Observe that (21) is equivalent after substitution to underestimators w_1, w_{11}, w_{15}, w_{16}, w_{19}, and w_{20} from Appendix B, and these are precisely the inequalities that depend only on u_{i2} and u_{i3}. In fact, all the inequalities except w_7, w_8, w_9, w_{12}, w_{13}, and w_{14} can be obtained using one underestimator each for x_1^2 and x_2^2. Interestingly, every inequality obtained by choosing such an underestimator is still facet-defining for the set with two underestimators, as can be checked using the listing in Appendix B. Next, we formalize and generalize this observation. Recall that P is defined as the product of polytopes defined in (2), each of which depends on underestimators for an inner-function.

Assume that we construct another polytope P' using a subset of the underestimators used to define P. Then, since P' is a projection of P and projection commutes with convexification, it follows readily that that $\text{conc}_P(\bar{\phi})$ projects to $\text{conc}_{P'}(\bar{\phi})$. Therefore, any inequalities valid for $\text{conc}_{P'}(\bar{\phi})$ are also valid for $\text{conc}_P(\bar{\phi})$. However, we will show that a stronger property holds. The facet-defining inequalities for $\text{conc}_{P'}(\phi)$ also define facets of $\text{conc}_P(\bar{\phi})$. Towards this end, we introduce some notation to describe a projection of P obtained by selecting a subset of underestimators. This subset is specified using a d-tuple $J = (J_1,\dots,J_d) \in \mathcal{J}$, where $\mathcal{J}$ is defined as in (5). Here, each tuple specifies which underestimators are selected. We denote the corresponding projection of P as P_J which is now the Cartesian product $P_{1J_1} \times \cdots \times P_{dJ_d}$, where P_{iJ_i} is the polytope defined in (2) using underestimators u_{iJ_i} with a vector of bounds a_{iJ_i}.

Proposition 5. *Assume* $\text{conc}_P(\bar{\phi})$ *is a polyhedral function and* $\phi \le \langle \alpha_J, u_J\rangle + b$ *is a facet-defining inequality of* $\text{conc}_{P_J}(\bar{\phi})(u_J)$. *Then,* $\phi \le \langle \alpha, u\rangle + b$ *is a facet-defining inequality of* $\text{conc}_P(\bar{\phi})(\cdot)$, *where* $\alpha := (\alpha_J, 0)$.

Proof. Since $\phi \le \langle \alpha_J, u_J\rangle + b$ is a facet-defining inequality of $\text{conc}_{P_J}(\bar{\phi})(\cdot)$ and $a_{i0} < a_{in}$, there is a point $\bar{u}_J \in P_J$ with $\text{conc}_{P_J}(\bar{\phi})(\bar{u}_J) = \langle \alpha_J, \bar{u}_J\rangle + b$ such that, for $i \in \{1,\dots,d\}$ and $j \neq 0$, $a_{i0} < \bar{u}_{ij}$. Let $\bar{u} = A(\bar{u}_J)$, where A is the affine transformation defined in (18). Then,

$$\langle \alpha, \bar{u}\rangle + b \le \text{conv}_P(\bar{\phi})(\bar{u}) \le \text{conc}_{A(P_J)}(\bar{\phi})(\bar{u}) = \text{conc}_{P_J}(\bar{\phi})(\bar{u}_J) = \langle \alpha, \bar{u}\rangle + b,$$

where the first inequality holds by the validity of $\langle \alpha, u\rangle + b$ and $\bar{u} \in P$, the second inequality follows from $A(P_J) \subseteq P$, the first equality holds because affine maps commute with convexification, and the last equality follows by the choice of $\bar{u}_J$ and the definition of α. Therefore, equalities hold throughout and, in particular, $\langle \alpha, \bar{u}\rangle + b = \text{conc}_P(\bar{\phi})(\bar{u})$, that is, $\bar{u} \in T_{\bar{\phi}}^{(\alpha,b)}(P)$. Now, let $i' \in \{1,\dots,d\}$ and $j' \notin J_{i'}$, and consider the face $P' := \{u \in P \mid u_{i'j'} = a_{i0}\}$. By definition, $\bar{u}_{i'j'} \neq a_{i0}$. Thus, we can construct $\hat{u} \in P'$ that matches $\bar{u}$ except that $\hat{u}_{i'j'} \neq \bar{u}_{i'j'}$. Such a $\hat{u}$ can be obtained using the same argument as above where $\hat{u} = B(\bar{u}_J)$ and B is defined similarly to A, except that $\hat{u}_{i'j'} = a_{i'0}$. It follows that $\hat{u} \in T_{\bar{\phi}}^{(\alpha,b)}(P)$ and, therefore, $e_{i'j'}$ is in the affine hull of $T_{\bar{\phi}}^{(\alpha,b)}(P)$. Then,

$$\dim(P) \ge \dim\big(T_{\bar{\phi}}^{(\alpha,b)}(P)\big) \ge \dim\big(T_{\bar{\phi}}^{(\alpha_J,b)}(P_J)\big) + \sum_{i=1}^{d}\big(n - |J_i|\big) = \dim(P_J) + \sum_{i=1}^{d}\big(n - |J_i|\big) = \dim(P),$$

where the first inequality is because $T_{\bar{\phi}}^{(\alpha,b)}(P) \subseteq P$, second inequality is because in our argument above the choice of $\bar{u}_J$ was arbitrary in $T_{\bar{\phi}}^{(\alpha_J,b)}(P_J)$ and the choice of (i',j') was arbitrary except that $j' \notin J_{i'}$, the first equality is because (α_J, b) defines a facet of $\text{conc}_{P_J}(\bar{\phi})$, and the second equality is by the definition of P. Therefore, equalities holds throughout and, in particular, $\dim(P) = \dim\big(T_{\bar{\phi}}^{(\alpha,b)}(P)\big)$. $\square$

3.2. On the strength of termwise relaxation of bilinear functions In this subsection, we consider a weighted graph $G = (V,E)$ with node set $V = \{1,\dots,d\}$ and edge set E. With this graph, we associate a bilinear function ϕ: $\mathbb{R}^d \to \mathbb{R}$ defined as $\phi(s_{1n},\dots,s_{dn}) = \sum_{e\in E} c_e \prod_{i\in e} s_{in}$, where, by $i \in e$, we mean that edge e is incident with node i. We assume that an edge exists only if the corresponding weight $c_e \neq 0$. We call an edge positive if $c_e > 0$ and negative if $c_e < 0$. Let $\phi^Q := \big\{(s,\phi) \bigm| \phi = \bar{\phi}(s_1,\dots,s_d), s \in Q\big\}$, and we will study whether its convex hull

$$\text{conv}\big(\phi^Q\big) = \big\{(s,\phi) \bigm| \text{conv}_Q(\bar{\phi})(s) \le \phi \le \text{conc}_Q(\bar{\phi})(s), s \in Q\big\}$$

is obtained by a simple relaxation, one obtained by convexifying each bilinear term separately. To formally define the latter relaxation, we associate with an edge $e \in E$ a bilinear function $\phi_e : \mathbb{R}^d \to \mathbb{R}$ defined as $\phi_e(s_{1n},\dots,s_{dn}) = c_e \prod_{i\in e} s_{in}$. Then, we construct the *termwise-relaxation* of

ϕ^Q by underestimating $\bar{\phi}(\cdot)$ with $\sum_{e\in E}\operatorname{conv}_Q(\bar{\phi}_e)(\cdot)$ and overestimating it with $\sum_{e\in E}\operatorname{conc}_Q(\bar{\phi}_e)(\cdot)$, where each term in the summation could be obtained by using Corollary 4, or more specifically, when there is one non-trivial underestimator for each inner function, using Corollary 5. Succinctly, the termwise relaxation is defined as follow:

$$\Psi := \Big\{(s,\phi)\,\Big|\, \sum_{e\in E}\operatorname{conv}_Q(\bar{\phi}_e)(s) \le \phi \le \sum_{e\in E}\operatorname{conc}_Q(\bar{\phi}_e)(s), s\in Q\Big\}.$$

Clearly, Ψ is convex superset of ϕ^Q and therefore also a superset of $\operatorname{conv}\big(\phi^Q\big)$. We show that, if the graph G satisfies some conditions, $\operatorname{conv}(\phi^Q)$ coincides with Ψ. Since the sign for all c_e can be reversed, it suffices to consider the equivalence $\sum_{e\in E}\operatorname{conc}_Q(\bar{\phi}_e)(s) = \operatorname{conc}_Q(\bar{\phi})(s)$.

We call an edge $e\in E$ is positive if $c_e > 0$ and negative otherwise. A (signed) graph is said to be *balanced* if every cycle has an even number of negative edges (see [16]). It is shown in Theorem 3 of [16] that a graph is balanced if and only if the vertex set $V(G)$ can be partitioned into subsets T_1 and T_2 so that each positive edge of G connects two nodes from the same subset and each negative edge connects two nodes from different subsets. We will argue, by switching the variables which correspond to one of the partitioned subsets, that we can transform $\phi(\cdot)$ into a supermodular function.

Lemma 4. *Consider a graph G and let $\phi(\cdot)$ be a bilinear function defined by the graph G. There exists a subset T of V so that, for $s(T)$ as defined in (20), the function $\phi(T)(s_1,\dots,s_d) = \phi(s(T)_{1n},\dots,s(T)_{dn})$ is supermodular when restricted to* $\operatorname{vert}(Q)$ *if and only if graph G is balanced.*

Proof. Assume $G=(V,E)$ is a balanced graph. Then, using Theorem 3 in [16], we partition V into subsets T_1 and T_2 such that positively signed edges connect nodes of the same subset and the negatively signed edges connect nodes of the different subsets. Then, to show that $\phi(T_1)(\cdot)$ is supermodular over $\operatorname{vert}(Q)$, it suffices to show that, for each edge e, $\phi_e(T_1)(\cdot)$ is supermodular over $\operatorname{vert}(Q)$. By (20) it follows that, for $i\in T_1$, $s_i \ge s'_i$ and $s(T_1)_i \le s'(T_1)_i$ whenever $z_i \ge z'_i$. Since $\phi_e(\cdot)$ is supermodular when $e\in T_1$ or $e \in T_2$, and submodular otherwise, it follows that, for each edge e, $\phi_e(T_1)(\cdot)$ is supermodular.

We now show the converse, *i.e.*, there does not exist a T such that $\phi(T)(\cdot)$ is supermodular when restricted to $\operatorname{vert}(Q)$. Since the graph is not balanced, there exists a cycle that contains an odd number of negative edges. Since this cycle leaves and enters T an even number of times, it follows that there is a negative edge either contained in T or in its complement. Let this edge be $e:=(k,l)$. Assume without loss of generality that $k,l\in T$ as the other case is similar. Consider vertices v' and v'' corresponding to grid points $(j'_i)_{i=1}^d$ and $(j''_i)_{i=1}^d$, where we assume that $j'_k = j''_l = 1$, $j'_l = j''_k = 2$, and $j'_i = j''_i$ otherwise. Then, it follows that

$$\begin{aligned}&\phi(T)(v'\vee v'') + \phi(T)(v'\wedge v'') - \phi(T)(v') - \phi(T)(v'')\\ &= c_e\big(a_{kn-2}a_{ln-2} + a_{kn-1}a_{ln-1} - a_{kn-2}a_{ln-1} + a_{kn-1}a_{ln-2}\big) < 0,\end{aligned}$$

where the inequality follows from the supermodularity of the bilinear product and $c_e < 0$. Therefore, it follows $\phi(T)(\cdot)$ is not supermodular. $\square$

In Theorem 3, we show that the balanced graphs are exactly the ones for which the termwise relaxation Ψ coincides with $\operatorname{conv}(\phi^Q)$. To prove this result, we need the following lemma. Recall that we say that the concave envelope of a function f is determined by a triangulation $\mathcal{K} = \{K_1,\dots,K_r\}$ if the concave envelope of f over $\bigcup_{i=1}^r K_i$ is $\min_{i=1}^r \chi^{K_i}(s)$, where χ^{K_i} is the affine function interpolating $\big(v, f(v)\big)$ for all $v\in\operatorname{vert}(K_i)$.

Lemma 5. *Consider a function $f:\operatorname{vert}(D)\to\mathbb{R}$ so that $f(s) = \sum_{j=1}^m f_j(s)$, where D is a polytope. If concave envelopes of $f_j(s)$, $j=1,\dots,m$, are determined by the same triangulation $\mathcal{K}$ of D then $\operatorname{conc}_D(f)(s) = \sum_{j=1}^m \operatorname{conc}_D(f_j)(s)$ for every $s\in D$. Moreover, if there does not exist a common triangulation which generates concave envelopes of f_j for all j, then there exists $s\in D$ such that $\operatorname{conc}_D(f)(s) < \sum_{j=1}^m \operatorname{conc}_D(f_j)(s)$.*

Lemma 5 follows as a special case from the proof of Corollary 3.9 in [38]. We include a direct proof of Lemma 5 in Appendix C for completeness. We remark that this result is also related to prior results on sum-decomposability of (concave) envelopes in [30, 37].

THEOREM 3. *Consider a graph G and a bilinear function $\phi(\cdot)$ defined on G. Then, $\sum_{e\in E}\operatorname{conc}_Q(\bar{\phi}_e)(s)=\operatorname{conc}_Q(\bar{\phi})(s)$ for every point $s\in Q$ if and only if G is balanced.*

Proof. Suppose that graph G is balanced. Then, we show that $\sum_{e\in E}\operatorname{conc}_Q(\bar{\phi}_e)(s)=\operatorname{conc}_Q(\bar{\phi})(s)$. By Lemma 4, there exists a subset T of V such that, for all $e\in E$, $\phi_e(T)(s_1,\dots,s_d)$ is supermodular when restricted to $\operatorname{vert}(Q)$. By Theorem 1.2 in [30], for all $e\in E$, $\bar{\phi}_e(\cdot)$ is concave-extendable from $\operatorname{vert}(Q)$. By Corollary 4, $\operatorname{conc}_Q(\bar{\phi}_e)(\cdot)$ is determined by the same switched staircase triangulation for all $e\in E$. So, by Lemma 5, we conclude that $\sum_{e\in E}\operatorname{conc}_Q(\bar{\phi}_e)(s)=\operatorname{conc}_Q(\bar{\phi})(s)$.

