Inclusion Certificates and Simultaneous Convexification of Functions

Mohit Tawarmalani*

September 5, 2010

Abstract

We define the inclusion certificate as a measure that expresses each point in the domain of a collection of functions as a convex combination of other points in the domain. Using inclusion certificates, we extend the convex extensions theory to enable simultaneous convexification of functions. We discuss conditions under which the domain of the functions can be reduced without affecting the simultaneous convex hull of the functions. For example, we show that a collection of multilinear functions can be convexified over a cartesian product of compact convex sets by restricting attention to the extreme points of the domain. We relate disjunctive programming to convex extensions theory and provide insights, generalizations, and shorter proofs for convexifying orthogonal disjunctions. Each convex envelope is naturally associated with an inclusion certificate. When the inclusion certificate matches across a collection of functions, we show that convexifying individual functions yields the simultaneous convex hull of a collection of functions. In particular, we show that it is easy to sequentially convexify submodular functions one variable at a time. Further, a collection of submodular functions is convexified by individually convexifying the functions. Various other ways of composing functions and their associated convex envelopes are also discussed.

1 Introduction

Branch-and-bound algorithms for global optimization rely on relaxations for nonconvex programs. In particular, an inequality of the form $f(x) \le 0$ (resp. $f(x) \ge 0$) is replaced by $\check{f}(x) \le 0$ (resp. $\hat{f}(x) \ge 0$) where $\check{f}$ (resp. $\hat{f}$) is a convex underestimator (resp. concave overestimator) of $f(x)$ over an outer-approximation of the feasible values for x, typically chosen to be a hypercube. Since such a relaxation yields a convex programming problem, it is solved in branch-and-bound algorithms to find provable lower bounds for global minimization problems. The best such underestimator (resp. overestimator) is called the convex (resp. concave) envelope of the associated function.

Finding the convex and/or the concave envelope of a nonlinear function is in general NP-Hard [11]. However, envelopes for special classes of functions have been found; see for example [1, 12, 25, 30, 29, 10, 6, 19, 4, 20, 28] for analysis of multilinear functions and [21, 38, 7, 22, 34, 16, 37] for other types of functions. We refer the reader to [13] for a survey of some of these techniques. In particular, [37] shows that many of the convex envelope formulae in [1, 12, 25, 30, 6, 10, 22, 28] can be unified under a common result concerning the convexification of convex-extendable submodular functions. This work also develops a technique to convexify certain disjunctive functions in the space of the original variables. By definition, the best convex relaxation for the epigraph of $f(x)$, $f(x) \le z$, is obtained by replacing $f(x)$ with its convex envelope. However, consider the set $\{(x, y, z) \mid z_1 \le \frac{x}{y},\ z_2 \le xy,\ (x, y) \in [1, 2]^2\}$. We will show later that $2z_1 + z_2 \le 3x$ is a valid inequality for this set. Presently, this can also be verified by noticing that $\frac{2x}{y} + xy - 3x = \frac{x}{y}(y - 2)(y - 1) \le 0$ for $x \ge 0$ and $y \in [1, 2]$. However, this inequality is not implied by relaxing the set to $\{(x, y, z) \mid z_1 \le f_1(x, y),\ z_2 \le f_2(x, y),\ (x, y) \in [1, 2]^2\}$ where $f_1(x, y)$ (resp. $f_2(x, y)$) are the concave envelopes of $\frac{x}{y}$ (resp. xy). In other words, finding the individual envelopes does not provide the same strength as simultaneous convexification of functions. Despite the potential of improving the strength of relaxations, the literature on simultaneous convexification tools is sparse, especially when the associated variables are continuous. In particular, [23, 32] study simultaneous convexification of bilinear functions of

*Krannert School of Management, Purdue University, 100 S. Grant Street, West Lafayette, IN 47907-2076, USA, mtawarma@purdue.edu.

0−1 variables. The seminal work on reformulation-linearization hierarchy, [31], studies the convexification of mathematical programs involving multilinear functions of 0−1 variables. In [35], the author proves that the convex hull of a systems of multilinear equations over a hypercube remains the same even if the variables are restricted to the extreme points of the hypercube. In [8, 2], the authors study the simultaneous convexification of quadratic functions of variables over a hypercube. In [17, 24], the authors consider convexification of mathematical programs with polynomial functions and develop infinite hierarchies of semidefinite relaxations that converge to the convex hull of the problem.

This paper extends convex extensions theory, initially proposed and studied as a convexification tool for epigraphs of individual functions (see [12, 25, 38, 39, 16]) to enable simultaneous convexification of functions. The essential ideas in this literature can be summarized as follows. First, the technique reduces the domain of the function such that the function can be expressed as a convex function over a disjunctive union of convex sets. Then, the technique proceeds in one of two ways. Either disjunctive programming is used to express the convex hull in a higher-dimensional space or a convex optimization problem is solved to find inequalities in the original variables that separate the convex hull of the epigraph from a given point. In Section 2, we extend the convex extensions theory to simultaneous convexification of functions generalizing the earlier results in [25, 39]. As an example, we show that a system of multilinear functions over a cartesian product of convex sets can be convexified by restricting attention to the extreme points of the domain. Then, we provide a short proof of the main result in orthogonal disjunctions theory (see [36]) and also relate the more general case of disjunctive programming to the convex extensions viewpoint. In Section 3.1, we develop a common unifying framework for the two primary constructions of [37] regarding the convex envelope of submodular functions and the convex envelope of certain disjunctive functions. This framework allows us to prove later that individual convex envelopes for these functions over a hypercube yield the simultaneous convex hull. In Section 3.2, we define the inclusion certificate as a way to express each point in the domain as a convex combination of points in the domain. For a given function, we consider underestimators/overestimators associated with an inclusion certificate. We show that when the convex envelopes of a collection of functions share the same inclusion certificate the functions are convexified simultaneously by convexifying the individual functions. We consider various other function compositions and provide results that enable construction of their simultaneous convex hulls. These results generalize the previously mentioned convex extendability of multilinear functions from the extreme points of the domain. We also show that our results generalize earlier results in [25, 22] that provide conditions under which the convex envelope of a sum of functions is the sum of their convex envelopes.

2 Simultaneous Convex Extensions

Consider the problem of convexifying $f(x) : P \mapsto \mathbb{R}$ over $\text{conv}(P)$, which is assumed to be compact. A technique that is useful in performing this convexification involves eliminating points x' from P if it can be shown that the vertices of the epigraph of f do not project to x'. We extend these ideas to address simultaneous convexification of functions. We also relate disjunctive programming to convex extensions.

2.1 Theoretical Framework

We denote the extreme points (resp. exposed points) of a convex set C as $\text{extr}(C)$ (resp. $\text{exps}(C)$), the projection of C to the space of x variables as $\text{proj}_x(C)$, the convex hull of C as $\text{conv}(C)$ and the closure of C as $\text{cl}(C)$. For a function $g(x)$, we denote its convex envelope over a convex set C by $\text{conv}_C(g)$ and its concave envelope by $\text{conc}_S(g)$. For a set C, the horizon cone C^∞ is defined as (see [27]):

$$C^\infty = \begin{cases} \{x \mid \exists x^\nu \in C, \lambda^\nu \downarrow 0, \text{ with } \lambda^\nu x^\nu \to x\} & \text{if } C \neq \emptyset \\ \{0\} & \text{otherwise.} \end{cases}$$

An indicator function of set C is defined in a standard way:

$$\delta(x \mid C) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

We denote the support function of C by $\delta^*(\cdot \mid C)$ and $[-\infty, \infty]$ by $\bar{\mathbb{R}}$. If $v_1, \dots, v_m$ are points in $\mathbb{R}^n$, $V = (v_i)_{i=1}^m$, and $g : \mathbb{R}^n \mapsto \mathbb{R}$ then we will denote $(g(v_1), \dots, g(v_m))$ by $g(V)$. For a function $g(x)$ we will say that the function is convex-extendable from a set C if restricting the domain of g to C does not alter the convex envelope. Unless specified otherwise, e will denote a vector of all ones. We will often use the same notation to denote a finite set of points and a matrix whose columns are the corresponding points.

Let $x \in \mathbb{R}^n$. Consider a non-empty closed set S^P of the following form: $S^P = \{(x,z) \mid H(x) \le z \le F(x), x \in P\}$, where, for $i \in I$, $h_i(x) : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ and $f_i(x) : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$, $z \in \mathbb{R}^{|I|}$, $H = (h_1, \dots, h_{|I|})$, $F = (f_1, \dots, f_{|I|})$, and P is a compact set. We assume that, for each i, either $h_i(x)$ (resp. $f_i(x)$) is $-\infty$ (resp. ∞) everywhere or it has an affine minorant (resp. majorant). We are interested in computing $\operatorname{conv}(S^P)$. Without loss of generality, we assume that, for each i, there exists an $x \in P$ such that either $h_i(x) > -\infty$ or $f_i(x) < +\infty$. Otherwise, S^P is a cartesian product of the real line with a smaller set and we may convexify the smaller set instead. We provide the main tool for convexifying S^P in Theorem 2.1. The result helps in constructing the convex hull by reducing the size of P. In particular, it identifies a subset of P, whose elements cannot be lifted to an extreme point of S^P. This result generalizes existing results in the literature concerning convex extensions of functions by allowing many functions to be treated simultaneously (see, for example, Theorem 1.2 in [25] and Theorem 7 and Corollary 5 in [39]).

Theorem 2.1. *Let $X \subseteq P$ be such that for each $x^0 \in X$, there exists a $\lambda(x^0) \in \mathbb{R}^m$ and $V(x^0) \subseteq P \backslash \{x^0\}$, where $|V(x^0)| = m$, that satisfy the following conditions:*

1. $H(V(x^0)) \ge F(V(x^0))$
2. $V(x^0)\lambda(x^0) = x^0$, $e^T\lambda(x^0) = 1$, $\lambda(x^0) \ge 0$,
3. $F(x^0) \le F(V(x^0))\lambda(x^0)$ *and* $H(x^0) \ge H(V(x^0))\lambda(x^0)$.

Then, $\operatorname{cl}\operatorname{conv}(S^P) = \operatorname{conv}(S^P) = \operatorname{conv}(S^{P\backslash X})$. *Further,* $\operatorname{cl}\operatorname{conv}(S^P)$ *does not contain any lines and the projection of* $\operatorname{extr}(\operatorname{conv}(S^P))$ *to the space of x variables does not intersect with X.*

Proof. Let $h_i^l(x)$ be the affine minorant of $h_i(x)$ and $f_i^l(x)$ be the affine majorant of $f_i(x)$. We assume $h_i^l(x) = -\infty$ if $h_i(x) = -\infty$ and $f_i^l(x) = \infty$ if $f_i(x) = \infty$. Let $q_i = 0$ (resp. $-\infty$) if $h_i(x)$ is greater than (resp. equal to) $-\infty$. Similarly, let $r_i = 0$ (resp. $+\infty$) if $f_i(x)$ is less than (resp. equal to) ∞.

First, we identify $S^{P\infty}$ as $\{(0,z) \mid q \le z \le r\}$ and, as a result, establish that $S^{P\infty}$ is pointed (since q_i and r_i are not $-\infty$ and $+\infty$ simultaneously). Clearly, $\operatorname{cl}\operatorname{conv}(S^P)$ can be outer-approximated by the closed convex set $S' = \{(x,z) \mid h_i^l(x) \le z \le f_i^l(x), x \in \operatorname{conv}(P)\}$. If $(x',z') \in S'^{\infty}$, then by compactness of P, and hence of $\operatorname{conv}(P)$, it follows that $x' = 0$. Clearly, $q \le z' \le r$. Otherwise, for any $(x,z) \in S^P$, it follows that $(x, z+\alpha z') \notin S'$ for a suitably large α. By Theorem 3.6 in [27], $S'^{\infty} \subseteq \{(0,z) \mid q \le z \le r\}$. It follows that $S^{P\infty} \supseteq \{(0,z) \mid q \le z \le r\} \supseteq S'^{\infty} \supseteq \operatorname{cl}\operatorname{conv}(S^P)^{\infty} \supseteq S^{P\infty}$, where the first inclusion follows since S^P is non-empty, second inclusion follows from the argument above, and the remaining inclusions follow since $S' \supseteq \operatorname{cl}\operatorname{conv}(S^P) \supseteq S^P$. Therefore, equality holds throughout.

Second, we show that $\operatorname{conv}(S^P)$ is closed and does not contain any lines. Observe that $\operatorname{cl}\operatorname{conv}(S^P) = \operatorname{conv}(S^P) + \operatorname{conv}(S^{P\infty}) = \operatorname{conv}(S^P + S^{P\infty}) = \operatorname{conv}(S^P)$, where the first equality follows from Corollary 3.46 in [27] because S^P is closed and we have shown that $S^{P\infty}$ is pointed, the second equality since for any sets A and B, $\operatorname{conv}(A+B) = \operatorname{conv}(A) + \operatorname{conv}(B)$, and the third equality since $S^P = S^P + S^{P\infty}$. Again by Corollary 3.46 in [27], $\operatorname{cl}\operatorname{conv}(S^P)^{\infty} = \operatorname{conv}(S^{P\infty}) = S^{P\infty}$. By the pointedness of $S^{P\infty}$, $\operatorname{cl}\operatorname{conv}(S^P)$ has no lines.

Third, we show that $X \subsetneq P$ and, in particular, that none of the extreme points of $\operatorname{cl}\operatorname{conv}(S^P)$ project to X. Since $\operatorname{cl}\operatorname{conv}(S^P)$ does not contain any lines, it follows from Corollary 18.5.3 in [26] that $\operatorname{cl}\operatorname{conv}(S^P)$ contains an extreme point, say (x',z'). Since $\operatorname{cl}\operatorname{conv}(S^P) = \operatorname{conv}(S^P)$, it follows easily that $(x',z') \in S^P$ (see, for example, Corollary 18.3.1 in [26]). Assume that x' satisfies the hypotheses for inclusion in X. Then, $H(V(x'))\lambda(x') \le H(x') \le z' \le F(x') \le F(V(x'))\lambda(x')$. Therefore, there exists a $\gamma \in [0,1]^{|I|}$ such that $z'_i = (\gamma_i f_i(V(x')) + (1-\gamma_i)h_i(V(x')))\,\lambda(x')$. Let $Z_i(x') = \gamma_i f_i(V(x')) + (1-\gamma_i)h_i(V(x'))$ and $Z(x') = (Z_1(x')^T, \dots, Z_{|I|}(x')^T)^T$. Then, it follows that $(x',z')^T = \begin{pmatrix} V(x') \\ Z(x') \end{pmatrix} \lambda(x')$. Since $V(x') \subseteq P \backslash \{x'\}$ and $\begin{pmatrix} V(x') \\ Z(x') \end{pmatrix} \subseteq S^P$ because $H(V(x')) \le F(V(x'))$, the inclusion of x' in X contradicts the extremality of (x',z') in $\operatorname{cl}\operatorname{conv}(S^P)$. Therefore, $x' \notin X$ and $S^{P\backslash X}$ is not empty.

Finally, we show that $\operatorname{conv}(S^P) = \operatorname{conv}(S^{P\backslash X})$. Since $S^{P\backslash X} \neq \emptyset$, $S^{P\backslash X\infty} = S^{P\infty}$. Therefore, $S^{P\backslash X}$ contains all the extreme points of $\operatorname{cl}\operatorname{conv}(S^P)$ and $S^{P\backslash X\infty}$ contains the recession directions and, hence, the extreme directions of $\operatorname{cl}\operatorname{conv}(S^P)$. It follows by Theorem 18.5 in [26], closedness of $\operatorname{conv}(S^P)$, and $S^{P\setminus X} = S^{P\setminus X} + S^{P\setminus X\infty}$, that $\operatorname{conv}(S^P) = \operatorname{conv}(S^{P\setminus X})$. □

The intuition for Theorem 2.1 is the following. The first condition in the statement of the result ensures that $V(x^0)$ can be lifted to feasible points in S^P. The second and third conditions guarantee that $S(x^0) = \{(x^0, z) \mid H(x^0) \le z \le F(x^0)\}$, the slice of S^P through x^0, is contained in the convex hull of the aforementioned liftings of $V(x^0)$ that belong to S^P. Then, the result states that x^0 can be removed from P without affecting the closure convex hull of S^P. Indeed, the case in Theorem 2.1 where, for all i, $-\infty < \inf_x h_i(x)$ and $\sup_x f_i(x) < \infty$ can be proved much more easily. In this case, S^P is compact. Therefore, $\operatorname{cl}\operatorname{conv}(S^P) = \operatorname{conv}(S^P)$ (see Corollary 2.30 in [27]) and, by definition, $S^{P\infty} = \emptyset$. Since the convex hull of a compact set is the convex hull of its extreme points (see Theorem 18.5 in [26]), it follows that any point that can be expressed as a convex combination of two other points can be removed from S^P without changing $\operatorname{conv}(S^P)$. In contrast, the proof of Theorem 2.1 allows for unbounded sets. We give a slight generalization of Theorem 2.1 next, where $V(x^0)$ and $\lambda(x^0)$ do not only depend on x^0 but also on the corresponding value of z variables.

Corollary 2.2. *Let $X \subseteq P$ be such that for each $x^0 \in X$ and $z^0 \in \{z \mid F(x^0) \le z \le H(x^0)\}$ there exists a $\lambda(x^0, z^0) \in \mathbb{R}^m_+$ and $V(x^0, z^0) \subseteq S^{P\setminus\{x^0\}}$, where $|V(x^0, z^0)| = m$, such that $e^T\lambda(x^0, z^0) = 1$. Then, $\operatorname{cl}\operatorname{conv}(S^P) = \operatorname{conv}(S^P) = \operatorname{conv}(S^{P\setminus X})$. Further, $\operatorname{cl}\operatorname{conv}(S^P)$ does not contain any lines and the projection of $\operatorname{extr}(\operatorname{conv}(S^P))$ to the space of x variables does not intersect with X.*

Proof. Note that any x^0 in X cannot be lifted to an extreme point of S^P since each lifting of x^0 can be expressed as a convex combination of points in S^P. □

It may be noted that the closedness of S^P cannot be relaxed. For example, consider $S^{[-1,1]} = \{(x, z) \mid -1 \le x \le 1, z \ge h(x)\}$, where $h(x) = 1$ if $x = 0$ and $|x|$ otherwise. It is easy to verify that for any $x \in (-1, 1)$, one can express $h(x) \ge \frac{1}{2}h(x + \epsilon) + \frac{1}{2}h(x - \epsilon)$ for small enough $\epsilon > 0$. However, $\operatorname{cl}\operatorname{conv}\big(S^{\{-1,1\}}\big) = \{(x, z) \mid -1 \le x \le 1, z \ge 1\} \subsetneq \{(x, z) \mid -1 \le x \le 1, z \ge |x|\} = \operatorname{cl}\operatorname{conv}\big(S^{[-1,1]}\big)$.

If P is not compact, $\operatorname{conv}(S^P)$ may not be closed even though S^P is closed. As an example, consider $S = \{(x, y, z) \mid z = xy, x \in \{0, 1\}\}$. Then, it can be easily verified that $(1, y, 0) \notin \operatorname{cl}\operatorname{conv}(S)$ for $y \ne 0$. However, $(1-\epsilon, y, 0) = \epsilon\left(0, \frac{y}{\epsilon}, 0\right) + (1-\epsilon)(1, 0, 0) \in \operatorname{conv}(S)$ for each $\epsilon > 0$. Therefore, $(1, y, 0) \in \operatorname{cl}\operatorname{conv}(S)\setminus\operatorname{conv}(S)$. Still, Theorem 2.1 can be used to show that for $S' = \{(x, y, z) \mid z = xy, x \in [0, 1]\}$, $\operatorname{conv}(S') = \operatorname{conv}(S)$. This is because one can apply Theorem 2.1 after fixing y. Note that Theorem 2.1 and Corollary 2.2 allow an $x^0 \in X$ to be used to expand X to include other points. Sometimes the following much simpler argument suffices.

Remark 2.3. *Clearly, if $S' \subseteq S \subseteq \operatorname{conv}(S')$ then $\operatorname{conv}(S) = \operatorname{conv}(S')$. Therefore, if one can express points in S^P as a convex combination of points in some set S' then $\operatorname{conv}(S^P) = \operatorname{conv}(S')$. Although the technique can again be applied to S', in contrast to Theorem 2.1, points in $S^P\setminus S'$ cannot be used to eliminate points in S'.*

Now, we provide a dual version of the exclusion argument in Theorem 2.1.

Corollary 2.4. *Assume $H(x) \le F(x)$ for all $x \in P$. Let $X \subseteq P$. If $x' \in X$, assume that there does not exist any $(\alpha, \beta) \in \mathbb{R}^{n+|I|}$ such that x' is the unique maximizer of*

$$L^P(\alpha, \beta): \qquad \sup\{\langle \alpha, x\rangle + \langle \beta^+, F(x)\rangle + \langle \beta^-, H(x)\rangle \mid x \in P\}.$$

Here, $\beta_i^+ = \max\{\beta_i, 0\}$ and $\beta_i^- = \min\{\beta_i, 0\}$. Then, $\operatorname{conv}(S^P) = \operatorname{cl}\operatorname{conv}(S^{P\setminus X})$. More generally, if for all (α, β) that have a unique minimizer in $L^P(\alpha, \beta)$, it is true that $L^P(\alpha, \beta) = L^{P\setminus X}(\alpha, \beta)$ then $\operatorname{conv}(S^P) = \operatorname{cl}\operatorname{conv}(S^{P\setminus X})$.

Proof. We show that $\operatorname{exps}(\operatorname{cl}\operatorname{conv}(S^P)) \subseteq S^{P\setminus X}$. Let (x', z') be an exposed point of $\operatorname{cl}\operatorname{conv}(S^P)$. Then, it is the unique maximizer for some (α', β') in $\delta^*(\cdot \mid \operatorname{cl}\operatorname{conv}(S^P))$. However, by Theorem 2.1, $(x', z') \in S^P$ and is, therefore, the unique maximizer for the support function $\delta^*((\alpha', \beta') \mid S^P)$. If $\beta_i' = 0$ and $h_i(x') < f_i(x')$ or if $h_i(x') < z_i' < f_i(x')$ then (x', z') cannot be the unique maximizer. Therefore, $z_i' \in \{h_i(x'), f_i(x')\}$. If $z_i' = h_i(x')$ (resp. $z_i' = f_i(x')$) then $\beta_i' \le 0$ (resp. $\beta_i' \ge 0$). Since $H(x) \le F(x)$ for all $x \in P$, it follows easily that x' is the unique maximizer of $L^P(\alpha', \beta')$ and, therefore, $x' \notin X$, $(x', z') \notin S^X$, and $\exp\mathrm{s}(\mathrm{cl}\,\mathrm{conv}(S^P)) \subseteq S^{P\setminus X}$. By Strasewicz's theorem and $H(x) \le F(x)$, $S^{P\setminus X} \ne \emptyset$. By Theorem 2.1, $S^{P\setminus X\infty} = S^{P\infty}$. Consequently, by Theorem 2.1, Theorem 18.7 in [26], and $S^{P\setminus X} + S^{P\setminus X\infty} = S^{P\setminus X}$, it follows that $\mathrm{conv}(S^P) = \mathrm{cl}\,\mathrm{conv}(S^{P\setminus X})$.

We now consider the last statement in the result. Let (x', z') be an exposed point of $\mathrm{cl}\,\mathrm{conv}(S^P)$ and therefore the unique optimal solution to $\delta^*((\alpha', \beta') \mid \mathrm{cl}\,\mathrm{conv}(S^P))$ for some (α', β'). We show that any other sequence (x^ν, z^ν) that attains $\delta^*((\alpha', \beta') \mid \mathrm{cl}\,\mathrm{conv}(S^P))$ in the limit must converge to (x', z'). Assume for deriving a contradiction that this is not true. Since (x', z') is the unique maximizer in $\delta^*((\alpha', \beta') \mid \mathrm{cl}\,\mathrm{conv}(S^P))$, (x^ν, z^ν) must not have an accumulation point. Consider $B' = (x', z') + B$, where B is a unit ball and consider the subsequence $(x^{\nu'}, z^{\nu'})$ that does not belong to B'. By convexity of $\mathrm{cl}\,\mathrm{conv}(S^P)$, it follows that the line segment joining (x', z') and $(x^{\nu'}, z^{\nu'})$ meets $\delta B'$ (the boundary of B') at some point, say $(\bar{x}^{\nu'}, \bar{z}^{\nu'})$, which belongs to $\mathrm{cl}\,\mathrm{conv}(S^P)$. Since $\delta B' \cap \mathrm{cl}\,\mathrm{conv}(S^P)$ is compact, a subsequence of $(\bar{x}^{\nu'}, \bar{z}^{\nu'})$ converges to a point different than (x', z') that must be optimal for $\delta^*((\alpha', \beta') \mid \mathrm{cl}\,\mathrm{conv}(S^P))$. By our assumption $L^{P\setminus X}(\alpha', \beta') = L^P(\alpha', \beta') = \delta^*((\alpha', \beta') \mid \mathrm{cl}\,\mathrm{conv}(S^P))$. Then, consider the sequence x^ν that converges to $L^{P\setminus X}(\alpha', \beta')$. For each i such that $\beta'_i \le 0$, let $z_i^\nu = h_i(x^\nu)$ and let $z_i^\nu = f_i(x^\nu)$ otherwise. Then, it follows that $\langle \alpha', x^\nu \rangle + \langle \beta', z^\nu \rangle$ converges to $\delta^*((\alpha', \beta') \mid \mathrm{cl}\,\mathrm{conv}(S^P))$. Therefore, as before, (x^ν, z^ν) converges to (x', z'). In other words, $\exp\mathrm{s}(\mathrm{cl}\,\mathrm{conv}(S^P)) \subseteq \mathrm{cl}(S^{P\setminus X})$. As before, it follows that $\mathrm{conv}(S^P) = \mathrm{cl}\,\mathrm{conv}(S^{P\setminus X})$. □

We provide an intuition into Corollary 2.4 by discussing its relation to biconjugation. It is well known [14] that as long as a function $\phi(x)$ has an affine minorant $\mathrm{cl}(\mathrm{conv}(\mathrm{epi}(\phi)) = \mathrm{epi}(\phi^{**})$; see [15]. This leads to the following simple observation.