Now, suppose that G is not balanced. We construct a point $s\in Q$ so that $\operatorname{conc}_Q(\bar{\phi})(s)<\sum_{e\in E}\operatorname{conc}_Q(\bar{\phi}_e)(s)$. Let $\bar{s}_i=\frac{1}{2}(0,a_{i1},\dots,a_{in-1},1)$ for all $i=1,\dots,d$. Then, we obtain

$$\operatorname{conc}_Q(\bar{\phi})(\bar{s}_1,\dots,\bar{s}_d)=\operatorname{conc}_{[0,1]^d}(\phi)\Big(\frac{1}{2},\dots,\frac{1}{2}\Big)<\sum_{e\in E}\operatorname{conc}_{[0,1]^2}(\phi_e)\Big(\frac{1}{2},\frac{1}{2}\Big)=\sum_{e\in E}\operatorname{conc}_Q(\bar{\phi}_e)(\bar{s}),$$

where first and last equality hold by Lemma 3 and strict inequality follows from Theorem 4 in [6]. (Alternately, the existence of a point that satisfies the strict inequality follows from Lemma 5, and that $\left(\frac{1}{2},\dots,\frac{1}{2}\right)$ is such a point is a consequence of strict supermodularity of the bilinear term). ☐

The hypercube $[0,1]^d$ arises as a special case of Q where $n=1$ and the variables $(s_{10},\dots,s_{d0})$ are projected out. In this case, Theorem 3 recovers the results of [6] and [26] regarding when McCormick envelopes [23] applied termwise suffice to obtain the concave envelope of a bilinear function ϕ over $[0,1]^d$.

COROLLARY 6 **(Theorem 3.10 in [26] and Theorem 4 in [6])**. *Consider the graph G associated with a bilinear function $\phi:[0,1]^d\to\mathbb{R}$. Then, the termwise relaxation of the hypograph of $\phi(\cdot)$ over $[0,1]^d$ coincides with the hypograph of $\operatorname{conc}_{[0,1]^d}(\phi)(\cdot)$ if and only if every cycle in G has an even number of negative edges.*

3.3. Tractable simultaneous convex hull We now extend our results to simultaneous convexification of a vector of functions $\theta\colon\mathbb{R}^d\to\mathbb{R}^\kappa$. Consider the hypograph of $\theta\colon\mathbb{R}^d\to\mathbb{R}^\kappa$ over a polytope $P:=P_1\times\cdots\times P_d$ defined as

$$\Theta^P:=\big\{(u,\theta)\in\mathbb{R}^{d\times(n+1)}\times\mathbb{R}^\kappa\bigm|\theta\le\theta(u_{1n},\dots,u_{dn}),\ u\in P\big\},$$

where P_i is the polytope defined in (2). For $k\in\{1,\dots,\kappa\}$, let $\Theta^P_k:=\big\{(u,\theta)\bigm|\theta_k\le\theta_k(u_{1n},\dots,u_{dn}),\ u\in P\big\}$ be the hypograph of $\theta_k(\cdot)$ over P. Since $\Theta^P\subseteq\bigcap_{k=1}^\kappa\operatorname{conv}(\Theta^P_k)$, it follows that $\operatorname{conv}(\Theta^P)$ is a subset of $\bigcap_{k=1}^\kappa\operatorname{conv}(\Theta^P_k)$, where the former will be referred to as the *simultaneous* convex hull of Θ^P, while the latter as the *individual* convex hull of Θ^P. Clearly, it is often the case that $\operatorname{conv}(\Theta^P)\subsetneq\bigcap_{k=1}^\kappa\operatorname{conv}(\Theta^P_k)$. Nevertheless, we will characterize conditions for which the simultaneous hull of Θ^P coincides with the individual hull of Θ^P.

THEOREM 4. *If concave envelopes of θ_k, $k=1,\dots,\kappa$, over Q are determined by the same triangulation $\mathcal{K}$ then $\operatorname{conv}(\Theta^P)=\bigcap_{k=1}^\kappa\operatorname{conv}(\Theta^P_k)$.*

Proof. Clearly, $\operatorname{conv}\big(\Theta^P\big)\subseteq\bigcap_{k=1}^\kappa\operatorname{conv}\big(\Theta^P_k\big)$ because $\Theta^P\subseteq\bigcap_{k=1}^\kappa\operatorname{conv}\big(\Theta^P_k\big)$ and the latter set is convex. To show $\bigcap_{k=1}^\kappa\operatorname{conv}\big(\Theta^P_k\big)\subseteq\operatorname{conv}\big(\Theta^P\big)$, we consider a point $(\bar{u},\bar{\theta})\in\bigcap_{k=1}^\kappa\operatorname{conv}\big(\Theta^P_k\big)$, *i.e.*, $(\bar{\theta}_1,\dots,\bar{\theta}_d)\le\big(\operatorname{conc}_P(\theta_1)(\bar{u}),\dots,\operatorname{conc}_P(\theta_d)(\bar{u})\big)$. Then, we lift $\bar{u}$ to the unique point $\bar{s}$ so that $(\bar{u},\bar{s})\in PQ'$, and define $J=(J_1,\dots,J_d)$, where $J_i:=\{j\mid\bar{u}_{ij}=\bar{s}_{ij}\}$. Let $D=(D_1,\dots,D_d)$ so that $D_i\subseteq\bar{J}_i:=\{0,\dots,n\}\setminus J_i$. Let $\bar{u}^D$ be a point of P so that, for $i\in\{1,\dots,d\}$, $\bar{u}^D_{ij}=\bar{s}_{ij}$ if $j\notin D_i$ and $\bar{u}^D_{ij}=\bar{u}_{ij}$

otherwise. Since $\bar{u}^{\bar{J}} = \bar{u}$ where $\bar{J}$ denotes $(\bar{J}_1, \dots, \bar{J}_d)$, the proof is complete if we show by induction on $|D| := \sum_{i=1}^{d} |D_i|$ that there is a set $M_D \subseteq \operatorname{vert}(P)$ and λ_v, for $v \in M_D$, that are independent of k and satisfy

$$\text{for } k \in \{1,\dots,\kappa\}, \quad \big(\bar{u}^D, \operatorname{conc}_P(\theta_k)(\bar{u}^D)\big) = \sum_{v \in M_D} \lambda_v\big(v, \theta_k(v_{\cdot n})\big), \quad \sum_{v \in M_D} \lambda_v = 1, \quad \lambda \geq 0, \tag{22}$$

where $v_{\cdot n}$ denotes $(v_{1n}, \dots, v_{dn})$. For the base case with $|D| = 0$, we have $\bar{u}^D = \bar{s} \in Q$. Since $\mathcal{K}$ is assumed to be the common triangulation that determines $\operatorname{conc}_Q(\theta_k)$ for every $k \in \{1,\dots,\kappa\}$, there exists $K \in \mathcal{K}$ and convex multipliers λ_v such that, for $k \in \{1,\dots,\kappa\}$, $\big(\bar{u}^D, \operatorname{conc}_Q(\theta_k)(\bar{u}^D)\big) = \sum_{v \in \operatorname{vert}(K)} \lambda_v\big(v, \theta_k(v_{\cdot n})\big)$. Therefore, the base case is established because $\operatorname{vert}(K) \subseteq \operatorname{vert}(Q) \subseteq \operatorname{vert}(P)$, and, by Corollary 1 in [17], we have that, for every $k \in \{1,\dots,\kappa\}$, $\operatorname{conc}_P(\theta_k)(\bar{u}^D) = \operatorname{conc}_Q(\theta_k)(\bar{u}^D)$. For the inductive step, consider $D = (D_1, \dots, D_d)$ so that $D_i \subseteq \bar{J}_i$ and assume that the result holds for any tuple D' such that $|D'| < |D|$. Since $|D| \neq 0$, there is a pair (i', j') so that $j' \in D_{i'}$. Let $D' := (D'_1, \dots, D'_d)$ so that $D'_i = D_i$ if $i \neq i'$ and $D'_{i'} = D_{i'} \setminus \{j'\}$. Note that from the induction hypothesis there exists a set $M_{D'} \subseteq \operatorname{vert}(P)$ and convex multipliers λ_v, one for each $v \in M_{D'}$, so that, for $k \in \{1,\dots,\kappa\}$, $\big(\bar{u}^{D'}, \operatorname{conc}_P(\theta_k)(\bar{u}^{D'})\big) = \sum_{v \in M^{D'}} \lambda_v\big(v, \theta_k(v_{\cdot n})\big)$. Consider an affine mapping A so that, for $i \neq i'$ and $j \neq j'$, $A(u)_{ij} = u_{ij}$ while $A(u)_{i'j'} = a_{i'0}$. Define $\gamma := \frac{u_{i'j'} - a_{i'0}}{s_{i'j'} - a_{i'0}}$.. Then, for $k \in \{1,\dots,\kappa\}$,

$$\begin{aligned}
\big(\bar{u}^D, \operatorname{conc}_P(\theta_k)(\bar{u}^D)\big) &= \gamma\big(\bar{u}^{D'}, \operatorname{conc}_P(\theta_k)(\bar{u}^{D'})\big) + (1-\gamma)\big(A(\bar{u}^{D'}), \operatorname{conc}_P(\theta_k)(\bar{u}^{D'})\big) \\
&= \gamma\Bigg(\sum_{v \in M_{D'}} \lambda_v\big(v, \theta_k(v_{\cdot n})\big)\Bigg) + (1-\gamma)\Bigg(\sum_{v \in M_{D'}} \lambda_v\big(A(v), \theta_k(v_{\cdot n})\big)\Bigg) \\
&= \gamma\Bigg(\sum_{v \in M_{D'}} \lambda_v\big(v, \theta_k(v_{\cdot n})\big)\Bigg) + (1-\gamma)\Bigg(\sum_{v \in M_{D'}} \lambda_v\Big(A(v), \theta_k\big(A(v)_{\cdot n}\big)\Big)\Bigg),
\end{aligned}$$

where the first equality holds because, by Corollary 1 in [17], $\operatorname{conc}_P(\theta_k)(\bar{u}^D) = \operatorname{conc}_P(\theta_k)(\bar{u}^{D'})$, the second equality follows from the induction hypothesis because $|D'| < |D|$ and exploits that convexification commutes with affine transformation, and the last equality is because $j' \neq n$. The induction step is established by observing that, for any $u \in \operatorname{vert}(P)$, $A(u) \in \operatorname{vert}(P)$. $\square$

Corollary 7. *Assume that, for every $k \in \{1,\dots,\kappa\}$, the extension $\bar{\theta}_k(\cdot)$ of $\theta_k(\cdot)$ is concave-extendable from $\operatorname{vert}(Q)$ and is supermodular when restricted to the vertices of Q. Then, $\operatorname{conv}\big(\Theta^P\big) = \bigcap_{k=1}^{\kappa} \operatorname{conv}\big(\Theta^P_k\big)$. Moreover, the facet generation problem of $\operatorname{conv}(\Theta^P)$ can be solved in $\mathcal{O}(\kappa d n \log d)$.*

Proof. It follows from Theorem 4 that $\operatorname{conv}\big(\Theta^P\big) = \bigcap_{k=1}^{\kappa} \operatorname{conv}\big(\Theta^P_k\big)$ because, by Proposition 4, concave envelopes of θ_k, $k = 1, \dots, \kappa$, are determined by a triangulation of Q, that does not depend on k. Now, we argue that the facet generation problem of $\operatorname{conv}(\Theta^P)$ can be solved by separating $\operatorname{conv}(\Theta^P_k)$ individually. Let $(\bar{u}, \bar{\theta}) \in \mathbb{R}^{d \times (n+1)} \times \mathbb{R}^{\kappa}$. Then, for each $k \in \{1,\dots,\kappa\}$, we call the procedure in Theorem 2 to solve the facet generation problem of $\operatorname{conv}(\Theta^P_k)$. If $(\bar{u}, \bar{\theta}) \in \bigcap_{k=1}^{\kappa} \operatorname{conv}(\Theta^P_k)$ then, as shown above, $(\bar{u}, \bar{\theta}) \in \operatorname{conv}(\Theta^P)$. Otherwise, without loss of generality, we assume that $(\bar{u}, \bar{\theta}) \notin \operatorname{conv}(\Theta^P_1)$ and the procedure outputs a facet-defining inequality $\theta_1 \leq \langle \alpha, u \rangle + b$ of $\operatorname{conc}_P(\theta_1)$ that is violated by $(\bar{u}, \bar{\theta})$. As $\operatorname{conv}(\Theta^P) \subseteq \bigcap_{k=1}^{\kappa} \operatorname{conv}(\Theta^P_k)$, this inequality is valid for $\operatorname{conv}(\Theta^P)$. To complete the proof, we will show that it defines a facet of $\operatorname{conv}(\Theta^P)$. Let $T := \big\{(u, \theta) \in \operatorname{conv}(\Theta^P) \bigm| \theta_1 = \langle \alpha, u \rangle + b\big\}$ and let $M := \big\{u \in \operatorname{vert}(P) \bigm| \langle \alpha, u \rangle + b = \theta_1(u_{\cdot n})\big\}$. Observe that $\big\{(u, \theta) \bigm| u \in M,\ \theta_1 = \theta_1(u_{\cdot n}),\ \theta_k \geq \theta_k(u_{\cdot n}) \text{ for } k = 2, \dots, \kappa\big\}$ is a subset of T. Therefore, $\dim\big(\operatorname{aff}(T)\big) \geq \dim(M) + \kappa - 1 = \dim(P) + \kappa - 1 \geq \dim(\Theta^P) - 1$, where the equality holds because $\theta_1 \leq \langle \alpha, u \rangle + b$ defines a facet of Θ_1. Thus, T is a facet of $\operatorname{conv}(\Theta^P)$. $\square$

4. Infinitely many estimators for inner functions Sections 2 to 3 considered composite functions and described a way to relax them while exploiting *finitely* many estimators for each inner-function. A natural follow-up question is to understand the limiting relaxation, one obtained using *infinitely* many estimators for each inner-function. We will explore the structure of this relaxation in Section 4. To begin, we review some basic concepts from probability theory, optimal transport, and stochastic order that we later use to characterize the structure of the limiting relaxation.