Remark 2.5. *Consider $\phi_1(x) : X_1 \mapsto \mathbb{R}$ and $\phi_2(x) : X_2 \mapsto \mathbb{R}$, where $X_1, X_2 \subseteq \mathbb{R}^n$. Assume that both ϕ_1 and ϕ_2 have affine minorants. If for all $\alpha \in \mathbb{R}^n$, $\phi_1^* = \sup_{x \in X_1}\{\langle x, \alpha\rangle - \phi_1(x)\} = \sup_{x\in X_2}\{\langle x, \alpha\rangle - \phi_2(x)\} = \phi_2^*$, then* $\mathrm{cl}(\mathrm{conv}(\phi_1)) = \mathrm{cl}(\mathrm{conv}(\phi_2))$. □

Let $\mathrm{dom}(\phi) = \{x \mid \phi(x) < \infty\}$. It can be easily verified that

$$\mathrm{conv}(\mathrm{dom}(\phi)) \subseteq \mathrm{dom}(\mathrm{cl}\,\mathrm{conv}(\phi)) \subseteq \mathrm{cl}(\mathrm{conv}(\mathrm{dom}(\phi))).$$

Therefore, $\mathrm{cl}(\mathrm{dom}(\mathrm{cl}\,\mathrm{conv}(\phi))) = \mathrm{cl}(\mathrm{conv}(\mathrm{dom}(\phi)))$. The case that relates to Corollary 2.4 is when ϕ_1 and ϕ_2 are indicator functions of X_1 and X_2 respectively. Indicator functions are non-negative and therefore have an affine minorant. Further, for an indicator function, ϕ, $\mathrm{dom}(\mathrm{cl}\,\mathrm{conv}(\phi))$ is already closed since $\mathrm{cl}\,\mathrm{conv}(\phi)(x) = 0$ if and only if $x \in \mathrm{cl}(\mathrm{conv}(\mathrm{dom}(\phi)))$. This is because for a set C, $\mathrm{cl}\,\delta(\cdot \mid \mathrm{conv}(C)) \le \delta(\cdot \mid \mathrm{cl}\,\mathrm{conv}(C)) \le \delta(\cdot \mid \mathrm{conv}(C))$ and the epigraph of $\delta(\cdot \mid \mathrm{cl}\,\mathrm{conv}(C))$ is closed. The following observation then follows from Remark 2.5.

Corollary 2.6. *Consider two sets X_1 and X_2 such that their support functions are identical. Then,* $\mathrm{cl}(\mathrm{conv}(X_1)) = \mathrm{cl}(\mathrm{conv}(X_2))$. □

Indeed, as is implicit in the proof of Corollary 2.4, $L^P(\alpha, \beta)$ is the support function of S^P if $H(x) \le F(x)$ for all $x \in P$. Therefore, it follows easily from Theorem 2.1 and Corollary 2.6 that if $L^P(\alpha, \beta) = L^{P\setminus X}(\alpha, \beta)$ for all (α, β) then $\mathrm{conv}(S^P) = \mathrm{cl}\,\mathrm{conv}(S^P) = \mathrm{cl}\,\mathrm{conv}(S^{P\setminus X})$. Corollary 2.4 allows one to check the equivalence of $L^P(\alpha, \beta)$ to $L^{P\setminus X}(\alpha, \beta)$ only for (α, β) that have unique maximizers.

We shall now demonstrate that the tools provided by Theorem 2.1, Corollary 2.4, and Corollary 2.6 yield generalizations of existing techniques to enable simultaneous convexification of many functions. To start with, we discuss an application of Corollary 2.6 for multilinear functions which was first presented in [35] and generalizes the polyhedrality of convex envelopes for individual multilinear functions discussed in [25]. In the following, we say a function $f(\prod_{i=1}^m x_i)$ is multilinear if, for each $i \in \{1, \ldots, m\}$, f is linear in x_i when $x_{i'}$ for each $i' \ne i$ is fixed at some value in its domain.

Corollary 2.7. *For each $i \in \{1, \ldots, m\}$, let $x_i \in \mathbb{R}^{n_i}$ and P_i be a compact convex set. Let $F = (f_1, \ldots, f_K)$. Assume that, for each k, $f_k(\prod_{i=1}^m x_i)$ is a multilinear function. For a set P, let M^P denote $\{(x, z) \mid x \in P,\ z_k = f_k(x_1, \ldots, x_m) \text{ for } k = 1, \ldots, K\}$. Then,* $\mathrm{conv}\left(M^{\prod_{i=1}^m P_i}\right) = \mathrm{conv}\left(M^{\prod_{i=1}^m \mathrm{extr}(P_i)}\right)$.

Proof. Consider the support function

$$\begin{aligned}
&\delta^*\left((\alpha,\beta) \mid M^{\prod_{i=1}^n P_i}\right) = \max\left\{\langle\alpha,x\rangle + \langle\beta,z\rangle \mid (x,z)\in M^{\prod_{i=1}^n P_i}\right\}\\
&= \max\left\{\langle\alpha,x\rangle + \left\langle\beta, F\left(\prod_{i=1}^m x_i\right)\right\rangle \;\middle|\; x_i\in P_i \text{ for } i=1,\ldots,m\right\}\\
&= \max\left\{\langle\alpha,x\rangle + \left\langle\beta, F\left(\prod_{i=1}^m x_i\right)\right\rangle \;\middle|\; x_i\in \operatorname{extr}(P_i) \text{ for } i=1,\ldots,m\right\}\\
&= \delta^*\left((\alpha,\beta) \mid M^{\prod_{i=1}^n \operatorname{extr}(P_i)}\right),
\end{aligned}$$

where the third equality follows inductively by fixing all x_i except one and observing that a linear function is optimized at an extreme point over a compact convex set; see Theorem 32.3 in [26]. Since the convex hull of a compact set is compact, the result then follows from Corollary 2.6. □

The above result implies that the reformulation-linearization hierarchy (see [31]) generates the convex hull of multilinear expressions of n variables over a hypercube at the n^{th} level. The proof of Corollary 2.7 extends to simultaneous convexification of epigraphs of functions that are non-negative and concave when the other variables are fixed, thus generalizing various results in the literature including Remark 1.4 in [25], Theorem 10 in [39], and Proposition 5 in [8]. Later, we will generalize Corollary 2.7 using Theorem 2.1 in a way that also contains the extension to non-negative concave functions.

Example 2.8. *Consider* $S = \big\{(x_1,x_2,x_3,z_{12},z_{13},z_{23}) \;\big|\; z_{12}=x_1x_2,\ z_{13}=x_1x_3,\ z_{23}=x_2x_3,\ (x_1,x_2,x_3)\in[0,1]^3\big\}$*. Then,* $\operatorname{conv}(S)$ *is given by the boolean quadric polytope:*

$$\left\{(x,z)\;\middle|\;\begin{array}{ll} z_{ij}\le\min\{x_i,x_j\} & i<j\\ z_{ij}\ge\max\{0,x_i+x_j-1\} & i<j\\ z_{12}+z_{13}\le x_1+z_{23} & \\ z_{12}+z_{23}\le x_2+z_{13} & \\ z_{13}+z_{23}\le x_3+z_{12} & \\ x_1+x_2+x_3\le 1+z_{12}+z_{13}+z_{23} & \end{array}\right\}.$$

Let $N=\{1,\ldots,n\}$ *and assume* $x\in\mathcal{H}=[-1,1]^n$*. This can be arranged when* x *belongs to a general hypercube by a linear transformation of the variables. For any set* $I\subseteq N$*, let* $m^I(x)=\prod_{i\in I}x_i$*. For* $J\subseteq N$*, we define* x^J *to be the extreme point of* $[-1,1]^n$ *such that* $x_j^J=1$ *if* $j\in J$ *and* $x_j^J=-1$ *otherwise. By Corollary 2.7,*

$$\operatorname{conv}\left\{z \mid z_I=m^I(x)\,\forall I\subseteq N,\ x\in\mathcal{H}\right\} = \operatorname{conv}\left\{z\;\middle|\; z\in\bigcup_{J\in 2^N}\left\{\left(m^I(x^J)\right)_{I\in 2^N}\right\}\right\}.$$

We denote $(m^I(x^J))_{J\in 2^N}$ *by* m^I*. For each variable* x_i *consider a random variable* X_i *that takes a value of* 1 *or* -1 *each with a probability* $\frac{1}{2}$*. Assume* X_i *and* X_j *are independent for* $i\neq j$*. Incidentally, it is easy to see that for* $I,I'\subseteq N$*, then* $\langle m^I,m^{I'}\rangle = 2^n\prod_{i\in I\cap I'}E[X_i^2]\prod_{i\in(I'\cup I)\setminus(I\cap I')}E[X_i]$*. Since, for all* i*,* $E[X_i]=0$ *and* $E[X_i^2]=1$*, the matrix* $\frac{1}{2^n}\left(m^I(x^J)\right)_{I\times J\in 2^N\times 2^N}$ *is orthogonal. (Its transpose is used as the design matrix in full factorial experimental designs.) Now, it follows easily that:*

$$\begin{aligned}
&\operatorname{conv}\left\{z\;\middle|\; z\in\bigcup_{J\in 2^N}\left\{\left(m^I(x^J)\right)_{I\in 2^N}\right\}\right\}\\
&= \operatorname{conv}\left\{z\;\middle|\;\sum_{I\in 2^N} m^I(x^J)z_I\ge 0\,\forall J\subseteq N,\ z_\emptyset=1\right\}.
\end{aligned}$$

Note that one can easily add other multilinear expressions to the above convex hull since any multilinear function can be written as a affine function of the monomials. □

2.2 Orthogonal Disjunctions and Disjunctive Programming

We now use Theorem 2.1 to interpret disjunctive programming. Towards this end, we first derive the main result of orthogonal disjunctions theory using Theorem 2.1; see [36]. We will then use Corollary 2.9 to relate the general case of disjunctive programming without orthogonality restrictions to Theorem 2.1.

Corollary 2.9. *Let $x_i \in \mathbb{R}^{n_i}$ and $C_i = \{x_i \mid -v(x_i,u_i) \le 1 \le t_i(x_i,u_i), w_i(x_i,u_i) \ge 0\}$ be convex sets for $i = 1,\ldots,k$. Assume that v_i, t_i, and w_i are positively-homogeneous functions and $C_i^\infty \supseteq \{(x_i \mid -v_i(x_i,u_i) \le 0 \le t_i(x_i,u_i), w_i(x_i,u_i) \ge 0\}$. Let $C'_i = \{(0,\ldots,0,x_i,0,\ldots,0) \mid x_i \in C_i\}$. Then, $\operatorname{conv}\left(\bigcup_{i=1}^k C'_i\right) \subseteq R \subseteq \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^k C'_i\right)$ where*

$$R = \left\{ x \;\middle|\; \begin{array}{l} \sum_{i=1}^k t_i(x_i,u_i) \ge 1 \\ \sum_{i=1}^k \min\{0, v_i(x_i,u_i)\} \ge -1 \\ t_i(x_i,u_i) + \min\{0, v_i(x_i,u_i)\} \ge 0,\ i=1,\ldots,k \\ t_i(x_i,u_i) \ge 0,\ i = 1,\ldots,k \\ w_i(x_i,u_i) \ge 0,\ i=1,\ldots,k \end{array} \right\}.$$

Further, if C'_i are closed, $\operatorname{cl}(R) = \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^k C'_i\right) = \operatorname{conv}\left(\bigcup_{i=1}^k C'_i\right) + \sum_{i=1}^k C_i'^\infty$.