4.1. Probability, optimal transport, and stochastic orders Each real-valued random variable A_i induces a probability measure on the real line $(-\infty,\infty)$ which can be described by its (cumulative) distribution function F_i, that is, $F_i(a_i) = \Pr\{A_i \le a_i\}$ for $a_i \in (-\infty,\infty)$. The expectation of random variable A_i is

$$\mathbb{E}[A_i] = \int_{-\infty}^{\infty} a_i \mathrm{d}F_i(a_i).$$

Any distribution function F_i has three properties; it is non-decreasing, right-continuous, and ranges from 0 to 1 with $\lim_{a_i\to-\infty} F_i(a_i) = 0$ and $\lim_{a_i\to\infty} F_i(a_i) = 1$. Conversely, any function satisfying these three properties is a distribution function for some random variable. The right-continuity of a non-decreasing function implies that the function is continuous except possibly at a finite or countable set of points where the graph of the distribution function has a vertical gap. Due to the vertical gaps, a distribution function F_i does not always have an inverse. To circumvent this issue, a generalized inverse is used instead that is defined for any $\lambda \in [0,1]$ as follows:

$$F_i^{-1}(\lambda) := \min\left\{a_i \;\middle|\; F_i(a_i) \ge \lambda\right\}.$$

The generalized inverse, $F_i^{-1}(\lambda)$, is non-decreasing and left-continuous on $[0,1]$. Like the distribution function F_i, the generalized inverse function, F_i^{-1}, can have at most countably many jumps where if it fails to be continuous. Observe that $F_i^{-1}(\lambda) \le a_i$ if and only if $\lambda \le F_i(a_i)$. This is because for any (a_i,λ) so that $F(a_i) \ge \lambda$, it follows by minimization and feasibility of a_i in the definition of $F_i^{-1}(\lambda)$ that $F_i^{-1}(\lambda) \le a_i$. Then, if $a_i' := F_i^{-1}(\lambda) \le a_i$, it follows that $F_i(a_i) \ge F_i(a_i') \ge \lambda$, where the first inequality is because $a_i \ge a_i'$ and the second is because F_i is right-continuous. In particular, we will consider real-valued random variables with support in a measurable subset of $[0,1]$. Let $\mathcal{F}$ be the set of all distribution functions with support in $[0,1]$. Then, $\mathcal{F}$ is a convex subset of $\mathcal{B}$, where the latter set denotes the convex cone whose elements are all bounded nonnegative univariate functions on $\mathbb{R}$. The convexity of $\mathcal{F}$ follows because the three properties characterizing functions in $\mathcal{F}$ are closed under convex combinations and any function satisfying these properties belongs to $\mathcal{F}$. The extreme set of $\mathcal{F}$, denoted as $\mathrm{ext}(\mathcal{F})$, is the set of distribution functions with Dirac measures over $[0,1]$, *i.e.*, $\mathrm{ext}(\mathcal{F}) := \left\{H_{\delta(a)} \;\middle|\; a \in [0,1]\right\}$, where $\delta(a)$ denotes the Dirac measure at point a and $H_{\delta(a)}$ denotes the corresponding distribution function, *i.e.*, $H_{\delta(a)}(x) = 0$ for $x < a$ and $H_{\delta(a)}(x) = 1$ for $x \ge a$.

For distribution functions $F_1,\dots,F_d$, we define $\Pi(F_1,\dots,F_d)$ as the set of all joint distribution functions on $\mathbb{R}^d$ whose marginals are $F_1,\dots,F_d$. Therefore, a distribution function F belongs to $\Pi(F_1,\dots,F_d)$ if and only if it satisfies the following properties. First, F is non-decreasing in each variable. Second, F is right-continuous in the sense that $\lim_{\delta\to0^+} F(a_1+\delta,\dots,a_d+\delta) = F(a_1,\dots,a_d)$. Third, $F(a_1,\dots,a_d) \to 0$ if $a_i \to -\infty$ for some i, and $F(a_1,\dots,a_d) = 1$ if $a_i \to \infty$ for all i. Finally, for each i and $a_i \in (-\infty,\infty)$, $F(\infty,\dots,a_i,\dots,\infty) = F_i(a_i)$. Let $\mathcal{B}^d$ denote the convex cone of bounded nonnegative functions on $\mathbb{R}^d$. Then, $\Pi(F_1,\dots,F_d)$ is a convex subset of $\mathcal{B}^d$ because the above properties are closed under taking convex combinations and any functions satisfying these properties belong to $\Pi(F_1,\dots,F_d)$. To clarify the joint distribution function, we add it as a subscript to the expectation operator so that, for a continuous function $\phi : \mathbb{R}^d \to \mathbb{R}$, its expectation under $F \in \Pi(F_1, \dots, F_d)$ is denoted as

$$\mathbb{E}_F\big[\phi(A_1, \dots, A_d)\big] := \int_{\mathbb{R}^d} \phi(a_1, \dots, a_d) \mathrm{d}F(a_1, \dots, a_d),$$

where $(A_1, \dots, A_d)$ follows F, which will be denoted as $(A_1, \dots, A_d) \sim F$. If the distribution functions $F_1, \dots, F_d$ are assumed to have supports in $[0,1]$ and $\phi(\cdot)$ is continuous on $[0,1]^d$, it follows that, for all $F \in \Pi(F_1, \dots, F_d)$, the expectation $\mathbb{E}_F\big[\phi(A_1, \dots, A_d)\big]$ is finite.

The limiting relaxation arises as the solution of an optimal transport problem. For distribution functions $F_1, \dots, F_d$ on the real line, the *multivariate Monge-Kantorovich* problem on the real line (Section 2 in [29]) is defined as the following optimization problem

$$\sup\Big\{\mathbb{E}_F\big[\phi(A_1, \dots, A_d)\big] \;\Big|\; F \in \Pi(F_1, \dots, F_d)\Big\}. \tag{23}$$

To express the limiting relaxation in an explicit form, we need to solve the multivariate Monge-Kantorovich problem.

THEOREM 5 (**[21] and Theorem 5 in [42]**). *Let $F_1, \dots, F_d$ be d probability distribution functions on the real line and let $F^*(a) = \min_i F_i(a_i)$. Then, for any continuous supermodular function $\phi : \mathbb{R}^d \to \mathbb{R}$,*

$$\sup_{F \in \Pi(F_1, \dots, F_d)} \int \phi \mathrm{d}F = \int \phi \mathrm{d}F^*$$

if $\phi \leq \varphi$ for some continuous function φ such that $\int \varphi \mathrm{d}F$ is finite and constant for all $F \in \Pi(F_1, \dots, F_d)$. $\square$

To establish the convexity of the limiting relaxation in the space of (x, f) variables, we will need to show that the optimal value of (23) changes monotonically as the univariate distribution functions $F_1, \dots, F_d$ vary in a specific manner. To this end, we review order relations over distribution functions. Let A_i and B_i two univariate random variables with distribution functions F_i and G_i, respectively. Then, A_i is said to be smaller than B_i in the *concave order* (denoted as $F_i \preceq G_i$) if $\mathbb{E}_{F_i}[\psi(A_i)] \leq \mathbb{E}_{G_i}[\psi(B_i)]$ for all concave functions $\psi : \mathbb{R} \to \mathbb{R}$, provided the expectations exist. The following two alternative characterizations from Theorem 3.A.1 and Theorem 3.A.5 in [34] will be useful in our context, that is, $F_i \preceq G_i$ if and only if

$$\mathbb{E}_{F_i}\big[\min\{A_i, a_i\}\big] \leq \mathbb{E}_{G_i}\big[\min\{B_i, a_i\}\big] \quad \text{for } a_i \in \mathbb{R} \qquad \text{and} \qquad \mathbb{E}_{F_i}[A_i] = \mathbb{E}_{G_i}[B_i], \tag{24}$$

or

$$\int_0^p F_i^{-1}(\lambda)\mathrm{d}\lambda \leq \int_0^p G_i^{-1}(\lambda)\mathrm{d}\lambda \quad \text{for } p \in [0,1] \qquad \text{and} \qquad \mathbb{E}_{F_i}[A_i] = \mathbb{E}_{G_i}[B_i]. \tag{25}$$

Next, we consider another order which is defined by dropping the second requirement in (24). Namely, we say $F_i \prec G_i$ if

$$\mathbb{E}_{F_i}\big[\min\{A_i, a_i\}\big] \leq \mathbb{E}_{G_i}\big[\min\{B_i, a_i\}\big] \quad \text{for } a_i \in \mathbb{R}, \tag{26}$$

This order is the *increasing concave order* of two random variables A_i and B_i, and is equivalently defined by requiring, $\mathbb{E}_{F_i}[\psi(A_i)] \leq \mathbb{E}_{G_i}[\psi(B_i)]$ for all increasing concave function $\psi : \mathbb{R} \to \mathbb{R}$ (see Theorem 4.A.2 in [34]). Moreover, Theorem 4.A.5 and (1.A.2) in [34] provide a useful alternative characterization, that is $F_i \prec G_i$ if and only if there exists a distribution function H_i such that

$$F_i^{-1}(\lambda) \leq H_i^{-1}(\lambda) \quad \text{for } \lambda \in [0,1] \qquad \text{and} \qquad H_i \preceq G_i. \tag{27}$$

To study how the optimal value of (23) changes as distribution functions $F_1, \dots, F_d$ change in (increasing) concave order, we need an integral inequality given in the following technical lemma.

LEMMA 6 **(Theorem 1 in [12])**. *Let $\psi(\lambda, u_1, \ldots, u_d)$ be a continuous function mapping from $[0,1]\times\mathbb{R}^d$ to $\mathbb{R}$. Then, we have*

$$\int_0^1 \psi(\lambda, \eta_1, \ldots, \eta_d)\mathrm{d}\lambda \le \int_0^1 \psi(\lambda, \gamma_1, \ldots, \gamma_d)\mathrm{d}\lambda$$

for each system of non-decreasing bounded univariate functions η_i, γ_i, $i=1,\ldots,d$, such that

$$\int_0^p \eta_i(\lambda)\mathrm{d}\lambda \ge \int_0^p \gamma_i(\lambda)\mathrm{d}\lambda \quad 0\le p\le 1 \qquad \text{and} \qquad \int_0^1 \eta_i(\lambda)\mathrm{d}\lambda = \int_0^1 \gamma_i(\lambda)\mathrm{d}\lambda, \tag{28}$$

if and only if the function ψ is convex in u_i when the other arguments are fixed, supermodular over $\mathbb{R}^d$ when λ is fixed, and

$$\int_0^\delta \Big(\psi(p+\delta+\lambda, u) - \psi(p+\delta+\lambda, u-he_i) + \psi(p+\lambda, u-he_i) - \psi(p+\lambda, u)\Big)\mathrm{d}\lambda \ge 0 \tag{29}$$

for all $0\le p\le 1-2\delta$, $\delta>0$, $h\ge 0$, $i=1,\ldots,d$, where e_i is the i^{th} standard basis vector in $\mathbb{R}^d$. □

4.2. Envelope characterization via optimal transport Consider a composite function $\phi\circ f : X\subseteq\mathbb{R}^m\to\mathbb{R}$ defined as $(\phi\circ f)(x)=\phi(f(x))$, where $f\colon \mathbb{R}^m\to\mathbb{R}^d$ is a vector of bounded functions over X and $\phi\colon \mathbb{R}^d\to\mathbb{R}$ is a continuous function. For each point (x,f), where $x\in X$ and $f=f(x)$, in Section 4.3 we will use underestimating functions of $f_i(x)$ to derive a marginal distribution function $F_i\in\mathcal{F}$. This marginal distribution will be such that the expected value of the corresponding random variable A_i, denoted as $\mathbb{E}_{F_i}[A_i]$, equals f_i. Consequently, to relax the hypograph of the composite function $\phi\circ f$, it will suffice to over-estimate $\phi\big(\mathbb{E}_{F_1}(A_1),\ldots,\mathbb{E}_{F_d}(A_d)\big)$. We now briefly discuss how this will be achieved. For notational brevity, we extend the outer-function $\phi(\cdot)$ to define $\tilde\phi(\cdot)$ so that, for any $(F_1,\ldots,F_d)\in\mathcal{F}^d := \mathcal{F}\times\cdots\times\mathcal{F}$,

$$\tilde\phi(F_1,\ldots,F_d) = \phi\big(\mathbb{E}_{F_1}[A_1],\ldots,\mathbb{E}_{F_d}[A_d]\big),$$

where $A_i\sim F_i$. To see the functional $\tilde\phi(\cdot)$ as an extension of $\phi(\cdot)$ from $[0,1]^d$ to $\mathcal{F}^d$, we map an $f\in[0,1]^d$ into $\mathcal{F}^d$ as $\big(H_{\delta(f_1)},\ldots,H_{\delta(f_d)}\big)$, where $H_{\delta(f_i)}$, as defined before, is the distribution function with its mass concentrated at the point f_i. In this subsection, we will derive the concave envelope of $\tilde\phi(\cdot)$ over its domain $\mathcal{F}^d$, that is the lowest concave overestimator of the extension $\tilde\phi(\cdot)$ over $\mathcal{F}^d$. This envelope will be denoted as $\mathrm{conc}_{\mathcal{F}^d}(\tilde\phi)$. More specifically, we show that when the outer-function $\phi(\cdot)$ satisfies certain conditions, $\mathrm{conc}_{\mathcal{F}^d}(\tilde\phi)$ is the solution to an optimal transport problem [45]. This solution can be derived explicitly when $\phi(\cdot)$ satisfies some additional requirements.

Before characterizing the concave envelope of $\tilde\phi(\cdot)$ over $\mathcal{F}^d$, we discuss how this setting relates to the discrete case. In Section 2 and 3, we introduced a sequence of mappings $(x,f)\to u\to s\to z$, where the first map evaluated underestimators, second map was defined by constructing two-dimensional concave envelopes, and the third map was via an affine transformation Z. It is the z-space that is intimately related to $\mathcal{F}^d$. More specifically, let $z=(z_1,\ldots,z_d)$, where $z_i=(z_{i0},\ldots,z_{in})\in\Delta_i$. Let F_i be defined as 1 at a_{in} and above, $1-z_{ik}$ in $[a_{ik-1},a_{ik})$ for $k\in\{1,\ldots,n\}$, and 0 below a_{i0}. Since the mapping from z to $(F_1,\ldots,F_d)$ is affine, results regarding the concave envelope over Δ translate to those about $\mathrm{conc}_{\mathcal{F}^d}(\tilde\phi)(F_1,\ldots,F_d)$. The following treatments will generalize the above discussion, allowing for more general distributions that are not necessarily supported at a finite set of discrete points.