Proof. First, we show that $R \subseteq \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^k C'_i\right)$. Choose an arbitrary $x \in R$. Then, $\max\{0, -v_i(x_i,u_i)\} \le t_i(x_i,u_i)$ and $\sum_{i=1}^k \max\{0,-v_i(x_i,u_i)\} \le 1 \le \sum_{i=1}^k t_i(x_i,u_i)$. Let $v = \sum_{i=1}^k \max\{0, -v_i(x_i,u_i)\}$ and $t = \sum_{i=1}^k t_i(x_i,u_i)$. If $t > v$, define $\lambda_i = \frac{t-1}{t-v}\max\{0,-v_i(x_i,u_i)\} + \frac{1-v}{t-v} t_i(x_i,u_i)$. Otherwise, let $\lambda_i = t_i(x_i,u_i)$. It can be easily verified that $\max\{0,-v_i(x_i,u_i)\} \le \lambda_i \le t_i(x_i,u_i)$, $\lambda_i \ge 0$, and $\sum_{i=1}^k \lambda_i = 1$. Let $I = \{i \mid \lambda_i > 0\}$ and $K = \{1,\ldots,k\}$ and note that $I \ne \emptyset$. By the definition of λ_i and positive-homogeneity of v_i, w_i, and t_i, it follows that whenever $i \in I$, $(x'_i,u'_i) = \left(\frac{x_i}{\lambda_i}, \frac{u_i}{\lambda_i}\right) \in C'_i$. On the other hand, if $i \in K\backslash I$ then $(x'_i,u'_i) = (x_i,u_i) \in C_i'^\infty$. Assume without loss of generality, by reordering indices if necessary, that $\lambda_1 = \max\{\lambda_i \mid i \in K\}$. Then, define $(x''_1,u''_1) = (x'_1,u'_1) + \sum_{i\in K\backslash I}(x'_i,u'_i)$ and $(x''_i,u''_i) = (x'_i,u'_i)$ for $i \in I\backslash\{1\}$. It follows that $(x,u) = \sum_{i\in I}\lambda_i(x''_i,u''_i)$ and for all $i \in I$, $(x''_i,u''_i) \in C'_i + \sum_{i'=1}^k C_i'^\infty$. Define $P'(x,u) = \bigcup_{i\in I}(x''_i,u''_i)$ and $P(x,u) = P'(x,u) \cup \{(x,u)\}$. Let $S = \bigcup_{i=1}^k \left(C'_i + \bigcup_{i'=1}^k C_{i'}'^\infty\right)$. Then,

$$\begin{aligned} \operatorname{conv}(R) &\subseteq \operatorname{conv}\left(\bigcup_{(x,u)\in R} \operatorname{conv}(P(x,u))\right) = \operatorname{conv}\left(\bigcup_{(x,u)\in R} P'(x,u)\right) \\ &\subseteq \operatorname{conv}(S) \subseteq \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^k C'_i\right), \end{aligned}$$

where the first inclusion follows from $(x,u) \in P(x,u)$, the equality holds because Theorem 2.1 applied to $\delta(\cdot \mid P(x,u))$ implies $\operatorname{conv}(P(x,u)) = \operatorname{conv}(P'(x,u))$, the second inclusion since $P'(x,u) \subseteq S$, and the last inclusion by Theorem 8.3 in [26]. Combining, $R \subseteq \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^k C'_i\right)$. Note that Theorem 2.1 can be replaced with the simpler Remark 2.3 in this case.

Now, we show that $R \supseteq \operatorname{conv}\left(\bigcup_{i=1}^k C'_i\right)$. Let $(x,u) \in \operatorname{conv}\left(\bigcup_{i=1}^k C'_i\right)$. Since C'_i is convex, it follows that there exists an $I \subseteq K$, $\{\lambda_i\}_{i\in I}$, and $\{(x'^i,u'^i)\}_{i\in I}$ such that (i) $\lambda_i > 0$ for all $i \in I$ and $\sum_{i\in I}\lambda_i = 1$, (ii) $(x'^i,u'^i) \in C'_i$ for all $i \in I$, and (iii) $(x,u) = \sum_{i\in I}\lambda_i(x'^i,u'^i)$. Note that for all $i \notin I$, $(x_i,u_i) = 0$ and therefore, by positive-homogeneity of t_i, v_i, and w_i, it follows that $t_i(x_i,u_i) = v_i(x_i,u_i) = w_i(x_i,u_i) = 0$. Combining these facts, the following holds:

$$\sum_{i=1}^k t_i(x_i,u_i) = \sum_{i\in I}\lambda_i t_i(x_i'^i, u_i'^i) \ge \sum_{i\in I}\lambda_i = 1.$$

Similarly, it can be shown that (x, u) satisfies the remaining constraints in R.

Note that $\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^k C_i'\right) \supseteq \operatorname{conv}\left(\bigcup_{i=1}^k C_i'\right) + \sum_{i=1}^k C_i'^{\infty} \supseteq \operatorname{conv}\left(\bigcup_{i=1}^k C_i'\right)$, where the first inclusion follows from Theorem 8.3 in [26] and the second inclusion from Theorem 8.1 in [26]. Therefore, we only need to show that $\operatorname{conv}\left(\bigcup_{i=1}^k C_i'\right) + \sum_{i=1}^k C_i'^{\infty}$ is closed. The following chain of equalities hold:

$$\begin{aligned}\operatorname{conv}\left(\bigcup_{i=1}^k C_i'\right) + \sum_{i=1}^k C_i'^{\infty} &= \operatorname{conv}\left(\bigcup_{i=1}^k C_i'\right) + \operatorname{conv}\left(\sum_{i=1}^k C_i'^{\infty}\right)\\ &= \operatorname{conv}\left(\bigcup_{i=1}^k C_i' + \sum_{i=1}^k C_i'^{\infty}\right) = \operatorname{conv}\left(\bigcup_{i=1}^k \left(C_i' + \sum_{i'=1}^k C_{i'}'^{\infty}\right)\right),\end{aligned}$$

where the first equality follows since $\sum_{i=1}^k C_i'^{\infty}$ is convex by Theorems 3.1 and 8.1 in [26], the second equality since for any sets A and B, $\operatorname{conv}(A+B) = \operatorname{conv}(A) + \operatorname{conv}(B)$, and the last equality since for sets A, B, and C, $(A \cup B) + C = (A + C) \cup (B + C)$. Then, the result follows by Corollary 9.8.1 in [26]. □

The expression for C_i in Corollary 2.9 only allows for one function each of the form v_i, t_i, and w_i. This assumption is, however, not restrictive since $\min\{f_1, f_2\}$ is a positively-homogeneous function if f_1 and f_2 are positively-homogeneous. Corollary 2.9 thus restates Theorem 1 of [36] and additionally characterizes $\operatorname{cl}\left(\bigcup_{i=1}^k C_i'\right)$ as $\operatorname{conv}\left(\bigcup_{i=1}^k C_i'\right) + \sum_{i=1}^k C_i'^{\infty}$.

We note that $\operatorname{conv}(C_1 \cup C_2) + C_1^{\infty} + C_2^{\infty}$ is not closed in general. Consider, for example, $C_1 = \{(x_1, x_2) \mid x_1x_2 \geq 1, x_1 \geq 0\}$ and $C_2 = \{(x_1, x_2) \mid x_1 \leq 0, x_2 = 2\}$. Then, $\operatorname{conv}(C_1 \cup C_2) + C_1^{\infty} + C_2^{\infty} = \{(x_1, x_2) \mid x_2 > 0\}$ which is not a closed set. The key that makes the characterization of $\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^k C_i'\right)$ possible in Corollary 2.9 is that the associated sets C_i' are orthogonal to each other. The assumption plays an important role in Corollary 2.9. In that sense, Corollary 2.9 only considers a special case of disjunctive programming. This special case is nevertheless interesting because the convex hull can be described in the space of the original problem variables. This was the motivation for its original presentation in [36]. However, we adopt a different view here and derive disjunctive programming technique as a consequence of Corollary 2.9. Consider the case where $C_1, \ldots, C_k$ are not necessarily orthogonal. Then, one can lift each C_i to C_i' to make the sets orthogonal. This requires creating replicas of variables that appear in multiple sets. For the sake of simplicity of notation, we assume that the sets share all the variables and, therefore, we define $C_i' = \{(0, \ldots, 0, x_i, 0, \ldots, 0 \mid x_i \in C_i\}$ where x_i is the i^{th} replica of the original variables. Then, Corollary 2.9 can be used to construct a set R that is sandwiched between $\operatorname{conv}\left(\bigcup_{i=1}^k C_i'\right)$ and $\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^k C_i'\right)$. Define $R' = \left\{y \mid y = \sum_{i=1}^k x_i, (x_i)_{i=1}^k \in R\right\}$. Let $\mathcal{C} = \bigcup_{i=1}^k C_i'$ and A be the linear transformation that maps $(x_i)_{i=1}^k$ to $\sum_{i=1}^k x_i$. Then, it follows that:

$$\operatorname{conv}(A\mathcal{C}) = A\operatorname{conv}(\mathcal{C}) \subseteq AR = R' \subseteq A\operatorname{cl}\operatorname{conv}(\mathcal{C}) \subseteq \operatorname{cl} A\operatorname{conv}(\mathcal{C}) = \operatorname{cl}\operatorname{conv}(A\mathcal{C})$$

where the first and last equalities follow since for any set C and a linear transformation A, $\operatorname{conv}(AC) = A\operatorname{conv}(C)$, first and second inclusion follow from Corollary 2.9, the second equality by the definition of R', and the last inclusion from Theorem 6.6 in [26]. Therefore, $\operatorname{conv}\left(\bigcup_{i=1}^k C_i\right) = \operatorname{conv}(A\mathcal{C}) \subseteq R' \subseteq \operatorname{cl}\operatorname{conv}(A\mathcal{C}) = \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^k C_i\right)$. In other words, by linearly transforming R we obtain a set, R', that is sandwiched between $\operatorname{conv}\left(\bigcup_{i=1}^k C_i\right)$ and $\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^k C_i\right)$. Therefore, disjunctive programming can be viewed as a technique where the convex multipliers are introduced to make the functions positively-homogeneous using right-scalar multiplication. Then, $C_i' = \{(0, \ldots, 0, x_i, 0, \ldots, 0, \lambda_i) \mid w_i(x_i, \lambda_i) \geq 0, \lambda_i \leq 1 \leq \lambda_i\}$ and, by applying Corollary 2.9,

$$R' = \left\{y \;\middle|\; y = \sum_{i=1}^k x_i,\ \sum_{i=1}^k \lambda_i = 1, w_i(x_i, \lambda_i) \geq 0\, \forall i\right\},$$

which matches what the disjunctive programming technique produces; see [3, 9] for details. The advantage of Corollary 2.9 is that if the provided functions are already positively-homogeneous then Corollary 2.9 describes the convex hull without introducing the convex multipliers.

Orthogonal disjunctions can also be related to Remark 2.3 without eliminating the convex multipliers. Define $K = \{1, \dots, k\}$ and

$$Q = \left\{ (x, z, p) \;\middle|\; \max\{0, -v_i(x_i)\} \le z_i \le t_i(x_i), i \in K,\ \sum_{i=1}^{k} z_i = p,\ w_i(x_i) \ge 0, i \in K \right\}.$$

The projection of Q to the space of (x, z) variables (viewed as a subspace with $p = 0$) is related to the set of our interest. The set Q and its projection (also their convex hulls) are related via the invertible linear transformation $(x, z, p) \mapsto \Big(x, z, p - \sum_{i=1}^{k} z_i\Big)$. Let $I = \{i \mid z_i > 0\}$ and $C_i(p) = \big\{(x_i, z_i) \;\big|\; \max\{0, -v_i(x_i)\} \le p \le t_i(x_i), z_i = p, w_i(x_i) \ge 0\big\}$. Define $C'_i = \{(x, z, p) \mid (x_i, z_i) \in C_i(p), (x_{i'}, z_{i'}) \in C_{i'}(0), i' \in K \backslash \{i\}\}$. Consider a point $(\bar{x}, \bar{z}, \bar{p}) \in Q$. If $\bar{p} = 0$, then for any i, $(\bar{x}, \bar{z}, \bar{p}) \in C'_i$. Now, assume that $\bar{p} > 0$. For each $i \in I$, define a point $(\hat{x}^i, \hat{z}^i, \hat{p}^i) \in C'_i$ as follows. Let $\hat{p}^i = \bar{p}$ and

$$(\hat{x}^i_{i'}, \hat{z}^i_{i'}) = \begin{cases} (0, 0) & \text{if } i' \in I \backslash \{i\} \\ (\bar{x}_{i'}, \bar{z}_{i'}) & \text{if } i' \notin I \\ \left(\dfrac{\bar{x}_{i'} \bar{p}}{\bar{z}_{i'}}, \bar{p} \right) & \text{if } i' = i. \end{cases}$$

Then, it can be verified easily that $(\bar{x}, \bar{z}, \bar{p}) = \sum_{i \in I} (\hat{x}^i, \hat{z}^i, \hat{p}^i) \frac{\bar{z}_i}{\bar{p}}$, where, for each $i \in I$, $\frac{\bar{z}_i}{\bar{p}} \ge 0$, and $\sum_{i \in I} \frac{\bar{z}_i}{\bar{p}} = 1$. Therefore, $Q \subseteq \operatorname{conv}\Big(\bigcup_{i=1}^{k} C'_i\Big)$. However, for each i, $C'_i \subseteq Q$. Therefore, as in Remark 2.3, $\operatorname{conv}(Q) = \operatorname{conv}\Big(\bigcup_{i=1}^{k} C'_i\Big)$.

3 Simultaneous vs. Individual Convexification

Let $E_i^P = \{(x, z) \mid z_i \ge h_i(x), x \in P\}$, where $i \in \{1, \dots, m\}$. Since $\bigcap_{i=1}^{m} \operatorname{conv}(E_i^P)$ is a convex superset of $\bigcap_{i=1}^{m} E_i^P$, it follows that $\operatorname{conv}\left(\bigcap_{i=1}^{m} E_i^P\right) \subseteq \bigcap_{i=1}^{m} \operatorname{conv}(E_i^P)$. Current global optimization software exploit this fact and relax $\bigcap_{i=1}^{m} E_i^P$ using the latter set instead of the former. We refer to the latter set as the set obtained from individual convexification and the former set as the one obtained from simultaneous convexification. It should not be surprising that $\operatorname{conv}\left(\bigcap_{i=1}^{m} E_i^P\right)$ is often a strict subset of $\bigcap_{i=1}^{m} \operatorname{conv}(E_i^P)$ because convexification operation does not, in general, distribute over intersection. We investigate the relationship between these sets in this section.