If $\phi(\cdot)$ is a univariate convex function, by Jensen's inequality, $\tilde\phi(F_1)=\phi\big(\mathbb{E}_{F_1}[A_1]\big)\le\mathbb{E}_{F_1}\big[\phi(A_1)\big]$. We now extend this idea to the multidimensional case. We will show in Lemma 7 that, as long as,

$\phi(\cdot)$ is convex in each argument when other arguments are fixed, there exists a joint distribution $F \in \Pi(F_1,\dots,F_d)$ so that

$$\tilde{\phi}(F_1,\dots,F_d) \le \mathbb{E}_F[\phi(A_1,\dots,A_d)]. \tag{30}$$

The above inequality immediately implies that

$$\operatorname{conc}_{\mathcal{F}^d}\big(\tilde{\phi}\big)(F_1,\dots,F_d) = \operatorname{conc}_{\mathcal{F}^d}\big(\tilde{\phi}|_{\operatorname{ext}(\mathcal{F}^d)}\big)(F_1,\dots,F_d), \tag{31}$$

i.e., it suffices to restrict $\tilde{\phi}(\cdot)$ to the extreme points for $\mathcal{F}^d$ for the purpose of constructing $\operatorname{conc}_{\mathcal{F}^d}\big(\tilde{\phi}\big)$. To see this, observe that the left hand side in (31) is at least as large as the right hand side. We now argue the converse relationship. Observe that

$$\begin{aligned}\tilde{\phi}(F_1,\dots,F_d) \le \mathbb{E}_F\big[\phi(A_1,\dots,A_d)\big] &= \int \tilde{\phi}(H_{\delta(a_1)},\dots,H_{\delta(a_d)})\mathrm{d}F(a) \\ &\le \operatorname{conc}_{\mathcal{F}^d}\big(\tilde{\phi}|_{\operatorname{ext}(\mathcal{F}^d)}\big)(F_1,\dots,F_d),\end{aligned} \tag{32}$$

where the first inequality holds by the hypothesis (30), and the equality is by definition of $\tilde{\phi}(\cdot)$. The last inequality holds by the concavity of $\operatorname{conc}_{\mathcal{F}^d}(\tilde{\phi}|_{\operatorname{ext}(\mathcal{F}^d)})$ and that $H_{\delta(a_i)}$ are the extreme points of $\mathcal{F}$, since, for each $i \in \{1,\dots,d\}$, $F_i(a_i) = \int H_{\delta(b_i)}(a_i)\mathrm{d}F_i(b_i)$. Then, it follows that the converse $\operatorname{conc}_{\mathcal{F}^d}(\tilde{\phi})(F_1,\dots,F_d) \le \operatorname{conc}_{\mathcal{F}^d}(\tilde{\phi}|_{\operatorname{ext}(\mathcal{F}^d)})(F_1,\dots,F_d)$ holds. The relation (31) shows that $\operatorname{conc}_{\mathcal{F}^d}(\tilde{\phi})$ is the lowest concave extension of $\tilde{\phi}(\cdot)$ restricted to $\operatorname{ext}(\mathcal{F}^d)$. We now establish that the inequality (30) holds under certain hypothesis on the structure of ϕ.

Lemma 7. *Let $A_1,\dots,A_d$ be independent random variables and $F_1,\dots,F_d$ be the corresponding distribution functions. If $\phi:\mathbb{R}^d \to \mathbb{R}$ is a continuous function which is convex in each argument when other arguments are fixed then $\tilde{\phi}(F_1,\dots,F_d) \le \mathbb{E}\big[\phi(A_1,\dots,A_d)\big]$, where the equality is attained when the function $\phi(\cdot)$ is multilinear.*

Proof. For any index set I, we will denote the joint distribution of $\{A_i \mid i \in I\}$ as F_I. The set of integers $\{i,\dots,j\}$ will be denoted $[i,j]$ so that the joint distribution of $\{A_i,\dots,A_d\}$ will be written as $F_{[i,j]}$. We prove the inequality in the statement of the result by induction on d. The base case $d=1$ follows from Jensen's inequality as was remarked earlier. For the inductive step, we have:

$$\begin{aligned}\tilde{\phi}(F_1,\dots,F_d) &\le \mathbb{E}_{F_{[1,d-1]}}\Big[\phi\Big(A_1,\dots,A_{d-1},\mathbb{E}_{F_d}\big[A_d\big]\Big)\Big] \\ &= \mathbb{E}_{F_{[1,d-1]}}\Big[\phi\Big(A_1,\dots,A_{d-1},\mathbb{E}_{F_d}\big[A_d \mid A_1,\dots,A_{d-1}\big]\Big)\Big] \\ &= \mathbb{E}_{F_{[1,d-1]}}\Big[\phi\Big(\mathbb{E}_{F_d}\big[A_1,\dots,A_{d-1},A_d \mid A_1,\dots,A_{d-1}\big]\Big)\Big] \\ &\le \mathbb{E}_{F_{[1,d-1]}}\Big[\mathbb{E}_{F_d}\Big[\phi\big(A_1,\dots,A_d\big) \,\Big|\, A_1,\dots,A_{d-1}\Big]\Big] \\ &= \mathbb{E}_{F_{[1,d]}}\Big[\phi\big(A_1,\dots,A_d\big)\Big],\end{aligned}$$

where the first inequality is by induction hypothesis, the first equality is by the independence of A_i, the second equality is because $\mathbb{E}[A_i \mid A_i] = A_i$, the second inequality is due to Jensen's inequality, and the last equality holds because of the law of iterated expectations. The proof is complete by observing that each inequality becomes an equality if ϕ is linear when all but one of its arguments are fixed. $\square$

Next, we relate the right hand side of (31) to the Monge-Kantorovich problem. Consider a functional $\hat{\phi}:\mathcal{F}^d \to \mathbb{R}$ defined as follows:

$$\hat{\phi}(F_1,\dots,F_d) := \sup\Big\{\mathbb{E}_F\big[\phi(A_1,\dots,A_d)\big] \,\Big|\, F \in \Pi(F_1,\dots,F_d)\Big\} \qquad \text{for } (F_1,\dots,F_d) \in \mathcal{F}^d, \tag{33}$$

where we recall that $\Pi(F_1,\dots,F_d)$ denotes the set of joint distributions with $F_1,\dots,F_d$ as marginals. Since $\phi(\cdot)$ is assumed to be continuous, it follows from Theorem 2.3.10 in [29] that there exists an optimal solution to (33). Now, we argue that (33) is a reformulation for the right hand side of (31).

PROPOSITION 6. *For $(F_1,\dots,F_d)\in\mathcal{F}^d$, $\hat\phi(F_1,\dots,F_d)=\operatorname{conc}_{\mathcal{F}^d}\big(\tilde\phi|_{ext(\mathcal{F}^d)}\big)(F_1,\dots,F_d)$. Moreover, if $\phi(\cdot)$ is continuous and convex in each argument when other arguments are fixed, for $(F_1,\dots,F_d)\in\mathcal{F}^d$, $\hat\phi(F_1,\dots,F_d)=\operatorname{conc}_{\mathcal{F}^d}\big(\tilde\phi\big)(F_1,\dots,F_d)$.*

Proof. Let $(F_1,\dots,F_d)\in\mathcal{F}^d$. It follows trivially that $\hat\phi(F_1,\dots,F_d)\le\operatorname{conc}\big(\tilde\phi|_{\mathrm{ext}(\mathcal{F}^d)}\big)(F_1,\dots,F_d)$ because, as we argued in (32), for $F\in\Pi(F_1,\dots,F_d)$, $\mathbb{E}_F[\phi(A_1,\dots,A_d)]\le\operatorname{conc}\big(\tilde\phi|_{\mathrm{ext}(\mathcal{F}^d)}\big)(F_1,\dots,F_d)$. To prove the converse, it suffices to show that $\hat\phi(\cdot)$ is concave in $\mathcal{F}^d$ because, for $(F_1,\dots,F_d)\in\mathrm{ext}(\mathcal{F}^d)$, each $F_i=H_{\delta(a_i)}$ for some $a_i\in[0,1]$ and it follows by considering the multidimensional Dirac distribution at $(a_1,\dots,a_d)$ that $\hat\phi(F_1,\dots,F_d)\ge\tilde\phi(F_1,\dots,F_d)$. To see that $\hat\phi(\cdot)$ is concave, let $(F_1,\dots,F_d)$ and $(G_1,\dots,G_d)$ be two points in $\mathcal{F}^d$ and let α be chosen to satisfy $0\le\alpha\le1$. Then, for any $F\in\Pi(F_1,\dots,F_d)$ and $G\in\Pi(G_1,\dots,G_d)$, let $A\sim F$, $B\sim G$, and $C\sim\alpha F+(1-\alpha)G$. We have,

$$\begin{aligned}\hat\phi\big(\alpha(F_1,\dots,F_d)+(1-\alpha)(G_1,\dots,G_d)\big)&\ge\mathbb{E}_{(\alpha F+(1-\alpha)G)}\big[\phi(C_1,\dots,C_d)\big]\\&=\alpha\mathbb{E}_F\big[\phi(A_1,\dots,A_d)\big]+(1-\alpha)\mathbb{E}_G\big[\phi(B_1,\dots,B_d)\big],\end{aligned}$$

where the inequality holds because $\alpha F+(1-\alpha)G$ is a feasible solution to (33) at $\alpha(F_1,\dots,F_d)+(1-\alpha)(G_1,\dots,G_d)$, and the equality holds because expectation of a mixture distribution is the mixture of the expectations under distributions being mixed. Since the inequality holds for every (F,G) in $\Pi(F_1,\dots,F_d)\times\Pi(G_1,\dots,G_d)$, it also holds for the supremum of $\alpha\mathbb{E}_F\big[\phi(A_1,\dots,A_d)\big]+(1-\alpha)\mathbb{E}_G\big[\phi(B_1,\dots,B_d)\big]$ over $\Pi(F_1,\dots,F_d)\times\Pi(G_1,\dots,G_d)$. Therefore, $\hat\phi\big(\alpha(F_1,\dots,F_d)+(1-\alpha)(G_1,\dots,G_d)\big)\ge\alpha\hat\phi(F_1,\dots,F_d)+(1-\alpha)\hat\phi(G_1,\dots,G_d)$, showing the concavity of $\hat\phi(\cdot)$.

The second statement in the result follows from the first statement because Lemma 7 implies (31). ☐

As a result, when the inequality in (30) is satisfied, the functional $\hat\phi(\cdot)$ coincides with the lowest concave overestimator of $\tilde\phi(\cdot)$ over $\mathcal{F}^d$. Then, to compute $\hat\phi(\cdot)$ at a given $(F_1,\dots,F_d)$ in $\mathcal{F}^d$, we need to solve the multivariate Monge-Kantorovich problem. By Theorem 5, the latter problem has an explicit solution under certain conditions. We apply this result in our setting to obtain an explicit integral representation of the functional $\hat\phi(\cdot)$.

THEOREM 6. *If $\phi:[0,1]^d\to\mathbb{R}$ is a continuous supermodular function then for $(F_1,\dots,F_d)\in\mathcal{F}^d$*

$$\hat\phi(F_1,\dots,F_d)=\mathbb{E}_{F^*}\big[\phi(A_1,\dots,A_d)\big]=\int_0^1\phi\big(F_1^{-1}(\lambda),\dots,F_n^{-1}(\lambda)\big)\mathrm{d}\lambda,$$

where $F^*(a)=\min\big\{F_1(a_1),\dots,F_d(a_d)\big\}$.

Proof. Let $(F_1,\dots,F_d)\in\mathcal{F}^d$. Since $A_1,\dots,A_d$ have supports in $[0,1]$ and the function $\phi(\cdot)$ is continuous, $\mathbb{E}_F\big[\phi(A_1,\dots,A_d)\big]$ is finite for all $F\in\Pi(F_1,\dots,F_d)$. We choose $\varphi(a)$ to be a constant c defined as $\max_{a'\in[0,1]^d}\big|\phi(a')\big|$. By definition, $\phi(f)\le c$ and $\int c\mathrm{d}F=c$ is finite for all $F\in\Pi(F_1,\dots,F_d)$. It follows from Theorem 5 that we have $\hat\phi(F_1,\dots,F_d)=\mathbb{E}_{F^*}[\phi(A_1,\dots,A_d)]$. Now, consider a random variable U that is uniformly distributed over $[0,1]$ and observe that, for all $(a_1,\dots,a_d)\in\mathbb{R}^d$,

$$\Pr\big(F_1^{-1}(U)\le a_1,\dots,F_d^{-1}(U)\le a_d\big)=\Pr\big(U\le F_1(a_1),\dots,U\le F_d(a_d)\big)=\min\big\{F_1(a_1),\dots,F_d(a_d)\big\},$$

where the first equality is because $F_i^{-1}(U)\le a_i$ if and only if $U\le F_i(a_i)$ and the second equality is because U is uniformly distributed. In other words, the distribution function of the random vector

$\big(F_1^{-1}(U),\dots,F_d^{-1}(U)\big)$ is $F^*(a)$. We can assume that the range of ϕ is $[0,1]$ since it is bounded. Let $A^* = (A_1^*,\dots,A_d^*) \sim F^*$, and, for $\delta = 2^{-m}$ and for $k = 0,\dots,2^m-1$, define

$$\begin{aligned} M_k &= \Pr\Big\{(A_1^*,\dots,A_d^*) \,\Big|\, k\delta \le \phi(A_1^*,\dots,A_d^*) < (k+1)\delta\Big\} \\ &= \Pr\Big\{\big(F_1^{-1}(U),\dots,F_d^{-1}(U)\big) \,\Big|\, k\delta \le \phi\big(F_1^{-1}(U),\dots,F_d^{-1}(U)\big) < (k+1)\delta\Big\}. \end{aligned}$$

Then, it follows that

$$\int \phi \mathrm{d}F^* = \lim_{m\to\infty} \sum_{k=0}^{2^m-1} k\delta M_k = \mathbb{E}_U\Big[\phi\big(F_1^{-1}(U),\dots,F_d^{-1}(U)\big)\Big],$$

where both equalities follow from the piecewise approximations of $\phi(\cdot)$ where it is replaced with $k\delta$ whenever it evaluates to a value in the range $\big[k\delta,(k+1)\delta\big)$ and the Dominated Convergence Theorem (see Theorem 16.4 in [5]), which applies because of the existence of $\varphi(\cdot)$. $\square$

Similar to Corollary 4, Theorem 6 can be used to characterize $\hat{\phi}(\cdot)$ for functions $\phi(\cdot)$ that become supermodular when their domain is transformed affinely, by using an operation such as the switching operation. Recall that the function $\phi(T)$, obtained by switching the domain of $\phi:[0,1]^d \to \mathbb{R}$, is described using a set $T \subseteq \{1,\dots,d\}$ so that $\phi(T)(f) = \phi\big(f(T)\big)$, where $f(T)_i = 1-f_i$ if $i \in T$ and $f(T)_i = f_i$ otherwise. We define a marginal distribution $F_i(T)$ so that

$$F_i(T)(b) = \Pr\{1-A_i \le b\} \quad \text{for } i \in T \qquad \text{and} \qquad F_i(T)(b) = \Pr\{A_i \le b\} \quad \text{otherwise.} \tag{34}$$

Then, it follows that for $i \in T$

$$\begin{aligned} F_i(T)^{-1}(\lambda) &= \min\big\{b \in [0,1] \,\big|\, F_i(T)(b) \ge \lambda\big\} = \min\Big\{b \in [0,1] \,\Big|\, 1 - \sup_{a<1-b} F_i(a) \ge \lambda\Big\} \\ &= 1 - \max\Big\{d \in [0,1] \,\Big|\, \sup_{a<d} F_i(a) \le 1-\lambda\Big\} = 1 - \sup\big\{a \in [0,1] \,\big|\, F_i(a) \le 1-\lambda\big\}. \end{aligned}$$

The following result explicitly characterizes $\hat{\phi}(\cdot)$ when $\phi(T)$, instead of ϕ, is supermodular.

Corollary 8. *If $\phi:[0,1]^d \to \mathbb{R}$ is a continuous function and there exists a $T \subseteq \{1,\dots,d\}$ so that $\phi(T)$ is supermodular then*

$$\hat{\phi}(F_1,\dots,F_d) = \int_0^1 \phi(T)\big(F_1(T)^{-1}(\lambda),\dots,F_d(T)^{-1}(\lambda)\big)\mathrm{d}\lambda.$$

Proof. We will show that $\hat{\phi}(F_1,\dots,F_d)$ equals $\widehat{\phi(T)}\big(F_1(T),\dots,F_d(T)\big)$ and then, the result follows directly from Theorem 6 using supermodularity of $\phi(T)$. To show $\hat{\phi}(F_1,\dots,F_d) \le \widehat{\phi(T)}\big(F_1(T),\dots,F_d(T)\big)$, consider $F \in \Pi(F_1,\dots,F_d)$, and let $A \sim F$. Define a random vector A' so that, for $S \subseteq [0,1]^d$, $\Pr\{A' \in S\} := \Pr\big\{A \in \{a(T) \mid a \in S\}\big\}$. Then, let F' be the cumulative distribution function of A', *i.e.*, $F'(b) = \Pr\big\{A \in \{a(T) \mid a \in [0,b]\}\big\}$. Assuming wlog that the range of ϕ is $[0,1]$, for $\delta = 2^{-m}$, let $S_k := \big\{a \in [0,1]^d \,\big|\, k\delta \le \phi(a) < (k+1)\delta\big\}$, and thus,

$$\begin{aligned} M_k := \Pr\big\{A \in S_k\big\} &= \Pr\Big\{A' \in \big\{a(T) \,\big|\, a \in S_k\big\}\Big\} \\ &= \Pr\Big\{A' \,\Big|\, k\delta \le \phi(T)(A') < (k+1)\delta\Big\}, \end{aligned} \tag{35}$$

where the second equality holds by the definition of A', and the third equality holds because $\phi(T)\big(a(T)\big) = \phi(a)$. Therefore,

$$\int \phi \mathrm{d}F = \lim_{m\to\infty} \sum_{k=0}^{2^m-1} k\delta M_k = \int \phi(T)\mathrm{d}F' \le \widehat{\phi(T)}\big(F_1(T),\dots,F_d(T)\big),$$

where the two equalities follow from the Dominated Convergence Theorem (see Theorem 16.4 in [5]) and the first and last equalities in (35), and the inequality holds because, using (34), the marginal distribution A'_i is given by $\Pr\{A'_i \le a_i\} = F_i(T)(a_i)$. Hence, $\hat{\phi}(F_1,\ldots,F_d) \le \widehat{\phi(T)}\big(F_1(T),\ldots,F_d(T)\big)$. Since the reverse inequality follows by a similar argument, the proof is complete. $\square$

4.3. Composite relaxations via random variables In this subsection, we will assume that, for each inner function, an underestimator, parametrized by its real-valued upper bound a_i, is available. For a given $x \in X$, the underestimator will vary with bound a_i and will be used to derive the marginal cumulative distribution function F_i used in Section 4.2. Then, we will use Theorem 6 to construct the composite relaxation. To relate the marginal distributions to the underestimating functions, we will find it useful to work with an alternate characterization of a distribution function with support over $[0,1]$ in terms of a concave function on the real line. To derive this function, we truncate the associated random variable to lie below a bound and study how the expectation varies with this bound. Formally, we define $E:\mathbb{R}\times\mathcal{F}$ as:

$$E(a_i,F_i) = \mathbb{E}_{F_i}\big[\min\{A_i,a_i\}\big] \qquad \text{for } a_i \in \mathbb{R} \text{ and } F_i \in \mathcal{F}, \tag{36}$$

where $A_i \sim F_i$. We will write $E_{F_i}(a_i)$ (resp. $E_{a_i}(F_i)$) when we wish to convey that F_i (resp. a_i) is fixed. It is the right derivative of $E_{F_i}(a_i)$ that relates to the cumulative distribution F_i. Recall that the left and right derivative of a univariate function $c(a)$ are defined as $c'_-(a) = \lim_{\delta\nearrow 0}\frac{c(a+\delta)-c(a)}{\delta}$ and $c'_+(a) = \lim_{\delta\searrow 0}\frac{c(a+\delta)-c(a)}{\delta}$. We adapt Theorem 1 in [32] for our purpose.

Lemma 8. *For a distribution function $F_i \in \mathcal{F}$, the univariate function $E_{F_i}(a_i)$ is non-decreasing concave such that*

$$E_{F_i}(0) = 0, \quad E_{F_i}(1) = \mathbb{E}[A_i], \quad (E_{F_i})'_-(a_i) = 1 \text{ for } a_i \le 0, \quad (E_{F_i})'_+(a_i) = 0 \text{ for } a_i \ge 1, \tag{37}$$

and the distribution function F_i can be recovered from E_{F_i} using $F_i(a_i) = 1-(E_{F_i})'_+(a_i)$. On the other hand, any concave function $c_i(a_i)$ on $\mathbb{R}$ with the properties that

$$c_i(0) = 0, \quad c_i(1) = \text{a finite value}, \quad (c_i)'_-(a_i) = 1 \text{ for } a_i \le 0, \quad (c_i)'_+(a_i) = 0 \text{ for } a_i \ge 1 \tag{38}$$

is $E_{F_i}(a_i)$ for some distribution function $F_i \in \mathcal{F}$.

Proof. For a distribution function $F_i \in \mathcal{F}$, the univariate function $E_{F_i}(a_i)$ is clearly non-decreasing. It is concave because, for $a'_i, a''_i \in \mathbb{R}$ and $\alpha\in[0,1]$, $\mathbb{E}_{F_i}[\min\{A_i, \alpha a'_i + (1-\alpha)a''_i\}] \le \alpha\mathbb{E}_{F_i}[\min\{A_i,a'_i\}] + (1-\alpha)\mathbb{E}_{F_i}[\min\{A_i,a''_i\}]$, where the inequality holds by the concavity of $\min\{a_i,a'_i\}$ in a'_i. Moreover, $E_{F_i}(0) = \int \min\{a_i,0\}\mathrm{d}F_i(a_i) = 0$ and $E_{F_i}(1) = \int\min\{a_i,1\}\mathrm{d}F_i(a_i) = \mathbb{E}_{F_i}[A_i]$. In addition, for $a_i \le 0$

$$\lim_{\delta\searrow 0}\frac{E_{F_i}(a_i) - E_{F_i}(a_i-\delta)}{\delta} = \lim_{\delta\searrow 0}\frac{a_i-(a_i-\delta)}{\delta} = 1;$$

and for $a_i \ge 1$

$$\lim_{\delta\searrow 0}\frac{E_{F_i}(a_i+\delta) - E_{F_i}(a_i)}{\delta} = \lim_{\delta\searrow 0}\frac{\mathbb{E}[A_i]-\mathbb{E}[A_i]}{\delta} = 0.$$

Last, it follows from Theorem 1 in [32] that $F_i(a_i)$ can be recovered using $1-(E_{F_i})'_+(a_i)$.

Now, let $c_i(a_i)$ be a concave function satisfying (38). It follows from $(c_i)'_-(0) = 1$, the concavity of c_i, and $c(0)=0$ that $c_i(a_i) \le a_i$. Similarly, it follows from $c_i(1) = M$ for some constant M, $(c_i)'_+(1) = 0$, and the concavity of c_i that $c_i(a_i) \le M$ for all a_i. Finally, since $(c_i)'_-(a_i) = 1$ for $a_i \le 0$, it follows that, for $b_i < 0$, $c_i(b_i) = c_i(b_i) - c_i(0) = \int_0^{b_i}(c_i)'_-(a_i) = b_i$. Therefore, $\lim_{a_i\to-\infty}\big(a_i - c_i(a_i)\big) = 0$. Thus, by Theorem 1 in [32], there exists a distribution function F_i with support over $\mathbb{R}$ such that $E_{F_i}(a_i) = c_i(a_i)$. The proof is complete if we show that $F_i(b_i) = 0$ for $b_i < 0$ and $F_i(b_i) = 1$ for

$b_i > 1$. Assume that there exists $b_i < 0$ such that $F_i(b_i) > 0$. Then, $c_i(0) = \int \min\{a_i, 0\}\mathrm{d}F_i < 0$, a contradiction. Similarly, suppose that $F_i(1) < 1$. This case also leads to a contradiction as follows, $\lim_{a_i\to\infty} c_i(a_i) = \int a_i \mathrm{d}F_i > \int \min\{a_i, 1\}\mathrm{d}F_i = M \geq \lim_{a_i\to\infty} c_i(a_i)$, where the last inequality follows because $c_i(a_i) \leq M$ for all a_i. $\square$

Lemma 8 establishes a connection between certain univariate concave functions and distribution functions. We now relate these univariate functions with certain underestimators of the inner functions. Formally, we consider a function $s_i : W \times \mathbb{R} \to \mathbb{R}$ such that, for $(x, f, a_i) \in W \times \mathbb{R}$,

$$s_i(x, f, a_i) \in \Big[\min\big\{a_i, f_i \mathbb{1}_{a_i \geq 1}\big\},\ \min\big\{a_i, f_i\big\}\Big], \tag{39}$$

where W outer-approximates the graph of inner function $f(x)$, and $\mathbb{1}_{\text{clause}}$ is one if the clause is true and 0 otherwise. For any (x, f, a_i), the range of values for $s_i(x, f, a_i)$ is non-empty because $f_i \mathbb{1}_{a_i\geq 1} \leq f_i$. In other words, (39) requires that, for $a_i \in [0,1]$, $s_i(x, f, a_i)$ underestimates $\min\{f_i, a_i\}$ over W, for $a_i \leq 0$, $s_i(x, f, a_i)$ is a_i, and, for $a_i \geq 1$, the function coincides with f_i. We construct one such function in the following remark.

Remark 3. Let W_i be a convex outerapproximation of the graph of inner function $f_i(x)$. With each constant $a_i \in \mathbb{R}$, associate a set $S_i(a_i) := \big\{(x, f_i, \rho_i) \bigm| \rho_i \geq \min\{a_i, f_i\},\ (x, f_i) \in W_i\big\}$. Define $s_i : W_i \times \mathbb{R} \to \mathbb{R}$ so that, for any $a_i \in \mathbb{R}$,

$$s_i\big(x, f_i, a_i\big) := \inf\Big\{\rho_i \Bigm| (x, f_i, \rho_i) \in \operatorname{conv}\big(S_i(a_i)\big)\Big\}. \tag{40}$$

To see that the function s_i satisfies the requirements in (39), consider a constant $a_i \in \mathbb{R}$. If $a_i \leq 0$ then $\min\{a_i, f_i\}$ equals a_i, the set $S_i(a_i)$ is convex, and $s_i(x, f_i, a_i) = a_i$. Similarly, if $a_i \geq 1$, $\min\{a_i, f_i\}$ equals f_i, $S_i(a_i)$ is convex, and $s_i(x, f_i, a_i) = f_i$. If $0 \leq a_i \leq 1$ then $0 \leq s_i(x, f_i, a_i) \leq \min\{a_i, f_i\}$, where the first inequality holds because, for each $(x, f_i, \rho_i) \in S_i(a_i)$, $0 \leq \min\{a_i, f_i\} \leq \rho_i$, and the second inequality holds because, for every $(x, f_i) \in W_i$, $\big(x, f_i, \min\{a_i, f_i\}\big) \in \operatorname{conv}\big(S_i(a_i)\big)$. Furthermore, $s_i(x, f_i, a_i)$ is a convex function over W_i; see Theorem 5.3 in [31]. In fact, for any fixed $a_i \in \mathbb{R}$, $s_i(x, f_i, a_i)$ is the convex envelope of the function $\min\{a_i, f_i\}$ over W_i. In contrast, for any fixed $(x, f_i) \in W_i$, $s_i(x, f_i, a_i)$ is a concave function in a_i. To see this, consider two distinct points a_i', a_i'' in $\mathbb{R}$, and define $\tilde{a}_i := \lambda a_i' + (1-\lambda)a_i''$ for some $\lambda \in (0,1)$. Since $\lambda s_i(x, f_i, a_i') + (1-\lambda) s_i(x, f_i, a_i'')$ (resp. $s_i(x, f_i, \tilde{a}_i)$) is a convex underestimator (resp. convex envelope) of $\min\{f_i, \tilde{a}_i\}$ over W_i, it follows that $\lambda s_i(x, f_i, a_i') + (1-\lambda) s_i(x, f_i, a_i'') \leq s_i(x, f_i, \tilde{a}_i)$. $\square$

In the following example, we consider the quadratic term, derive the underestimator (40) explicitly, and illustrate that this underestimator, treated as a function of a_i, is concave and, via the transformation discussed in Lemma 8, yields a distribution function.

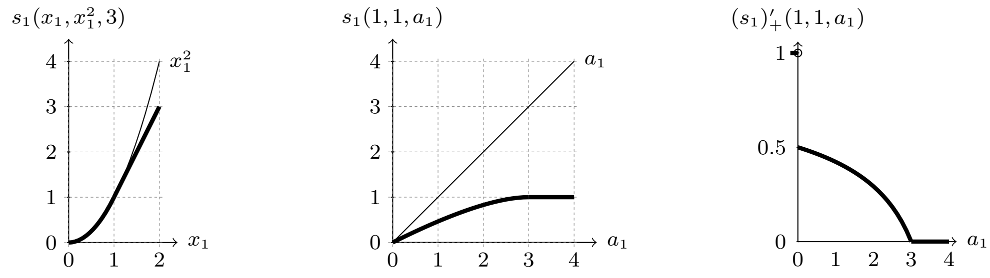
Example 3. Consider the quadratic term x_1^2 over the interval $[0,2]$. Here, the quadratic term varies over $[0,4]$, while in our formal treatment, we have assumed that $f_i \in [0,1]$. However, as we discussed before, this does not pose any issues since an affine transformation of the function can be used to normalize any bounded range to $[0,1]$. In our current setting, we could use $\frac{1}{4}x_1^2$ as the inner function instead of x_1^2. The function $s_1(x_1, x_1^2, a_1)$ defined in (40) can be computed explicitly as follows

$$s_1(x_1, x_1^2, a_1) = \begin{cases} a_1 & a_1 \leq 0 \\ \big(-4 + 2\sqrt{-a_1 + 4}\big)(2 - x_1) + a_1 & 0 \leq 2 - \sqrt{-a_1 + 4} \leq x_1 \leq 2 \\ x_1^2 & \text{otherwise.} \end{cases} \tag{41}$$

Figure 5a illustrates the function at $a_1 = 3$, where we see that this function is the largest convex underestimator of x_1^2 over $[0,2]$ bounded by 3. In contrast, Figure 5b depicts $s_1(x_1, x_1^2, a_1)$ as a function of a_1 at $x_1 = 1$ and it is easily verified that this function is concave and satisfies the requirements in (38). For general a_1, the right derivative of $s_1(x_1, x_1^2, a_1)$ with respect to a_1 is as follows:

$$(s_1)'_+(x_1, x_1^2, a_1) = \begin{cases} 1 & a_1 < 0, \\ 1 - \frac{2-x_1}{\sqrt{-a_1+4}} & 0 \le 2 - \sqrt{-a_1+4} < x_1 \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

By Lemma 8, $(s_1)'_+(1,1,a_1)$ is a survival function ($1-$ distribution function) for a random variable with support in $[0,4]$ and is depicted in Figure 5c. $\square$



(a) (b) (c)

FIGURE 5. (a) the largest convex underestimator of x_1^2 over $[0,2]$ that is bounded from above by 3. (b) a concave function which satisfies (38). (c) the right derivative of $s_1(1,1,a_1)$ is a survival function.