In [37], the authors recently proposed two techniques for individually convexifying functions in the space of the original variables. The first technique constructs convex envelopes of submodular functions that are convex extendable from the vertices of a hypercube and the second technique characterizes the convex envelopes of functions of the form $xg(y)$ over $[0, 1]^{n+1}$ where $x \in [0, 1]$, $y \in [0, 1]^n$, and $g(\cdot)$ is a non-increasing convex function. It was shown in [37], that the above results generalize existing results concerning convex envelopes in [25, 30, 10, 6, 22, 28]. In Section 3.1, we provide a construction that unifies these seemingly disparate techniques thereby developing a common framework to understand many individual convexification schemes in the literature. In Section 3.2, we compare $\operatorname{conv}\left(\bigcap_{i=1}^{m} E_i^P\right)$ with $\bigcap_{i=1}^{m} \operatorname{conv}(E_i^P)$. In particular, we show that if h_i are submodular functions and $P = \{0, 1\}^n$, then $\operatorname{conv}\left(\bigcap_{i=1}^{m} E_i^P\right) = \bigcap_{i=1}^{m} \operatorname{conv}(E_i^P)$. For general h_i, this is, however, not true. Nevertheless, we show that if similar convex extensions based arguments are used in the convexification of each h_i, the simultaneous convex hull can often be obtained using disjunctive programming techniques and results in a relaxation that is typically tighter than the one obtained from individual convexification.

3.1 Convex Envelopes of Submodular and Disjunctive Functions

Consider a function $g(x, y)$ where $x \in \{0, 1\}$ and $y \in [0, 1]^n$ and assume that the convex envelope of $g(x, y)$ over $[0, 1]^{n+1}$ is lower-semicontinuous. We will assume that the convex envelopes of $g(0, y)$ and $g(1, y)$ are known. The convex envelope of $g(x, y)$ can therefore be described using disjunctive programming techniques in a higher-dimensional space; see [33]. However, our current interest is in functions whose convex envelope can be described easily in the space of the original variables.

Let $\bar{x} \in [0,1]$ and $\bar{y} \in [0,1]^n$. Let $V \in \mathbb{R}^{n+1} \times \mathbb{R}^k$ be a matrix whose columns will be interpreted as points in $\mathbb{R}^{n+1}$. Assume V is such that $\mathrm{conv}_{[0,1]^{n+1}}\, g(\bar{x}, \bar{y}) = g(V)^T\lambda$, $V\lambda = (\bar{x}, \bar{y})$, $e^T\lambda = 1$, and $\lambda > 0$, *i.e.*, the convex envelope of g at $(\bar{x}, \bar{y})$ is determined by the points in V. Let $V = (V', V'')$ and $\lambda = (\lambda', \lambda'')$ where $V' \in \mathbb{R}^{n+1} \times \mathbb{R}^m$ and $\lambda' \in \mathbb{R}^m$. Let $e_{V'}$ (respectively, $e_{V''}$) denote a vector of all ones with a dimension equal to the number of columns of V' (respectively, columns of V''). Then,

$$\begin{aligned}
&\mathrm{conv}_{[0,1]^{n+1}}(g)(\bar{x}, \bar{y}) = \mathrm{conv}_{[0,1]^{n+1}}(g)(V\lambda) \\
&= \mathrm{conv}_{[0,1]^{n+1}}(g)\left(e_{V'}^T\lambda' \frac{V'\lambda'}{e_{V'}^T\lambda'} + e_{V''}^T\lambda'' \frac{V''\lambda''}{e_{V''}^T\lambda''}\right) \\
&\le e_{V'}^T\lambda' \,\mathrm{conv}_{[0,1]^{n+1}}(g)\left(\frac{V'\lambda'}{e_{V'}^T\lambda'}\right) + e_{V''}^T\lambda'' \,\mathrm{conv}_{[0,1]^{n+1}}\left(\frac{V''\lambda''}{e_{V''}^T\lambda''}\right) \\
&\le g(V')\lambda' + g(V'')\lambda'' = \mathrm{conv}_{[0,1]^{n+1}}(g)(\bar{x}, \bar{y}),
\end{aligned} \tag{1}$$

where the first equality follows since $(\bar{x}, \bar{y}) = V\lambda$, the first inequality from convexity of $\mathrm{conv}_{[0,1]^{n+1}}(g)$, the second inequality since $\mathrm{conv}_{[0,1]^{n+1}}(g)$ is convex and $\mathrm{conv}_{[0,1]^{n+1}}(g)$ underestimates g. Therefore, equality holds throughout and in particular,

$$\mathrm{conv}_{[0,1]^{n+1}}(g)\left(\frac{V'\lambda'}{e^T\lambda'}\right) = g(V')^T\frac{\lambda'}{e^T\lambda'}, \tag{2}$$

i.e., the convex envelope of $\frac{V'\lambda'}{e^T\lambda'}$ is determined by the function values at V'. Assume that, for every $(\bar{x}, \bar{y}) \in [0,1]^{n+1}$, it is possible to select a V such that for every pair of points $v', v'' \in V$, where $v' = (0, y^0)$ and $v'' = (1, y^1)$, (y^0, y^1) satisfies a property $\mathcal{P}$. Then, we say that the convex combinations of g satisfy $\mathcal{P}$ pairwise. Let y_I denote the vector $(y_i)_{i\in I}$. In particular, if there exists an $I \subseteq N$ such that $y^0_I = 0$ and $y^1_{I^c} = e_{I^c}$, where $I^c = N\backslash I$ and e_{I^c} is a vector of ones in $\mathbb{R}^{|I_c|}$, then we say that the convex combinations of g are *pairwise complementary*. Now, if $V' \subseteq V$ such that each column of V' is of the form $(0, y)$ and V' contains all such columns, then it follows by choosing $(0, y') = \frac{V'\lambda'}{e_{V'}^T\lambda'}$ that $y'_I = 0$. Similarly, by defining $(0, y'') = \frac{V''\lambda''}{e_{V''}^T\lambda''}$, it follows that $y''_{I^c} = e_{I^c}$. Assume $0 < \bar{x} < 1$. Since $\bar{y} = (1-\bar{x})y' + \bar{x}y''$, it follows that $y'_{I^c} = \frac{\bar{y}_{I^c} - \bar{x}e_{I^c}}{1-\bar{x}}$ and $y''_I = \frac{\bar{y}_I}{\bar{x}}$. We note that $[0, y'] \in [0,1]^{n+1}$ only if $0 \le \bar{y}_{I^C} - \bar{x} \le 1 - \bar{x}$ and $[1, y''] \in [0,1]^{n+1}$ only if $0 \le \bar{y}_I \le \bar{x}$. In fact, given $(\bar{x}, \bar{y})$, y' and y'' are uniquely defined and can be identified by choosing $I = \{i \mid \bar{y}_i \le \bar{x}\}$. Underlying this construction is the fact that $\bigcup_{I\subseteq N} S_I$ where $S_I = \{(x, y) \mid 0 \le y_I \le x \le y_{I^c} \le 1\}$ forms a polyhedral subdivision of $[0,1]^{n+1}$; see [37] for further details. Therefore, we have proved the following result.

Proposition 3.1. *Consider a function $g : [0,1]^{n+1} \mapsto \bar{\mathbb{R}}$ whose convex combinations are pairwise complementary. Let $(x, y) \in [0,1]^{n+1}$. Define $I = \{i \mid y_i \le x\}$. If $x > 0$, let*

$$y' = \begin{cases} 0 & \text{if } i \in I \\ \dfrac{y_i - x}{1 - x} & \text{otherwise,} \end{cases} \qquad \text{and} \qquad y'' = \begin{cases} \dfrac{y_i}{x} & \text{if } i \in I \\ 1 & \text{otherwise.} \end{cases}$$

Otherwise, for all i, let $y'_i = y_i$ and $y''_i = 1$. Then, $\mathrm{conv}_{[0,1]^{n+1}}(g)(\bar{x}, \bar{y}) = (1-\bar{x})\,\mathrm{conv}_{[0,1]^{n+1}}\, g(0, y') + \bar{x}\,\mathrm{conv}_{[0,1]^{n+1}}\, g(1, y'')$. □

Now, we discuss practical applications of Proposition 3.1. First, consider a function $g(x, y) : \{0,1\}^{n+1} \mapsto \mathbb{R}$ that is submodular, *i.e.*,

$$g(x', y') + g(x'', y'') \ge g(x' \vee x'', y \vee y'') + g(x' \wedge x'', y' \wedge y''),$$

where $\vee$ and $\wedge$ denote component-wise maximum/minimum; see [40]. We will show that the convex combinations of submodular functions are pairwise complementary. It is well-known that $g(x, y)$ is submodular if and only if $g(0, y)$ and $g(1, y)$ are submodular and $g(1, y) - g(0, y)$ is non-increasing. Further, $g'(x, y) = g(0, y) + x\big(g(1, y) - g(0, y)\big)$ equals $g(x, y)$ when $x \in \{0,1\}$ and is submodular over $[0,1] \times \{0,1\}^n$; see [37].

The advantage of the transformation from g to g' is that g' is submodular also when $x \notin \{0,1\}$. Further, g' is linear in x and therefore convex extendable from $x \in \{0,1\}$. It follows then that $\mathrm{conv}_{[0,1]^{n+1}}(g) = \mathrm{conv}_{[0,1]^{n+1}}(g')$. By induction, for a function $g(x, y) : \{0,1\}^{n+1} \mapsto \mathbb{R}$ that is submodular, we obtain a multilinear function $g'(x,y)$ that is submodular over $[0,1]^{n+1}$ and its convex envelope matches that of $g(x,y)$. The following result also follows from the characterization of the convex envelope of a submodular function in terms of the Kuhn's triangulation of $[0,1]^{n+1}$; see [21, 37]. The typical proof of the latter result relies on Edmond's characterization of greedy solutions for polymatroids. Here, we provide a direct primal argument that gives an alternate proof of the characterization of the envelope in terms of Kuhn's triangulation.

Proposition 3.2. *Let $x \in [0,1]$ and $y \in [0,1]^n$. Consider $g(x,y) : [0,1]^{n+1} \mapsto \mathbb{R}$ and assume that $g(x,y)$ is submodular when restricted to $\{0,1\}^{n+1}$ and convex-extendable from $\{0,1\}^{n+1}$. Then, two-point convex combinations of g are pairwise complementary.*

Proof. By replacing $g(x,y)$ with the multilinear function derived above, we may assume that $g(x,y)$ is submodular over $[0,1]^{n+1}$ and is linear when all but one variable is fixed. Consider a point $(\bar{x},\bar{y})$ such that $\mathrm{conv}_{[0,1]^{n+1}}\, g(\bar{x},\bar{y})$ is determined by V and the multipliers for each point in V are non-zero. We first show that V can be chosen so that whenever $v' = (0,y^0) \in V$ and $v'' = (1,y^1) \in V$, it holds that $y^0 \le y^1$. In this case, we say that the pairwise combinations are non-decreasing. Assume for deriving a contradiction that this is not true. First, let $\lambda' \le \lambda''$, where λ' (resp. λ'') is the convex multiplier associated with v' (resp. v''). We show that v' and v'' can be replaced in V with $v^1 = (0, y^0 \wedge y^1)$, v'', and $v^2 = (1, y^0 \vee y^1)$. Note that $\lambda' v' + \lambda'' v'' = \lambda'(v^1 + v^2) + (\lambda'' - \lambda')v''$. However,

$$\begin{aligned}&\lambda' g(v') + \lambda'' g(v'') \ge \lambda'(g(v^1) + g(v^2)) + (\lambda'' - \lambda')g(v'')\\ &\ge \mathrm{conv}_{[0,1]^{n+1}}(g)\left(\frac{v'\lambda' + v''\lambda''}{\lambda' + \lambda''}\right)(\lambda' + \lambda''),\end{aligned}$$

where the first inequality follows by submodularity. Therefore, by (2), equality holds throughout. Similarly, it can be shown that if $\lambda' > \lambda''$, then v' and v'' can be replaced with v^1, v', and v^2. We carry out the above exchange iteratively. Note that all the vectors generated belong to a finite lattice spanned by vectors in V. Further, at each step of the process if $\lambda' \le \lambda''$ a vector of the form $(0,y)$ decreases strictly or if $\lambda' > \lambda''$ then a vector of the form $(1,y)$ increases strictly. In addition, any introduced point of the form $(1,y)$ is at least as large as an existing point of the same form and any introduced point of the form $(0,y)$ is no larger than an existing point. Therefore, if we choose a maximal y^0 and a minimal y^1 at each step, the process converges finitely with points that are such that pairwise-combinations are non-decreasing.