In the following result, we formally relate the underestimator $s_i(x,f,a_i)$ with a distribution function. Assume that we are given a function $s_i(x,f,a_i)$ that satisfies (39). We denote by $(s_i)'_+(x,f,a_i)$ the right derivative of $s_i(x,f,a_i)$ with respect to a_i. We will fix $(x,f) \in W$ and characterize $1-(s_i)'_+(x,f,a_i)$ as a distribution function, which, for notational brevity, we denote as $S_i^{x,f}(a_i)$.

PROPOSITION 7. *Let $s_i : W \times \mathbb{R}$ be a function satisfying (39), and define $S_i^{x,f}(a_i) := 1 - (s_i)'_+(x,f,a_i)$. If, for any given $(x,f) \in W$, $s_i(x,f,a_i)$ is concave in a_i then $S_i^{x,f}$ is a distribution function so that $\int a_i \mathrm{d}S_i^{x,f}(a_i) = f_i$. Moreover, $s_i(x,f,a_i) = E_{S_i^{x,f}}(a_i)$.*

Proof. It is easy to verify that function $s_i(x,f,a_i)$ that is concave in a_i and satisfies the requirement in (39) also satisfies the four conditions in (38). In particular, since $s_i(x,f,a_i)$ equals a_i (resp. f_i) when $a_i \le 0$ (resp. $a_i \ge 1$), it follows that $(s_i)'_-(x,f,a_i)$ (resp. $(s_i)'_+(x,f,a_i)$) equals 1 (resp. 0). Then, by the second part of Lemma 8, there exists a distribution function $F_i \in \mathcal{F}$ such that $s_i(x,f,a_i) = E_{F_i}(a_i)$. By the first part of Lemma 8, $F_i(a_i) = 1 - (E_{F_i})'_+(a_i) = 1 - (s_i)'_+(x,f,a_i) = S_i^{x,f}(a_i)$. Moreover, $\int a_i \mathrm{d}S_i^{x,f}(a_i) = \lim_{a_i \to +\infty} s_i(x,f,a_i) = f_i$, where the first equality is because $s(x,f,a_i)$ equals $E_{S_i^{x,f}}(a_i)$, which in turn approaches the right hand side as $a_i \to \infty$, and the second equality follows directly from (39). $\square$

Equipped with Proposition 6 and Proposition 7, we are ready to derive the limiting relaxation for the hypograph of $\phi \circ f$ as follows. For each $i \in \{1,\dots,d\}$, let $S_i^{x,f}(a_i)$ be a function defined as in Proposition 7. Then, if the outer-function $\phi(\cdot)$ satisfies (30), for example, as in Lemma 7, if $\phi(\cdot)$ is continuous and convex in each argument when other arguments are fixed, we obtain that, for every $(x,f) \in W$,

$$\phi(f) = \tilde{\phi}(S_1^{x,f},\dots,S_d^{x,f}) \le \hat{\phi}\big(S_1^{x,f},\dots,S_d^{x,f}\big),$$

where the first equality holds by Proposition 7 and the definition of $\tilde{\phi}(\cdot)$, and the inequality holds because, by Proposition 6, $\hat{\phi}(\cdot)$ is the lowest concave overestimator of $\tilde{\phi}(\cdot)$ over $\mathcal{F}^d$. We will show that the limiting relaxation $\hat{\phi}(S_1^{x,f},\dots,S_d^{x,f})$ has a convex representation in the space of variables (x,f). Before providing a formal discussion, we illustrate the ideas on an example.

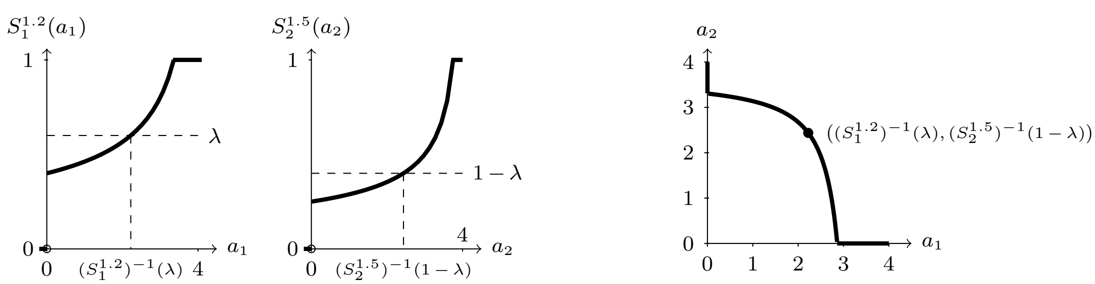
EXAMPLE 4. Consider $x_1^2x_2^2$ over the rectangle $[0,2]^2$. We use Proposition 7 to derive distribution functions from underestimators of x_i^2. For underestimator $s_i(x,x_i^2,a_i)$ given in (41), $1-(s_i)'_+(x,x_i^2,a_i)$ is easily computed using the right derivative in (42). For notational brevity, let $S_i^{x_i}(a_i)$ denote $1-(s_i)'_+(x,x_i^2,a_i)$ since it depends only on the i^{th} coordinate of x. For any $x_i\in[0,2]$, it follows from Proposition 7 that $S_i^{x_i}$ is a distribution function of a random variable $A_i^{x_i}$ such that $\mathbb{E}[A_i^{x_i}]=x_i^2$. Let $D(a_1,a_2):=S_1^{x_1}(a_1)S_2^{x_2}(a_2)$ and $G^*(a_1,a_2):=\max\{0,S_1^{x_1}(a_1)+S_2^{x_2}(a_2)-1\}$. Then, in this example setting, our construction is essentially derived from the following argument:

$$\begin{aligned}x_1^2x_2^2&=\mathbb{E}[A_1^{x_1}]\mathbb{E}[A_2^{x_2}]=\mathbb{E}_D[A_1^{x_1}A_2^{x_2}]\geq\inf_G\left\{\mathbb{E}_G[A_1^{x_1}A_2^{x_2}]\;\middle|\;G\in\Pi(S_1^{x_1},S_2^{x_2})\right\}\\&=\mathbb{E}_{G^*}[A_1^{x_1}A_2^{x_2}]=\mathbb{E}_U\left[(S_1^{x_1})^{-1}(U)(S_2^{x_2})^{-1}(1-U)\right],\end{aligned}\tag{43}$$

where the first equality is because $\mathbb{E}[A_i^{x_i}]=x_i^2$, the second equality holds since D is constructed by coupling $A_1^{x_1}$ and $A_2^{x_2}$ independently, and the first inequality holds because the product distribution D has $S_1^{x_1}$ and $S_2^{x_2}$ as marginals. The third equality holds because G^* is feasible to the optimization problem on the left hand side, and because, for two marginals $S_1^{x_1},S_2^{x_2}\in\mathcal{F}$, $\Pr\{(A_1^{x_1}>a_1)\cup(A_2^{x_2}>a_2)\}\leq\Pr\{A_1^{x_1}>a_1\}+\Pr\{A_2^{x_2}>a_2\}$. This implies that, for $G\in\Pi(S_1^{x_1},S_2^{x_2})$, $G(a_1,a_2)\geq G^*(a_1,a_2)$, where the right hand side is known as Hoeffding-Fréchet lower bound [18, 13], and, thus, by [11] $\mathbb{E}_G[A_1^{x_1}A_2^{x_2}]\geq\mathbb{E}_{G^*}[A_1^{x_1}A_2^{x_2}]$ since the bilinear term is a correlation affine function. The third equality also follows from the more general result in Corollary 8 choosing either $T=\{1\}$ or $T=\{2\}$. The last equality holds because the distribution function of the random vector $\big((S_1^{x_1})^{-1}(U),(S_2^{x_2})^{-1}(1-U)\big)$ is G^*. We depict in Figure 6a the marginal distributions $S_1^{1.2}$ and $S_2^{1.5}$. For a given $U=\lambda$, we compute their inverse values to locate a point on the curve that is the locus of support points for G^*; see Figure 6b. We evaluate the last term in (43) by integrating the function value at points on this curve to derive the following limiting composite relaxation:

$$\begin{aligned}x_1^2x_2^2&\geq\max\left\{0,\int_{1-\frac{x_1}{2}}^{\frac{x_2}{2}}\left(\frac{4\lambda^2-4+4x_1-x_1^2}{\lambda^2}\right)\left(\frac{4\lambda^2-8\lambda+4x_2-x_2^2}{(1-\lambda)^2}\right)\mathrm{d}\lambda\right\}\\&=\max\Big\{0,-2\ln(2-x_2)(-2+x_2)^2(-2+x_1)^2+2\ln(-x_2)(-2+x_2)^2(-2+x_1)^2\\&\qquad-2\ln(-2+x_1)(-2+x_2)^2(-2+x_1)^2+2\ln(x_1)(-2+x_2)^2(-2+x_1)^2\\&\qquad-4x_1^2x_2-4x_1x_2^2+12x_1^2+32x_1x_2-48x_1+12x_2^2-48x_2+48\Big\}\end{aligned}$$

whose convexity, although not directly apparent from the resulting formula, is a consequence of Corollary 9 proved later. □



(a) marginal distributions (b) support of optimal plan

FIGURE 6. Underestimating $x_1^2x_2^2$ at the point $x=(1.2,1.5)$ via optimal transport

To show that the the limiting relaxation is convex in (x, f), we will need certain monotonicity properties of the optimal functional $\hat{\phi}(\cdot)$. Given two tuples of distribution functions $(F_1, \ldots, F_d)$ and $(G_1, \ldots, G_d)$, we say $(F_1, \ldots, F_d) \preceq (G_1, \ldots, G_d)$ (resp. $(F_1, \ldots, F_d) \prec (G_1, \ldots, G_d)$) if, for each $i \in \{1, \ldots, d\}$, $F_i \preceq G_i$ (resp. $F_i \prec G_i$), which is defined as in (24) (resp. in (26)).

Proposition 8. *Let $\phi : [0,1]^d \to \mathbb{R}$ be a continuous function which is convex in each argument when the other arguments are fixed, and assume that there exists a $T \subseteq \{1, \ldots, d\}$ so that $\phi(T)$ is supermodular. If $(F_1, \ldots, F_d) \preceq (G_1, \ldots, G_d)$ then $\hat{\phi}(F_1, \ldots, F_d) \geq \hat{\phi}(G_1, \ldots, G_d)$. The weaker condition $(F_1, \ldots, F_d) \prec (G_1, \ldots, G_d)$ suffices to show $\hat{\phi}(F_1, \ldots, F_d) \geq \hat{\phi}(G_1, \ldots, G_d)$ if $\phi(T)$ is also non-increasing in each argument.*

Proof. Assume that $(F_1, \ldots, F_d) \preceq (G_1, \ldots, G_d)$. We will invoke Lemma 6 to show that $\hat{\phi}(F_1, \ldots, F_d) \geq \hat{\phi}(G_1, \ldots, G_d)$. Let $\eta_i(\lambda) := F_i(T)^{-1}(\lambda)$ and define $\psi(\lambda, \eta_1, \ldots, \eta_d) = \phi(T)(\eta_1, \ldots, \eta_d)$. Observe that the hypothesis on $\psi(\cdot)$ in Lemma 6 is satisfied by this definition since $\phi(T)$ is independent of λ and is assumed to be supermodular and convex in each argument when the others are fixed. Moreover, since the first argument of $\psi(\cdot)$ is ignored, the condition (29) holds trivially. Now, we show that (28) is satisfied with our definition. Since, for any concave univariate function $\psi(\cdot)$, $F_i \preceq G_i$ implies that $\mathbb{E}[\psi(1 - A_i)] \leq \mathbb{E}[\psi(1 - B_i)]$ where $A_i \sim F_i$ and $B_i \sim G_i$, it follows that $F_i(T) \preceq G_i(T)$, where $F_i(T)$ and $G_i(T)$ are defined as in (34). It follows from (25) (see Theorem 3.A.5 in [34]) that $F_i(T) \preceq G_i(T)$ if and only if

$$\int_0^p F_i(T)^{-1}(\lambda) \mathrm{d}\lambda \leq \int_0^p G_i(T)^{-1}(\lambda) \mathrm{d}\lambda \quad \text{for } p \in [0,1],$$

with equality achieved at $p = 1$. Last, observe that $F_i(T)^{-1}(\lambda)$ and $G_i(T)^{-1}(\lambda)$ are non-decreasing. Therefore, it follows from Lemma 6 that

$$\int_0^1 \phi(T)\big(F_1(T)^{-1}(\lambda), \ldots, F_d(T)^{-1}(\lambda)\big) \mathrm{d}\lambda \geq \int_0^1 \phi(T)\big(G_1(T)^{-1}(\lambda), \ldots, G_d(T)^{-1}(\lambda)\big) \mathrm{d}\lambda.$$

Hence, by Corollary 8, we conclude that $\hat{\phi}(F_1, \ldots, F_d) \geq \hat{\phi}(G_1, \ldots, G_d)$.

Assume now that $\phi(\cdot)$ is non-increasing. We prove that $\hat{\phi}(F_1, \ldots, F_d) \geq \hat{\phi}(G_1, \ldots, G_d)$ under the weaker condition $(F_1, \ldots, F_d) \prec (G_1, \ldots, G_d)$. Clearly, $F_i \prec G_i$ implies $F_i(T) \prec G_i(T)$. Then, it follows from 27 (see Theorem 4.A.6 in [34]) that $F_i(T) \prec G_i(T)$ if and only if there exists a distribution function $D_i \in \mathcal{F}$ such that $F_i(T)^{-1}(\lambda) \leq D_i^{-1}(\lambda)$ for all $\lambda \in [0,1]$ and $D_i \preceq G_i(T)$. Therefore,

$$\begin{aligned}
\int_0^1 \phi(T)\big(F_1(T)^{-1}(\lambda), \ldots, F_d(T)^{-1}(\lambda)\big) \mathrm{d}\lambda &\geq \int_0^1 \phi(T)\big(D_1^{-1}(\lambda), \ldots, D_d^{-1}(\lambda)\big) \mathrm{d}\lambda \\
&\geq \int_0^1 \phi(T)\big(G_1(T)^{-1}(\lambda), \ldots, G_d(T)^{-1}(\lambda)\big) \mathrm{d}\lambda,
\end{aligned}$$

where the first inequality holds because $\phi(T)$ is non-increasing and, for every $i \in \{1, \ldots, d\}$ and $\lambda \in [0,1]$, $F_i(T)^{-1}(\lambda) \leq D_i^{-1}(\lambda)$, and second inequality was established above. Thus, the result follows from Corollary 8. $\square$

Now, we derive the limiting composite relaxation. Our construction will be based on three key ideas: (a) for each (x, f), the underestimator will be used as in Proposition 7 to derive marginal distributions $S_1^{x,f}, \ldots, S_d^{x,f}$; (b) under the technical condition (30), Proposition 6 will be used to relax $\phi(f)$ by $\hat{\phi}(S_1^{x,f}, \ldots, S_d^{x,f})$; (c) $\hat{\phi}(S_1^{x,f}, \ldots, S_d^{x,f})$ will be further relaxed to $\hat{\phi}(F_1, \ldots, F_d)$, where $(F_1, \ldots, F_d) \preceq (S_1^{x,f}, \ldots, S_d^{x,f})$ using Proposition 8 without sacrificing the quality of the relaxation. We present these ideas formally in the next result and establish the convexity of the resulting relaxation.

THEOREM 7. *Let $\phi \circ f$ be a composite function, where $\phi : [0,1]^d \to \mathbb{R}$ is a continuous function which is convex in each argument when other arguments are fixed, and $f : \mathbb{R}^m \to [0,1]^d$ is a vector of functions over a subset X of $\mathbb{R}^m$. For each $i \in \{1,\dots,d\}$, let $s_i : W \times \mathbb{R} \to \mathbb{R}$ be a function which satisfies (39) and is concave in a_i. Define $S_i^{x,f}(a_i) := 1 - (s_i)'_+(x,f,a_i)$. Then, $\operatorname{proj}_{(x,\phi)}(R)$ is a relaxation of hypograph of $\phi \circ f$, where:*

$$R := \Big\{(x,f,\phi,F_1,\dots,F_d) \;\Big|\; \phi \le \hat{\phi}(F_1,\dots,F_d),\ (x,f) \in W,\ F_i \in \mathcal{F},\ S_i^{x,f} \preceq F_i,\ i = 1,\dots,d\Big\}, \tag{44}$$

Moreover,
1. *if, for each fixed a_i, $s_i(x,f,a_i)$ is convex in (x,f) and W is convex then R is convex;*
2. *if $\phi(\cdot)$ is a supermodular function then*

$$\operatorname{proj}_{(x,f,\phi)}(R) = \Big\{(x,f,\phi) \;\Big|\; (x,f) \in W,\ \phi \le \int_0^1 \phi\Big(\big(S_1^{x,f}\big)^{-1}(\lambda),\dots,\big(S_d^{x,f}\big)^{-1}(\lambda)\Big)\mathrm{d}\lambda\Big\}; \tag{45}$$

3. *if $\phi(\cdot)$ is supermodular and non-increasing in each argument and, for fixed $a_i \in \mathbb{R}$, $s_i(x,f(x),a_i)$ is convex in X then we obtain a convex relaxation of the hypograph of $\phi \circ f$:*

$$\Big\{(x,\phi) \;\Big|\; x \in X,\ \phi \le \int_0^1 \phi\Big(\big(S_1^{x,f(x)}\big)^{-1}(\lambda),\dots,\big(S_d^{x,f(x)}\big)^{-1}(\lambda)\Big)\mathrm{d}\lambda\Big\}. \tag{46}$$

Proof. We first prove that $\operatorname{hyp}(\phi \circ f) \subseteq \operatorname{proj}_{(x,\phi)}(R)$, where $\operatorname{hyp}(\phi \circ f)$ denotes the hypograph of $\phi \circ f$. Let $(x,\phi) \in \operatorname{hyp}(\phi \circ f)$. Define $f := f(x)$ and observe that $(x,f) \in W$ because W is a relaxation of $\operatorname{gr}(f)$. Moreover, let $F_i := S_i^{x,f}$. By Proposition 7, F_i is the distribution function of a random variable A_i such that $\mathbb{E}[A_i] = f_i$. By $(x,\phi) \in \operatorname{hyp}(\phi \circ f)$, $f = f(x)$, $F_i = S_i^{x,f}$, and Proposition 6, we obtain $\phi \le \phi\big(f(x)\big) = \phi(f) = \tilde{\phi}(S_1^{x,f},\dots,S_d^{x,f}) = \tilde{\phi}(F_1,\dots,F_d) \le \hat{\phi}(F_1,\dots,F_d)$. Therefore, we conclude that $(x,f,\phi,F_1,\dots,F_d) \in R$. In other words, $\operatorname{hyp}(\phi \circ f) \subseteq \operatorname{proj}_{(x,\phi)}(R)$.

We now show that under the three conditions in the statement the claimed structure for the relaxation holds. We begin with Condition 1. Assume that, for each a_i, $s_i(x,f,a_i)$ is convex in (x,f) and W is convex. To show that R is convex, it suffices to argue that constraint $S_i^{x,f_i} \preceq F_i$ defines a convex set because W is assumed to be convex and, by Proposition 6, the functional $\hat{\phi}(\cdot)$ is concave over $\mathcal{F}^d$. It follows from (24) that $S_i^{x,f} \preceq F_i$ is equivalent to

$$E_{a_i}\big(S_i^{x,f}\big) \le E_{a_i}(F_i) \quad \forall a_i \in \mathbb{R} \qquad \text{and} \qquad \int a_i \mathrm{d}S_i^{x,f}(a_i) = \int a_i \mathrm{d}F_i(a_i). \tag{47}$$

For each $a_i \in \mathbb{R}$, by Proposition 7, we have $E_{a_i}(S_i^{x,f}) - E_{a_i}(F_i) = s_i(x,f,a_i) - E_{a_i}(F_i)$. Since we assumed that, for each a_i, $s_i(x,f,a_i)$ is a convex function, the convexity of the inequality in (47) follows if $E_{a_i}(F_i)$ is a concave function for each a_i. The latter follows because, for $\alpha \in [0,1]$ and $F_i, G_i \in \mathcal{F}$, $E_{a_i}\big(\alpha F + (1-\alpha)G\big) = \int \min\{a'_i, a_i\}\mathrm{d}\big(\alpha F_i(a'_i) + (1-\alpha)G_i(a'_i)\big) = \alpha E_{a_i}(F) + (1-\alpha)E_{a_i}(G)$. The second equation in (47) defines a convex set because, by Proposition 7, $\int a_i \mathrm{d}S_i^{x,f}(a_i) = f_i$, and, for $F_i, G_i \in \mathcal{F}$ so that $\int a_i \mathrm{d}F_i = \int a_i \mathrm{d}G_i = f_i$, $\int a_i \mathrm{d}\big(\alpha F_i(a_i) + (1-\alpha)G_i(a_i)\big) = \alpha \int a_i \mathrm{d}F_i(a_i) + (1-\alpha)\int a_i \mathrm{d}G_i(a_i) = f_i$. It follows that the set described in (47) is convex in the space of $(x,f,F_1,\dots,F_d)$ variables.

Next, we prove Condition 2. Let $R' := \big\{(x,f,\phi) \;\big|\; (x,f) \in W,\ \phi \le \hat{\phi}(S_1^{x,f},\dots,S_1^{x,f})\big\}$, which, by Theorem 6, is the set in the right hand side of (45). To show $R' \subseteq \operatorname{proj}_{(x,f,\phi)}(R)$, we consider a point $(x,f,\phi) \in R'$ and define $(F_1,\dots,F_d) = (S_1^{x,f},\dots,S_d^{x,f})$. Then, by Proposition 7, we have $\int a_i \mathrm{d}F_i(a_i) = f_i$. Since $S_i^{x,f} \preceq F_i$ holds trivially, it follows that $(x,f,\phi,F_1,\dots,F_d) \in R$, showing that $R' \subseteq \operatorname{proj}_{(x,f,\phi)}(R)$. To prove that $\operatorname{proj}_{(x,f,\phi)}(R) \subseteq R'$, we consider a point $(x,f,\phi,F_1,\dots,F_d)$ of R and show that $(x,f,\phi) \in R'$. It follows readily that $\phi \le \hat{\phi}(F_1,\dots,F_d) \le \hat{\phi}\big(S_1^{x,f},\dots,S_d^{x,f}\big)$, where second inequality holds because $\big(S_1^{x,f},\dots,S_d^{x,f}\big) \preceq (F_1,\dots,F_d)$ and, by Proposition 8, the functional

$\hat{\phi}(\cdot)$ is non-increasing under the order $\preceq$. Since $(x,f)\in W$, we conclude that $(x,f,\phi)\in R'$ and $\mathrm{proj}_{(x,f,\phi)}(R)\subseteq R'$.

Last, we prove Condition 3. Assume that $s_i(x,f(x),a_i)$ is convex in x. We start by showing $\mathrm{proj}_{(x,\phi)}(\tilde{R})$ is a convex relaxation of the hypograph of $\phi\circ f$, where

$$\tilde{R}:=\bigl\{(x,\phi,F_1,\dots,F_d)\bigm| \phi\le\hat{\phi}(F_1,\dots,F_d),\ (F_1,\dots,F_d)\in\mathcal{F}^d,\ S_i^{x,f(x)}\prec F_i,\ i=1,\dots,d\bigr\}. \tag{48}$$

Then, we observe that $\mathrm{hyp}(\phi\circ f)\subseteq\mathrm{proj}_{(x,\phi)}(R)\subseteq\mathrm{proj}_{(x,\phi)}(\tilde{R})$, where the first containment was shown above and the second containment holds because, by the definition in (24), the constraint $S_i^{x,f(x)}\preceq F_i$ imposes an additional condition on $S_i^{x,f(x)}\prec F_i$. The convexity of $\tilde{R}$ follows because $\hat{\phi}(\cdot)$ is concave over $\mathcal{F}^d$, and the constraint $S_i^{x,f(x)}\prec F_i$ is convex because, by the alternative characterization of increasing concave order in (26), it can be imposed using $s_i\bigl(x,f(x),a_i\bigr)\le E_{a_i}(F_i)$ for every $a_i\in\mathbb{R}$. This defines a convex set because, for every $a_i\in\mathbb{R}$, $s_i\bigl(x,f(x),a_i\bigr)$ is convex and $E_{a_i}(F_i)$ was shown to be linear in the proof of Condition 1. Now, let R'' be the set defined by (46). We will show $\mathrm{proj}_{(x,\phi)}(\tilde{R})=R''$. We have $R''=\mathrm{proj}_{(x,\phi)}(R)\subseteq\mathrm{proj}_{(x,\phi)}(\tilde{R})$, where the first equality holds by Condition 2 and $W:=\mathrm{gr}(f)$, and the second equality holds because $R\subseteq\tilde{R}$. Now, to prove $\mathrm{proj}_{(x,\phi)}(\tilde{R})\subseteq R''$, we consider a point $(x,\phi,F_1,\dots,F_d)$ of $\tilde{R}$ and show $(x,\phi)\in R''$. It follows readily that $\phi\le\hat{\phi}(F_1,\dots,F_d)\le\hat{\phi}\bigl(S_1^{x,f(x)},\dots,S_d^{x,f(x)}\bigr)$, where the second inequality holds by $\bigl(S_1^{x,f(x)},\dots,S_d^{x,f(x)}\bigr)\prec(F_1,\dots,F_d)$ and because Proposition 8 shows that $\hat{\phi}(\cdot)$ is non-increasing in $\prec$ under the assumed properties of $\phi(\cdot)$. Thus, by Theorem 6, $(x,\phi)\in R''$ and, therefore, $\mathrm{proj}_{(x,\phi)}(\tilde{R})\subseteq R''$. $\square$

REMARK 4. We remark that $\hat{\phi}\bigl(S_1^{x,f},\dots,S_d^{x,f}\bigr)$ coincides with the composite function $\phi\circ f$ when the underestimating function $s_i(x,f,a_i)$ equals $\min\{f_i(x),a_i\}$. To see this, consider an $x\in X$ and let $f=f(x)$. It follows that, for $i=1,\dots,d$, $s_i(x,f,a_i)=a_i$ if $a_i\le f_i$ and $s_i(x,f,a_i)=f_i$ otherwise. Therefore, $S_i^{x,f}(a_i)=0$ if $a_i\le f_i$ and 1 otherwise. In other words, $S_i^{x,f}$ corresponds to the distribution function of a Dirac measure with all its mass at f_i. Therefore, the only joint distribution, feasible in the optimal transport formulation (33), is the distribution of a Dirac measure with all its mass at $(f_1,\dots,f_d)$. In other words, $\hat{\phi}(S_1^{x,f},\dots,S_d^{x,f})=\phi(f)$. $\square$

Observe that since the locus of points over which the integral (45) or (46) is taken is independent of the function $\phi(\cdot)$, (45) or (46) can be used to simultaneously treat a vector of functions θ_k, $k\in\{1,\dots,\kappa\}$. Using arguments similar to those in Conditions 2 and 3 of Theorem 7, we can extend the result to treat functions that become supermodular after switching. We record this result for its use in applications such as Example 4.

COROLLARY 9. *Let R be the set defined in (44). Assume the same setup as Theorem 7 except that the assumed properties on $\phi(\cdot)$ apply to $\phi(T)$, where T is some subset of $\{1,\dots,d\}$. Then, Condition 2 applies with the definition of $\mathrm{proj}_{(x,f,\phi)}(R)$ replaced with:*

$$\mathrm{proj}_{(x,f,\phi)}(R)=\left\{(x,f,\phi)\,\middle|\,(x,f)\in W,\ \phi\le\int_0^1\phi(T)\Bigl(\bigl(S_1^{x,f}\bigr)(T)^{-1}(\lambda),\dots,\bigl(S_d^{x,f}\bigr)(T)^{-1}(\lambda)\Bigr)\mathrm{d}\lambda\right\}.$$

Similarly, Condition 3 applies, where the relaxation is replaced with:

$$\left\{(x,\phi)\,\middle|\,x\in X,\ \phi\le\int_0^1\phi(T)\Bigl(\bigl(S_1^{x,f(x)}\bigr)(T)^{-1}(\lambda),\dots,\bigl(S_d^{x,f(x)}\bigr)(T)^{-1}(\lambda)\Bigr)\mathrm{d}\lambda\right\}.\quad\square$$

5. Conclusions In this paper, we developed new tractable relaxations for composite functions. Our relaxations leverage the composite relaxation framework recently proposed in [17] that involves convexifying the outer-function over a polytope P. The polytope P encodes the structure of inner-functions using n estimators for each function. The structure of P generalizes that of a hypercube; the set used in factorable relaxations for a similar purpose and derived using bounds on the inner-function. Although convexifying general outer-functions over P is NP-Hard, we showed that when the outer-function is supermodular and concave-extendable, its concave envelope over P is determined by the staircase triangulation of a subset Q of P. Using this result, we found exponentially many inequalities describing the concave envelope of the outer-function over P. Since the polyhedral subdivision of P is invariant with the outer-function, we could convexify simultaneously the hypograph of a vector of composite functions. We also derived various inequalities regarding the structure of inequalities for the special case where the outer-function is multilinear.