Now, consider a convex combination $v' = (0,y^0)$ and $v'' = (0,y^1)$ where $y^0 \le y^1$. Consider an $i \in \{1,\dots,n\}$ such that $y^0_i > 0$ and $y^1_i < 1$. Then, $\lambda' v' + \lambda'' v'' = \lambda'(v' - \lambda''\epsilon e_i) + \lambda''(v'' + \lambda'\epsilon e_i)$. Let $w^0 = v' + (1-y^0_i)e_i$, $w^1 = v'' + (1-y^1_i)e_i$, $z^0 = v' - y^0_i e_i$, and $z^1 = v'' - y^1_i e_i$. Then, choose $\epsilon = \min\left\{\frac{y^0_i}{\lambda''}, \frac{1-y^1_i}{\lambda'}\right\} > 0$ such that

$$\begin{aligned}&0 \le \lambda'(g(v' - \lambda''\epsilon e_i) - g(v')) + \lambda''(g(v'' + \lambda'\epsilon e_i) - g(v''))\\ &= \lambda'\lambda''\epsilon(g(z^0) - g(w^0) + g(w^1) - g(z^1)) \le 0.\end{aligned}$$

Here, the first inequality follows from the choice of ϵ, (2), and $\lambda'(v' - \lambda''\epsilon e_i) + \lambda''(v'' - \lambda'\epsilon e_i) = \lambda' v' + \lambda'' v''$, the first equality by linearity of g when all variables except y_i are fixed and the second inequality by submodularity of g. Therefore, equality holds throughout. In other words. the i^{th} coordinate of y^1 can be increased and that of y^0 decreased until the first reaches one or the second reaches zero. ☐

We can apply Proposition 3.2 inductively along with Proposition 3.1 to develop the convex envelope of a function that is submodular and convex-extendable from the vertices. As discussed before Proposition 3.2, the polyhedral subdivision associated with pairwise complementary combinations is of the form $\{0 \le y_i \le x \le y_j \le 1, i \in I, j \in I^c\}$. On the other hand, Kuhn's triangulation, $\mathcal{K} = \{\Delta_1,\dots,\Delta_{n!}\}$, expresses $[0,1]^{n+1}$ as a union of simplices, where for each permutation $\pi = \{\pi(1),\dots,\pi(n+1)\}$ of $\{1,\dots,n+1\}$, Δ_π is the simplex $\{x \mid 0 \le x_{\pi(n+1)} \le \dots \le x_{\pi(1)} \le 1\}$; see [18] for details. The convex envelope of a submodular function is linear over each simplex of the Kuhn's triangulation; see [37] for details. Observe that in Proposition 3.1, $y'_i \le y'_j$ and $y''_i \le y''_j$ if $y_i < y_j$. It can then be easily verified that a recursive application of Proposition 3.2 yields the Kuhn's triangulation.

In [37], the authors also considered a class of disjunctive functions of the form $xf(y)$, where $f(y)$ is non-increasing and convex, and developed their convex envelope over $[0,1]^{n+1}$. We now consider a slight generalization of this class of functions. In particular, the class of functions we consider includes $g(x,y)$, where $g(0,y)$ is non-decreasing convex function and $g(1,y)$ is a non-increasing convex function and $x \in \{0,1\}$. We show that combinations of g are pairwise-complementary. Then, the convex envelope of $g(x,y)$ also follows from Proposition 3.1.

Proposition 3.3. *Consider $g(x,y)$ where $x \in \{0,1\}$ and $y \in [0,1]^n$. Assume $g(0,y)$ is non-decreasing in y_i and $g(1,y)$ is non-increasing function in y_i when all $y_{i'}$ with $i' \neq i$ are fixed. Then, the pairwise combinations of g are complementary.*

Proof. Consider $(\bar{x},\bar{y})$ such that the convex envelope of g at $(\bar{x},\bar{y})$ is determined by V. Let $v' \in V$ such that $v' = (0,y^0)$ and $v'' = (1,y^1) \in V$. Let $i \in \{1,\dots,n\}$ such that $y_i^0 > 0$ and $y_i^1 < 1$. For a sufficiently small t, $v^0(t) = v' - \lambda'' t e_i$ and $v^1(t) = v'' + \lambda' t e_i$ are both feasible. Then, $h(t) = \lambda' g(v^0(t)) + \lambda'' g(v^1(t))$ is non-increasing in t. Further, $\lambda' v^0(t) + \lambda'' v^1(t) = \lambda' v' + \lambda'' v''$. Choose $t = t'$ where $t' = \min\left\{\frac{y_i^0}{\lambda'}, \frac{1-y_i^1}{\lambda''}\right\}$. Since $\lambda' g(v^0(t')) + \lambda'' g(v^1(t')) \leq \lambda' g(v') + \lambda'' g(v'')$, v' and v'' can be replaced with $v^0(t')$ and $v^1(t')$. Let $V' = \{v \in V \mid v = (0,y), y_i > 0\}$ and $V'' = \{v \in V \mid v = (1,y), y_i < 1\}$. If $V' \neq \emptyset$ and $V'' \neq \emptyset$, one can use the argument above to reduce $|V'| + |V''|$. Therefore, after a finite number of steps either $|V'|$ or $|V''|$ equals zero. □

3.2 Inclusion Certificates and Simultaneous Convexification

The primary idea in Theorem 2.1 can be expressed succinctly using probability measures. The set $V(x^0)$ and the multipliers $\lambda(x^0)$ can be visualized as a measure defined on P, say $\mu(x^0)$ such that $E_{\mu(x^0)}[x] = x^0$, $E_{\mu(x^0)}[F(x)] \geq F(x^0)$, and $E_{\mu(x^0)}[H(x)] \leq H(x^0)$, where the subscript denotes the measure used in the computation of the expectation. We assume here, for ease of presentation, that each measure is discrete and has a finite support. However, most of our results go through for more general measures. Motivated by this discussion, we define an inclusion certificate as follows.

Definition 3.4. *Let P be a compact set. For each point $x^0 \in P'$, where $P' \subseteq \operatorname{conv}(P)$, an inclusion certificate is a measure $\mu(x^0)$ with its support in P such that $E_{\mu(x^0)}[x] = x^0$.*

To capture the essence of Theorem 2.1, we say an inclusion certificate proves exclusion of x^0 from S^P if $E_{\mu(x^0)}[F(x)] \geq F(x^0)$, and $E_{\mu(x^0)}[H(x)] \leq H(x^0)$ and $\mu(x^0)$ is not a Dirac measure. For a polytope P, any point can be expressed as a convex combination of $\operatorname{extr}(P)$. In this case, the inclusion certificate is equivalent to barycentric coordinates.

Example 3.5. *Let P_i be a $n_i - 1$-dimensional simplex in $\mathbb{R}^{n_i}$ for $i = 1,\dots,k$. Since the barycentric coordinates of a point are invariant under an invertible linear transformation of a polytope, we assume without loss of generality that $P_i = \{x_i \mid x_i \geq 0, e^T x_i = 1\}$. Let $x_i^0 \in P_i$. We identify $\prod_{i=1}^k e_{ij_i}$, where $j_i \in \{1,\dots,n_i\}$ and e_{ij_i} is the j_i^{th} principal vector in $\mathbb{R}^{n_i}$ with χ^J, where $J = \{j_1,\dots,j_k\}$. Observe that $x_i^0 = \sum_{j=1}^{n_i} x_{ij}^0 e_{ij}$. Therefore, for $x^0 = (x_1^0,\dots,x_k^0) \in \prod_{i=1}^k P_i$, the barycentric coordinate for x^0 associated with the extreme point χ^J may be chosen as $\prod_{i=1}^k x_{ij_i}^0$.*

The special case dealing with $[0,1]^n$ is particularly interesting. We establish a one-to-one correspondence between a subset of $N = \{1,\dots,n\}$, say I, and an extreme point of $[0,1]^n$ by using the indicator vector of I, denoted as χ^I. Then, the barycentric coordinates for an $x \in [0,1]^n$ can be gleaned from $x = \sum_{I \subseteq N} \chi^I \left(\prod_{i\in I} x_i \prod_{i \in I^c}(1-x_i)\right)$ as the product-factors used in the reformulation-linearization technique. □

We now discuss the connection of inclusion certificates, disjunctive programming, and Theorem 2.1. Let P be a polytope with extreme points $(v^1,\dots,v^r)$. Let $\mu_i(x^0)$ be the barycentric coordinate of $x^0 \in P$ associated with v^i. Then, it follows that $\mu_i(x^0) = \sum_{i'=1}^r \mu_i(v^{i'})\mu_i(x^0)$ since $\mu_i(v^{i'})$ is 1 if $i = i'$ and 0 otherwise. For any set $C \subseteq P$, let $Y^C = \{(x,z) \mid z_j = \sum_{i=1}^r \alpha_{ij}\mu_i(x), j = 1,\dots,m,\ x \in C\})$, where $\alpha_{ij} \in \mathbb{R}$ for $i = 1,\dots,r$ and $j = 1,\dots,m$. Since $L\operatorname{conv}(C) = \operatorname{conv}(LC)$ for a linear transformation L and a set C, it follows from Theorem 2.1 that

$$\operatorname{conv}(Y^P) = \operatorname{conv}(Y^{\operatorname{extr}(P)}). \tag{3}$$

Assume without loss of generality that $P_i' \in \mathbb{R}^{k_i}$ is a full-dimensional polytope for each $i = 1,\dots,k$. Let $V_i = \operatorname{extr}(P_i')$. As before, we also denote by V_i the matrix obtained by arranging the elements of V_i as columns. Then, $V_i P_i = P_i'$, where P_i is as defined in Example 3.5. Similarly, any linear function $\langle a, x_i\rangle$ where $x_i \in P_i'$

can be extended to a linear function $\langle aV_i, y_i\rangle$ where $y_i \in P_i$. In this way, a multilinear function in $\prod_{i=1}^k \mathbb{R}^{k_i}$ extends to a multilinear function in $\prod_{i=1}^k \mathbb{R}^{n_i}$. Note that by utilizing $\sum_{j=1}^{n_i} x_{ij} = 1$, any multilinear function in $\prod_{i=1}^k \mathbb{R}^{n_i}$ can be expressed as a linear function of the barycentric coordinates presented in Example 3.5. Then, (3) provides a different proof of the fact that one can restrict x to $\prod_{i=1}^k \operatorname{extr}(P'_i)$ while convexifying multilinear functions over $\prod_{i=1}^k P'_i$.

The above argument can be easily extended to disjunctive programming. Let $K = \{1,\dots,k\}$. For each $i \in K$, let C_i be a convex set and let W be such that $\bigcup_{i=1}^k C_i \subseteq W \subseteq \operatorname{conv}\left(\bigcup_{i=1}^k C_i\right)$. Consider any $x \in W$ and for each $i \in \{1,\dots,k\}$ define $\gamma_i(x)$ and $y_i(x)$ such that $\gamma_i(x) \ge 0$, $\sum_{i=1}^k \gamma_i(x) = 1$, $y_i(x) \in C_i$, and $x = \sum_{i=1}^k \gamma_i(x) y_i(x)$. In other words, $\gamma_i(x)$ and $y_i(x)$ define an inclusion certificate for x in W with support in $\bigcup_{i=1}^k C_i$. Assume further that for $x \in C_i$, $\gamma_i(x) = 1$ and, therefore, $y_i(x) = x$. For any set C, let $Q^C = \{(x, z_1, \dots, z_k, \lambda) \mid z_i = \gamma_i(x) y_i(x), i \in K,\ \lambda_i = \gamma_i(x), i \in K,\ x \in C\}$. Then, it follows from Theorem 2.1 that $Q^W = \operatorname{conv}\left(Q^{\bigcup_{i=1}^k C_i}\right)$ because $z_i = \gamma_i(x) y_i(x) = \sum_{i'=1}^k \gamma_i(x)\gamma_i(y_{i'}(x)) y_i(y_{i'}(x))$ and $\lambda_i = \gamma_i(x) = \sum_{i'=1}^k \gamma_i(x)\gamma_i(y_{i'}(x))$.

Now, we consider the situation where the domain is not necessarily a disjunctive union of convex sets. Also, we allow more general compositions of functions. Let $X_i \subseteq P_i$, $i = 1,\dots,n$. For each $x_i^0 \in P_i$, assume that we have inclusion certificate that has a support on $P_i \backslash X_i$ and is a Dirac measure only if $x_i^0 \in P_i\backslash X_i$. For $x^0 \in \prod_{i=1}^n P_i$, we induce a measure $\mu(x^0)$ by assuming independence of the above measures. Then, it can be easily verified that $\mu(x^0)$ is an inclusion certificate and is the Dirac measure only if $x^0 \in \prod_{i=1}^n P_i\backslash X_i$. We define $P = \prod_{i=1}^n P_i$, and $X = P\backslash \left(\prod_{i=1}^n P_i\backslash X_i\right)$.