We extended our results to the case with infinitely many estimators for each inner-function, by assuming that the outer-function is convex in each argument. For this extension, we described a marginal distribution for each inner-function by considering how underestimating function varies as a function of its upper bound. We then reformulated the concave envelope construction to an optimal transport problem and showed that the problem has an explicit solution when the outer-function is supermodular. Moreover, when the outer-function is non-decreasing, we exploited monotonicity properties of the explicit solution for the optimal transport problem with respect to a certain stochastic order to show that, as long as the underestimating functions were convex, we can derive a convex relaxation for the composite function in the space of the original problem variables.

Appendix A: Proof of Proposition 3 Clearly, $\mathrm{vert}(\Delta_i) = \{\zeta_{ij}\}_{j=0}^n$, where $\zeta_{ij} = \sum_{j'=0}^{j} e_{ij'}$, where $e_{ij'}$ is the j'-th standard basis vector in the space spanned by variables $(z_{i0},\dots,z_{in})$. Then, $\mathrm{vert}(\Delta)$ forms a lattice. Let $\bar{z}\in\Delta$. We require the first d entries, when $\bar{z}$ is sorted in non-increasing order, to be $\bar{z}_{i0}$, $i=1,\dots,d$. Then, if the $(d+k)^{\text{th}}$ variable in this order is $\bar{z}_{ij}$, we associate with $\pi_k = i$, that is a movement which steps from $j-1$ to j along the i^{th} direction. Thus, movement vector π describes a simplex S of the staircase triangulation of Δ that contains $\bar{z}$. Using Corollary 3.4 in [39], $\mathrm{conc}_\Delta(\eta)(\bar{z})$ can be obtained as an affine interpolation of $\eta(\cdot)$ over S, that is, $\mathrm{conc}_\Lambda(\eta)(\bar{z}) = \eta^{S}(\bar{z})$. $\square$

Appendix B: The explicit description of envelopes in Example 2 The concave envelope over P is given by

$$\max\begin{cases} w_1 := 0\\ w_2 := u_{11}+u_{21}-1\\ w_3 := u_{12}+3u_{21}-3\\ w_4 := 3u_{11}+u_{22}-3\\ w_5 := 4u_{11}+u_{23}-4\\ w_6 := u_{13}+4u_{21}-4\\ w_7 := 2u_{11}+u_{12}+2u_{21}+u_{22}-5\\ w_8 := 3u_{11}+u_{12}+2u_{21}+u_{23}-6\\ w_9 := 2u_{11}+u_{13}+3u_{21}+u_{22}-6\\ w_{10} := 3u_{11}+u_{13}+3u_{21}+u_{23}-7\\ w_{11} := 3u_{12}+3u_{22}-9\\ w_{12} := u_{11}+3u_{12}+2u_{22}+u_{23}-10\\ w_{13} := 2u_{12}+u_{13}+u_{21}+3u_{22}-10\\ w_{14} := u_{11}+2u_{12}+u_{13}+u_{21}+2u_{22}+u_{23}-11\\ w_{15} := 4u_{12}+3u_{23}-12\\ w_{16} := 3u_{13}+4u_{22}-12\\ w_{17} := 3u_{12}+u_{13}+u_{21}+3u_{23}-13\\ w_{18} := u_{11}+3u_{13}+3u_{22}+u_{23}-13\\ w_{19} := u_{12}+3u_{13}+u_{22}+3u_{23}-15\\ w_{20} := 4u_{13}+4u_{23}-16 \end{cases}$$

Appendix C: Proof of Lemma 5 Let $\mathcal{K}=\{K_1,\ldots,K_r\}$. Define $f^{\mathcal{K}}(s):=\min_{i=1}^r \chi^{K_i}(s)$ (resp. $f_j^{\mathcal{K}}(s):=\min_{i=1}^r \chi_j^{K_i}(s)$), where χ^{K_i} (resp. $\chi_j^{K_i}$) affinely interpolates $\big(v,f(v)\big)$ (resp. $\big(v,f_j(v)\big)$ for all $v\in \mathrm{vert}(K_i)$. Let $s\in K_i$. Since $\mathcal{K}$ is the triangulation of D, it follows that for some $\lambda\geq 0$ such that $\sum_v \lambda_v = 1$

$$\begin{aligned}\sum_{j=1}^m \mathrm{conc}_D(f_j)(s) \geq \mathrm{conc}_D(f)(s) \geq \chi^{K_i}(s) &= \sum_{v\in\mathrm{vert}(K_i)} \lambda_v \chi^{K_i}(v) = \sum_{v\in\mathrm{vert}(K_i)} \lambda_v f(v)\\ &= \sum_{v\in\mathrm{vert}(K_i)} \lambda_v\Big(\sum_{j=1}^m f_j(v)\Big) = \sum_{j=1}^m \sum_{v\in\mathrm{vert}(K_i)} \lambda_v f_j(v) = \sum_{j=1}^m \chi_j^{K_i}(s) \geq \sum_{j=1}^m \mathrm{conc}_D(f_j)(s),\end{aligned}$$

where the first inequality is because $\sum_{j=1}^m \mathrm{conc}_D(f_j)$ is a concave overestimator of f, the second inequality is because of Jensen's inequality and $\mathrm{conc}_D(f)$ is a concave function, the first equality is by definition of $\chi^{K_i}(s)$, the second equality is because $\chi^{K_i}(v)=f(v)$ for all $v\in\mathrm{vert}(K_i)$, the third equality is because of definition of f_j, the fourth equality is by interchanging the order of summation, the last equality is by the definition of $\chi_j^{K_i}(s)$ and $f_j(v)=\chi_j^{K_i}(v)$, and the last inequality is because $\mathrm{conc}(f_j)(x)=\min_{i=1}^r \chi_j^{K_i}(s)$. Therefore, equality holds throughout and $\mathrm{conc}_D(f)(s)=\chi^{K_i}(s)=\sum_{j=1}^m \mathrm{conc}_D(f_j)(s)$.

Now, we consider the case when a common triangulation does not exist. Let $\mathcal{K}=\{K_1,\ldots,K_r\}$ be the triangulation associated with the concave envelope of f and let j be such that the concave envelope of f_j is not associated with $\mathcal{K}$. Clearly, $\mathrm{conc}_D(f_j)(s)\geq \chi_j^{K_i}(s)$ for all $s\in K_i$. But, there must exist an i and an $s\in K_i$ such that $\mathrm{conc}_D(f_j)(s)>\chi_j^{K_i}(s)$. Otherwise, as shown in (8) $\mathrm{conc}_D(f_j)(s)=\min_{i=1}^r \chi_j^{K_i}(s)$, which contradicts the assertion that the concave envelope of f_j is not associated with the triangulation $\mathcal{K}$. Let $s=\sum_{v\in\mathrm{vert}(K_i)} v\lambda_v$ express s as a convex combination of vertices of K_i. It follows that

$$\begin{aligned}\mathrm{conc}_D(f)(s) &= \sum_{v\in\mathrm{vert}(K_i)} f(v)\lambda_v = \sum_{v\in\mathrm{vert}(K_i)}\sum_{j'=1}^m f_{j'}(v)\lambda_v\\ &= \sum_{j'=1}^m \sum_{v\in\mathrm{vert}(K_i)} f_{j'}(v)\lambda_v = \sum_{j'=1}^m \chi_{j'}^{K_i}(s) < \sum_{j'=1}^m \mathrm{conc}_D(f_{j'})(s),\end{aligned}$$

where the first equality is because $\mathcal{K}$ is the triangulation associated with $\mathrm{conc}_D(f)$, the second equality is by definition of f, the third equality is by interchanging the summations, the fourth equality is by the definition of $\chi_j^{K_i}(s)$ and the strict inequality is because for $j'\neq j$, $\mathrm{conc}_D(f_{j'})(s)\geq \chi_{j'}^{K_i}(s)$ and we have chosen s so that $\mathrm{conc}_D(f_j)(s)>\chi_j^{K_i}(s)$. $\square$

Acknowledgments. We thank the anonymous referees for the suggestions that helped us improve the presentation of the paper. The research of the second author was funded in part by NSF CMMI 1727989.

References

[1] Al-Khayyal FA, Falk JE (1983) Jointly constrained biconvex programming. *Mathematics of Operations Research* 8(2):273–286.

[2] Balas E, Ceria S, Cornuéjols G (1993) A lift-and-project cutting plane algorithm for mixed 0–1 programs. *Mathematical Programming* 58:295–324.

[3] Belotti P, Lee J, Liberti L, Margot F, Wächter A (2009) Branching and bounds tightening techniques for non-convex MINLP. *Optimization Methods & Software* 24(4-5):597–634.

[4] Benson HP (2004) Concave envelopes of monomial functions over rectangles. *Naval Research Logistics* 51(4):467–476.

[5] Billingsley P (2008) *Probability and measure* (John Wiley & Sons).

[6] Boland N, Dey SS, Kalinowski T, Molinaro M, Rigterink F (2017) Bounding the gap between the McCormick relaxation and the convex hull for bilinear functions. *Mathematical Programming* 162(1-2):523–535.

[7] Cafieri S, Lee J, Liberti L (2010) On convex relaxations of quadrilinear terms. *Journal of Global Optimization* 47(4):661–685.

[8] Crama Y (1989) Recognition problems for special classes of polynomials in 0–1 variables. *Mathematical Programming* 44(1):139–155.

[9] De Loera J, Rambau J, Santos F (2010) *Triangulations: Structures for Algorithms and Applications*. Algorithms and Computation in Mathematics (Springer Berlin Heidelberg).

[10] Del Pia A, Khajavirad A (2018) The multilinear polytope for acyclic hypergraphs. *SIAM Journal on Optimization* 28(2):1049–1076.

[11] Epstein LG, Tanny SM (1980) Increasing generalized correlation: a definition and some economic consequences. *Canadian Journal of Economics* 16–34.

[12] Fan K, Lorentz G (1954) An integral inequality. *The American Mathematical Monthly* 61(9):626–631.

[13] Fréchet M (1951) Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, serie 3, Sciences, Sect. A* 14:53–77.

[14] Graham RL (1972) An efficient algorithm for determining the convex hull of a finite planar set. *Information Processing Letters* 1(4):132–133.

[15] Gupte A, Kalinowski T, Rigterink F, Waterer H (2020) Extended formulations for convex hulls of some bilinear functions. *Discrete Optimization* 36:100569.

[16] Harary F, et al. (1953) On the notion of balance of a signed graph. *The Michigan Mathematical Journal* 2(2):143–146.

[17] He T, Tawarmalani M (2020) A new framework to relax composite functions in nonlinear programs. *Mathematical Programming* URL http://dx.doi.org/10.1007/s10107-020-01541-x.

[18] Höffding W (1940) Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und Instituts fur Angewandte Mathematik der Universitat Berlin* 5:181–233.

[19] Knuth DE (1997) *The art of computer programming*, volume 3 (Pearson Education).

[20] Lasserre JB (2001) Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization* 11(3):796–817.

[21] Lorentz G (1953) An inequality for rearrangements. *The American Mathematical Monthly* 60(3):176–179.

[22] Lovász L (1983) Submodular functions and convexity. *Mathematical Programming The State of the Art*, 235–257 (Springer).

[23] McCormick GP (1976) Computability of global solutions to factorable nonconvex programs: Part i — Convex underestimating problems. *Mathematical Programming* 10(1):147–175.

[24] Meyer CA, Floudas CA (2004) Trilinear monomials with mixed sign domains: Facets of the convex and concave envelopes. *Journal of Global Optimization* 29(2):125–155.

[25] Misener R, Floudas CA (2014) ANTIGONE: Algorithms for continuous/integer global optimization of nonlinear equations. *Journal Global Optimization* 59(2-3):503–526.

[26] Misener R, Smadbeck JB, Floudas CA (2015) Dynamically generated cutting planes for mixed-integer quadratically constrained quadratic programs and their incorporation into GloMIQO 2. *Optimization Methods and Software* 30(1):215–249.

[27] Padberg M (1989) The boolean quadric polytope: some characteristics, facets and relatives. *Mathematical programming* 45(1):139–172.

[28] Parrilo PA (2003) Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming* 96(2):293–320.

[29] Rachev ST, Rüschendorf L (1998) *Mass Transportation Problems: Volume I: Theory* (Springer Science & Business Media).

[30] Rikun AD (1997) A convex envelope formula for multilinear functions. *Journal of Global Optimization* 10(4):425–437.

[31] Rockafellar RT (2015) *Convex analysis* (Princeton university press).

[32] Rockafellar RT, Royset JO (2014) Random variables, monotone relations, and convex analysis. *Mathematical Programming* 148(1-2):297–331.

[33] Seacrest T, Su FE (2018) A lower bound technique for triangulations of simplotopes. *SIAM Journal on Discrete Mathematics* 32(1):1–28.

[34] Shaked M, Shanthikumar JG (2007) *Stochastic orders* (Springer Science & Business Media).

[35] Sherali HD (1997) Convex envelopes of multilinear functions over a unit hypercube and over special discrete sets. *Acta mathematica vietnamica* 22(1):245–270.

[36] Sherali HD, Adams WP (1990) A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics* 3(3):411–430.

[37] Tardella F (2008) Existence and sum decomposition of vertex polyhedral convex envelopes. *Optim. Lett.* 2(3):363–375.

[38] Tawarmalani M (2010) Inclusion certificates and simultaneous convexification of functions, URL http://www.optimization-online.org/DB_HTML/2010/09/2722.html, working paper.

[39] Tawarmalani M, Richard JPP, Xiong C (2013) Explicit convex and concave envelopes through polyhedral subdivisions. *Mathematical Programming* 138(1-2):531–577.

[40] Tawarmalani M, Sahinidis NV (2002) Convex extensions and envelopes of lower semi-continuous functions. *Mathematical Programming* 93(2):247–263.

[41] Tawarmalani M, Sahinidis NV (2005) A polyhedral branch-and-cut approach to global optimization. *Mathematical Programming* 103(2):225–249.

[42] Tchen AH (1980) Inequalities for distributions with given marginals. *The Annals of Probability* 814–827.

[43] Topkis DM (2011) *Supermodularity and complementarity* (Princeton university press).

[44] Vigerske S, Gleixner A (2018) SCIP: global optimization of mixed-integer nonlinear programs in a branch-and-cut framework. *Optimization Methods and Software* 33(3):563–593.

[45] Villani C (2008) *Optimal transport, old and new* (Berlin: Springer).