Theorem 3.6. *Let $H_i = (h_{i1}, \dots, h_{ij_i})$ and $F_i = (f_{i1}, \dots, f_{ij_i})$ be vectors of functions for $i = 1,\dots,n$. Further, for $z_i \in \mathbb{R}^{j_i}$, let $u(z_1,\dots,z_n,x)$ (resp. $v(z_1,\dots,z_n)$) be a vector of functions each of which is non-decreasing in z (i.e., $u(z', x) \ge u(z'', x)$ and $v(z', x) \ge v(z'', x)$ if $z' \ge z''$) and concave (resp. convex) for each (x_i, z_i) when $(x_{i'}, z_{i'})$ are fixed for all $i' \neq i$ at some values in $\operatorname{dom}(F_{i'}) \times P_i$. Let $S^P = \{(x, w) \mid u(H_1(x_1),\dots,H_n(x_n),x) \le w \le v(F_1(x_1),\dots,F_n(x_n),x), x \in P\}$, where $P = \prod_{i=1}^n P_i$. Let $\mu_i(x_i)$ be an inclusion certificates that proves exclusion of X_i from $\{(x_i, z_i) \mid H_i(x_i) \le z_i \le F_i(x_i),\ x \in P_i\}$. Then, the inclusion certificate for P, $\mu(x)$, obtained using independence of $\mu_i(x_i)$ proves exclusion of X from S^P. If for some i and all $x_i^0 \in P_i$, $E_{\mu_i(x_i^0)}[H_i(x_i)] = H_i(x_i^0)$ then u need not be non-decreasing in z_i. Similarly, if $E_{\mu_i(x_i^0)}[F_i(x_i)] = F_i(x_i^0)$ then v need not be non-decreasing in z_i. In particular, if $H_i(x_i) = F_i(x_i)$ then $E_{\mu_i(x_i^0)}[H_i(x_i)] = H_i(x_i^0)$ and $E_{\mu_i(x_i^0)}[F_i(x_i)] = F_i(x_i^0)$.*

Proof. Choose an arbitrary $x^0 \in P$. We prove, that

$$\begin{aligned}&E_{\mu(x^0)}[u(H_1(x_1),\dots,H_n(x_n),x)] \le u\big(H_1(x_1^0),\dots,H_n(x_n^0),x^0\big) \text{ and}\\ &E_{\mu(x^0)}[v(F_1(x_1),\dots,F_n(x_n),x)] \ge v\big(F_1(x_1^0),\dots,F_n(x_n^0),x^0\big).\end{aligned} \tag{4}$$

Then, the result follows from Theorem 2.1. We show the first relation by showing the following sequence of inequalities:

$$\begin{aligned}&E_{\mu(x^0)}[u(H_1(x_1),\dots,H_n(x_n),x)]\\ &\le u\big(E_{\mu(x^0)}[H_1(x_1)],\dots,E_{\mu(x^0)}[H_n(x_n)],x^0\big)\\ &\le u(H_1(x_1^0),\dots,H_n(x_n^0),x^0).\end{aligned} \tag{5}$$

We prove the first inequality above by induction. Let $H'_i(x_i)$ denote $E_{\mu(x^0)}[H_i(x_i)]$, $I = \{i,\dots,n\}$, and $I' = \{i+1,\dots,n\}$. We show that:

$$\begin{aligned}&E_{\mu(x^0)}[u(H_1(x_1),\dots,H_n(x_n),x) \mid x_j = x'_j, j \in I]\\ &\le u(H'_1(x_1),\dots,H'_{i-1}(x_{i-1}),H_i(x'_i)\dots,H_n(x'_n),x_1^0,\dots,x_{i-1}^0,x'_i,\dots,x'_n)\end{aligned}$$

and observe that the first inequality in (5) follows when $i = n+1$. The base case with $i = 1$ follows trivially since all the variables are fixed. For the inductive step:

$$\begin{aligned}&E_{\mu(x^0)}[u(H_1(x_1),\dots,H_n(x_n),x) \mid x_j = x'_j, j \in I']\\ &= E_{\mu(x^0)}[E_{\mu(x^0)}[u(H_1(x_1),\dots,H_n(x_n),x) \mid x_j = x''_j, j \in I] \mid x''_j = x'_j, j \in I']\\ &\le E_{\mu(x^0)}[u(H'_1(x_1),\dots,H'_{i-1}(x_{i-1}),H_i(x''_i),\dots,H_n(x''_n),x_1^0,\dots,x_{i-1}^0,x''_i,\dots,x''_n)\\ &\qquad\qquad \mid x''_j = x'_j, j \in I']\\ &\le u(H'_1(x_1),\dots,H'_i(x_i),H_{i+1}(x'_{i+1})\dots,H_n(x'_n),x_1^0,\dots,x_i^0,x'_{i+1}\dots,x'_n),\end{aligned}$$

where the first equality holds by the law of iterated expectations, the first inequality by the induction hypothesis, and the second inequality by the assumed independence in the definition of $\mu(x^0)$ and Jensen's inequality since u is concave in (z_i, x_i).

Now, we show the second inequality in (5). Note that u is non-decreasing in each z_i. Further, for all i, it follows from the definition of $\mu(x^0)$ and $\mu_i(x_i^0)$ that $E_{\mu(x^0)}[H_i(x_i)] = E_{\mu_i(x_i^0)}[H_i(x_i)] \le H_i(x_i^0)$. Observe that if $E_{\mu_i(x_i^0)}[H_i(x_i)] = H_i(x_i^0)$, u does not need to be non-decreasing for the second inequality to hold. Similarly, $E_{\mu(x^0)}[v(F_1(x_1), \ldots, F_n(x_n), x)] \ge v\big(F_1(x_1^0), \ldots, F_n(x_n^0), x^0\big)$. Therefore, $\mu(x^0)$ is the desired inclusion certificate.

The last statement in the theorem follows by observing that if $E_{\mu_i(x_i^0)}(H_i(x_i)) \le H_i(x_i^0) = F_i(x_i^0) \le E_{\mu_i(x_i^0)}(F_i(x_i)) = E_{\mu_i(x_i^0)}(H_i(x_i))$, where the last equality follows since $F_i(x_i) = H_i(x_i)$. Then, equality holds throughout. □

We now consider applications of Theorem 3.6. Let $H_i(x_i)$ be a vector of functions $(h_{i1}(x_i), \ldots, h_{iK}(x_i))$ and let $F_i(x_i) = H_i(x_i)$. Consider $g(x) = \sum_{k=1}^K \prod_{i=1}^m h_{ik}(x_i)$, where, for each i and $x_i^0 \in P_i$, there exists an inclusion certificate $\mu_i(x_i^0)$ such that $E_{\mu_i(x_i^0)}[h_{ik}(x_i)] = h_{ik}(x_i^0)$. For each i, let $X_i = \{x \mid \mu_i(x_i) \text{ is not a Dirac measure}\}$. Now, consider the case when x_i are fixed for all $i \ne i'$. Then, $h(x) = \sum_{k=1}^K c_k h_{i'k}(x_{i'})$ where, for each k, c_k is some constant. Since summation is a linear function, it follows from Theorem 3.6 that for convexifying $g(x)$ it suffices to limit attention to the points in $\prod_{i=1}^n P_i \backslash X_i$. Now, let $F_i(x_i) = (f_{i1})(x_i), \ldots, f_{iK}(x_i))$ where $f_{ik}(x_i) = \infty$ and assume that h_{ik} is non-negative over P_i. Then, c_k is also non-negative. In this case, it suffices to find an inclusion certificate $\mu_i(x_i^0)$ such that $E_{\mu(x_i^0)}[h_{ik}(x_i)] \le h_{ik}(x_i^0)$.

If each $h_{ik}(x)$ is concave then, by Jensen's inequality, any inclusion certificate satisfies $E_{\mu(x_i^0)}[h_{ik}(x_i)] \le h_{ik}(x_i^0)$. Therefore, if, for all i and k, $h_{ik}(x)$ is non-negative whenever $x \in P_i$, then for convexifying $g(x)$ it suffices to restrict attention to $x \in \prod_{i=1}^K \operatorname{extr}(P_i)$. It is instructive to view Corollary 2.7 as a special case of Theorem 3.6. Note that each multilinear function can be written in the form of $g(x)$ with all $h_{ik}(x_i)$ being linear functions. Then, by linearity of h_{ik}, it follows that every inclusion certificate yields $E_{\mu_i(x_i^0)}[h_{ik}(x_i)] = h_{ik}(x_i)$. Since every point in a compact set can be expressed as a convex combination of the extreme points, there exists an inclusion certificate with support on $\operatorname{extr}(P_i)$. Then, Corollary 2.7 follows from Theorem 3.6.

Theorem 3.6 can sometimes be used even if H_i and H_j for $i \ne j$ depend on some common variables. For example, if $S = \{(x, y, w) \mid u(H_1(x_1, y), \ldots, H_n(x_n, y), x, y) \le w \le v(F_1(x_1, y), \ldots, F_n(x_n, y), x, y), x \in P, y \in C\}$, where C is an arbitrary set, then one may fix y before applying Theorem 3.6. In fact, Theorem 3.6 itself can be viewed as an inductive application of the above principle. In particular, for i' increasing from 1 to n, we fix all x_i, $i \ne i'$ and realize that $x_{i'}$ can be restricted to $P_{i'} \backslash X_{i'}$. We illustrate an application of these ideas in the next example.

Example 3.7. *Let $x \in \mathbb{R}^3$, $y \in \mathbb{R}^2$, $H(x) = (h_i(x))_{i=1}^4 = \left(\frac{x_1 x_2}{y_1}, \frac{x_1 - x_1 x_2}{y_1 y_2}, x_1 x_2, x_1 x_2 x_3\right)$. Let $\mathcal{H} = [0,1]^3 \times [y^L, y^U]$, where $y^L \in \mathbb{R}^2_{++}$ and $y^U \ge y^L$. Let $S^{\mathcal{H}} = \{(x, y, z) \mid (z_1, z_2) \ge (h_1(x), h_2(x)),\ (z_3, z_4) = (h_3(x), h_4(x)),\ (x, y) \in \mathcal{H}\}$. Now, let us fix $y = \bar{y}$. The resulting set satisfies the conditions of Theorem 3.6 (in fact, it reduces to the special case discussed in Corollary 2.7). Fixing x_2 and x_3, all the associated functions become linear. Therefore, it suffices to restrict $x_1 \in \{0, 1\}$. Note that $S^{\mathcal{H}} \cap \{(x, y, z) \mid x_1 = 0\}$ is convex. On the other hand, consider $x_1 = 1$. In this case, applying the above argument inductively, we fix $x_2 \in \{0, 1\}$. The resulting set is convex for both fixings. Therefore, $\operatorname{conv}(S^{\mathcal{H}}) = \operatorname{conv}\left(\bigcup_{i=1}^3 S^i\right)$, where $S^1 = \{(x, y, z) \mid x_1 = 0,\ z \ge 0,\ (z_3, z_4) = (0, 0),\ (x, y) \in \mathcal{H}\}$, $S^2 = \big\{(x, y, z) \bigm| (x_1, x_2) = (1, 0),\ z_1 \ge 0, (z_3, z_4) = 0,\ z_2 \ge \frac{1}{y_1 y_2},\ (x, y) \in \mathcal{H}\big\}$, and $S^3 = \big\{(x, y, z) \bigm| (x_1, x_2) = (1, 1),\ z_1 \ge \frac{1}{y_1},\ z_2 = 1,\ z_3 = x_3,\ z_4 \ge 0,\ (x, y) \in \mathcal{H}\big\}$. In fact, since each of these sets is second-order cone representable, and the above sets are closed and bounded, $\operatorname{conv}(S^{\mathcal{H}})$ is second-order cone representable as well; see Proposition 3.3.5 in [5].* □

The key assumption in Theorem 3.6 is that $\mu_i(x_i^0)$ be the same for each h_{ij} and f_{ij}. We illustrate the impact of this assumption using a simple example.

Example 3.8. *Let $\mathcal{H} = [x^L, x^U] \times [y^L, y^U]$, where we assume that $x_L \ge 0$, $y_L > 0$, $x^U > x^L$ and $y^U > y^L$. Let $f_1(x, y) = \frac{x}{y}$ and $f_2(x, y) = xy$. Consider the set $S = \{(x, y, z_1, z_2) \mid z_i = f_i(x, y), i = 1, 2,\ (x, y) \in \mathcal{H}\}$.*

The current relaxation techniques relax each function separately. In particular, S may be relaxed to

$$S' = \{(x, y, z_1, z_2) \mid \mathrm{conv}_{\mathcal{H}}(f_i)(x, y) \le z_i \le \mathrm{conc}_{\mathcal{H}}(f_i)(x, y), i = 1, 2, (x, y) \in \mathcal{H}\};$$

see for example [38]. By Theorem 3.6, it is clear that $\mathrm{conv}(S) = \{(x, y, z_1, z_2) \mid z_i = f_i(x, y), i = 1, 2, (x, y) \in \{x^L, x^U\} \times [y^L, y^U]\}$. *Let S^U (resp. S^L) be the set S where x is restricted to x^U (resp. x^L). Since f_2 is linear when $x = x^U$ or when $x = x^L$, it suffices to convexify f_1 in order to convexify S^U and S^L. Then,* $\mathrm{conv}(S) = \mathrm{conv}(\mathrm{conv}(S^L) \cup \mathrm{conv}(S^U))$, *where the latter set can be described (in a higher-dimensional space) using disjunctive programming. We show that* $\mathrm{conv}(S) \subsetneq S'$. *It can be easily verified that the inequality*

$$z_1 y^L y^U + z_2 - x(y^L + y^U) \le 0 \tag{6}$$

is valid for $\mathrm{conv}(S)$ *since the left-hand-side simplifies to $-\frac{x}{y}(y - y^L)(y^U - y)$ if we expand z_1 and z_2 to $\frac{x}{y}$ and xy respectively. However, we will show that (6) is not valid for S'. In particular, consider* $(x', y') = \left(\frac{x^L+x^U}{2}, \frac{y^L+y^U}{2}\right)$. *Then, by expressing (x', y') as a convex combination of (x^L, y^U) and (x^U, y^L) it follows that* $\mathrm{conc}_{\mathcal{H}}(f_1)(x', y') \ge \frac{1}{2}\left(\frac{x^L}{y^U} + \frac{x^U}{y^L}\right) = z_1' \ge \mathrm{conv}_{\mathcal{H}}(f_1)(x', y')$. *Similarly, by expressing (x', y') as a convex combination of (x^L, y^L) and (x^U, y^U) it follows that* $\mathrm{conc}_{\mathcal{H}}(f_2)(x', y') \ge \frac{x^L y^L + x^U y^U}{2} = z_2' \ge \mathrm{conv}_{\mathcal{H}}(x', y')$. *In fact, since f_1 is convex over the line segment joining (x^L, y^U) and (x^U, y^L) and f_2 is convex over the line segment joining (x^L, y^L) and (x^U, y^U), it follows that $z_1' \ge f_1(x', y')$ and $z_2' \ge f_2(x', y')$. It can be verified that that $z_1' y^L y^U + z_2' - x'(y^L + y^U) = \frac{1}{2}(x^U - x^L)(y^U - y^L) > 0$. However, $(x', y', z_1', z_2') \in S'$. Therefore, (6) is not valid for S'. The key reason why S' is a weaker relaxation compared to* $\mathrm{conv}(S)$ *is that while constructing S', the overestimators of f_1 and f_2 were obtained by expressing (x', y') as a convex combination in two different ways whereas this is not permissible in the construction of* $\mathrm{conv}(S)$. *This fact also explains why the statement of Theorem 3.6 requires that for each i, $\mu_i(x_i^0)$ is independent of j.* □

As in Theorem 2.1, if given a point x^0 and a vector of functions $H(x)$ there exists a non-Dirac measure $\mu(x^0)$ such that $E_{\mu(x^0)}(H(x)) \le H(x^0)$ and $E_{\mu(x^0)}(x) = x^0$, then x^0 can be excluded from the domain while convexifying $H(x)$ simultaneously. Let $H(x)$ be a vector of polyhedral functions, *i.e.*, functions that have a polyhedral convex envelope. Further, assume that the elements of $H(x)$ share the same polyhedral subdivision, *i.e.*, the subdivision of the convex hull of the domain into polyhedra, such that the convex envelope is affine over each polyhedron. Since, the polyhedral subdivision can be refined into a triangulation (see [18]) and the convex multipliers for each point in a simplex are uniquely determined, it follows that for every point that is not the vertex of this triangulation, there exists an inclusion certificate such that all the functions in $H(x)$ are simultaneously underestimated. Therefore, one may restrict attention to the vertices of the triangulation for the purpose of simultaneously convexifying $H(x)$. In particular, the simultaneous convex envelope of $H(x)$ is polyhedral. In fact, a slightly stronger result, namely Corollary 3.9, can be derived in this case.

Corollary 3.9. *Let $H(x) : P \mapsto \bar{\mathbb{R}}^m$ (resp. $F(x) : P \mapsto \bar{\mathbb{R}}^m$) be a vector of m functions $(h_1, \ldots, h_m)$ (resp. $(f_1, \ldots, f_m)$). For each $x^0 \in \mathrm{conv}(P)$, let $\mu(x^0)$ be the inclusion certificate associated with the convex envelopes of h_i and concave envelopes of f_i, i.e., $\mu(x^0)$ has its support in P and for all $i \in \{1, \ldots, m\}$,* $E_{\mu(x^0)}(h_i(x)) = \mathrm{conv}_{\mathrm{conv}(P)} h_i(x)$ *and* $E_{\mu(x^0)}(f_i(x)) = \mathrm{conc}_{\mathrm{conv}(P)} f_i(x)$. *Let*

$$C^P = \left\{(x, z) \;\middle|\; \operatorname*{conc}_{\mathrm{conv}(P)}(f_i)(x) \ge z_i \ge \operatorname*{conv}_{\mathrm{conv}(P)}(h_i)(x), i = 1, \ldots, m,\ x \in P\right\}$$

and $S^P = \{(x, z) \mid F(x) \ge z \ge H(x), x \in P\}$. *Then,* $C^{\mathrm{conv}(P)} = \mathrm{conv}(S^P)$. *In particular, the relation holds if, P is a polytope, f_i and h_i have polyhedral envelopes for all i, and the polyhedral subdivisions of P associated with* $\mathrm{conv}_{\mathrm{conv}(P)}(h_i(x))$ *and* $\mathrm{conc}_{\mathrm{conv}(P)}(f_i(x))$ *are the same.*

Proof. Let $X = \{x \mid \mu(x) \text{ is not a Dirac measure}\}$. The result follows from the following chain of arguments:

$$\mathrm{conv}(S^P) \subseteq C^{\mathrm{conv}(P)} = \mathrm{conv}(C^{P\setminus X}) = \mathrm{conv}(S^{P\setminus X}) \subseteq \mathrm{conv}(S^P).$$

The first inclusion follows since $C^{\mathrm{conv}(P)}$ is convex and $S^P \subseteq C^{\mathrm{conv}(P)}$. The first equality follows from the definition of X and Theorem 2.1 since we have assumed that, for all i, $E_{\mu(x^0)}(h_i(x)) = \mathrm{conv}_{\mathrm{conv}(P)}(h_i)(x^0)$, and $E_{\mu(x^0)}(f_i(x)) = \mathrm{conc}_{\mathrm{conv}(P)}(f_i)(x^0)$. The second equality follows since for each $x^0 \in P\backslash X$, $h_i(x^0) = E_{\mu(x^0)}(h_i(x)) = \mathrm{conv}_{\mathrm{conv}(P)}(h_i)(x^0)$, and $f_i(x^0) = E_{\mu(x^0)}(f_i(x)) = \mathrm{conc}_{\mathrm{conv}(P)}(f_i)(x^0)$. The last inclusion follows since $S^P \supseteq S^{P\backslash X}$.

Now, consider the case where, for all i, f_i and h_i have polyhedral envelopes and the polyhedral subdivisions associated with $\mathrm{conv}_{\mathrm{conv}(P)}(h_i(x))$ and $\mathrm{conc}_{\mathrm{conv}(P)}(f_i(x))$ are the same. A polyhedral subdivision can be refined into a triangulation without adding new vertices; see [18]. Further, the expression for each point in a simplex in terms of the extreme points is unique; see Exercise 2.28(c) in [27]. By definition, the same expression yields the convex envelopes of h_i and concave envelopes of f_i satisfying the conditions on the measure in the earlier part of the result. □

Consider $P = [0,1]^n$. If $H(x)$ (resp. $F(x)$) is a vector of submodular (resp. supermodular) functions when restricted to $\{0,1\}^n$, and each component function of $H(x)$ (resp. $F(x)$) is convex extendable (resp. concave extendable) from the vertices of $[0,1]^n$, then these functions share the same triangulation; see Proposition 3.2 and the following discussion. Therefore, the convex envelope of S^P is easily obtained by forming the convex and concave envelopes of the elements of $H(x)$ and $F(x)$ respectively. As another application of Corollary 3.9, let $x \in [0,1]$ and $y \in [0,1]^n$. Let $H(x,y) : [0,1]^{n+1} \mapsto \mathbb{R}^m$ be a vector of m functions $(h_1(x,y),\ldots,h_m(x,y))$. Assume that each h_i is convex extendable from $x \in \{0,1\}$, convex non-decreasing when $x = 0$, and convex non-increasing when $x = 1$. Then, it follows from Proposition 3.3 and Corollary 3.9 that the epigraphs of functions in $H(x,y)$ are simultaneously convexified by taking convex envelopes of the constituent functions, which in turn can be easily constructed using Proposition 3.1. Functions of these sort appear in commonly in nonlinear programs; see [37] for various examples of the functions discussed above. In Corollary 3.9, we assumed that all the functions depend on the same variables. However, its conclusion extends to situations where the functions do not depend on the same variables. In this context, we derive the inclusion certificate using conditional independence.

Corollary 3.10. *Let $h_1(u,v)$ and $h_2(u,w)$ be two functions defined over $P = U \times V \times W$ where $u \in U$, $v \in V$, and $w \in W$. Given a $(u^0,v^0,w^0) \in P$, let $\mu^{h_1}(u^0,v^0)$ (resp. $\mu^{h_2}(u^0,w^0)$) be an inclusion certificate associated with the convex envelope of h_1 (resp. h_2), i.e., $E_{\mu^{h_1}(u^0,v^0)}(h_1(u,v)) = \operatorname*{conv}_{U\times V}(h_1)(u^0,v^0)$ and $E_{\mu^{h_2}(u^0,w^0)}(h_2(u,w)) = \operatorname*{conv}_{U\times W}(h_2)(u^0,w^0)$. Let $H(u,v,w) = \{h_1(u,v), h_2(u,w)\}$. Then,*

$$\begin{aligned}&\mathrm{conv}\{(z,u,v,w) \mid z_1 \geq h_1(u,v),\ z_2 \geq h_2(u,w),\ (u,v,w) \in P\}\\ &= \{(z,u,v,w) \mid z_1 \geq \operatorname*{conv}_{U\times V}(h_1)(u,v), z_2 \geq \operatorname*{conv}_{U\times W}(h_2)(u,w), (u,v,w) \in \mathrm{conv}(P)\}\end{aligned}$$

if $\Pr_{\mu^{h_1}(u^0,v^0)}(u \in A) = \Pr_{\mu^{h_2}(u^0,w^0)}(u \in A)$ for all $A \subseteq U$.

Proof. Since $P = U \times V \times W$, for any $A \subseteq U$, $B \subseteq V$, and $C \subseteq W$, it follows that $A \times B \times C \subseteq P$. Define $\Pr_{\mu(u^0,v^0,w^0)}(u \in A, v \in B, w \in C) = \Pr_{\mu^{h_1}(u^0,v^0)}(u \in A)\Pr_{\mu^{h_1}(u^0,v^0)}(v \in B \mid u \in A)\Pr_{\mu^{h_2}(u^0,w^0)}(w \in C \mid u \in A)$, *i.e.*, $\mu(u^0,v^0,w^0)$ is defined from $\mu^{h_1}(u^0,v^0)$ and $\mu^{h_2}(u^0,w^0)$ using conditional independence, in particular, assuming independence after conditioning on u. Then, since f_1 does not depend on w and $\Pr_{\mu^{h_2}(u^0,w^0)}(w \in W \mid u \in A) = 1$, it follows that $E_{\mu(u^0,v^0,w^0)}(h_1(u,v)) = E_{\mu^{h_1}(u^0,v^0)}(f_1(u,v)) = \operatorname*{conv}_{U\times V}(h_1)(u^0,v^0)$. Since we assume that $\Pr_{\mu^{h_1}(u^0,v^0)}(u \in A) = \Pr_{\mu^{h_2}(u^0,w^0)}(u \in A)$ for all $A \subseteq U$, it follows similarly that $E_{\mu(u^0,v^0,w^0)}(h_2(u,w)) = E_{\mu^{h_2}(u^0,w^0)}(h_2(u,w)) = \operatorname*{conv}_{U\times W}(h_2)(u^0,w^0)$. Now, μ satisfies the hypothesis of Corollary 3.9 and the result follows. □

Remark 3.11. *Assume $\alpha \in \mathbb{R}^m_+$. Note that for any linear transformation, L, $\mathrm{conv}(LS^P) = L\,\mathrm{conv}(S^P)$. Therefore, it follows from Corollary 3.9 that if for each $x^0 \in P$, there exists an inclusion certificate $\mu(x^0)$ such that $E_{\mu(x^0)}(h_i(x)) = \mathrm{conv}(h_i(x))$ for all i then*

$$\mathrm{conv}_P\left(\sum_{i=1}^m \alpha_i h_i(x)\right) = \sum_{i=1}^m \alpha_i\,\mathrm{conv}_P(h_i)(x),$$

In other words, if the inclusion certificate can be chosen to be common across all the functions h_i, then individual convex envelopes yield the convex envelope of any non-negative linear combination of functions.

Corollary 3.10 extends the application of this result, where each function need not depend on all the variables. In particular, Corollary 3.10 extends Theorem 1.4 in [25] and Proposition 2 in [22]. □

Corollary 3.9 and Remark 3.11 show that if the convex envelope of two functions share the same triangulation, then the convex envelope of a non-negative linear combination of these functions is obtained as the combination of their convex envelopes. Ordinarily, the convex envelopes of the constituent functions do not yield the convex envelope of a non-negative linear combination in this way. To see this, assume that $f_1(x)$ and $f_2(x)$ are two functions whose convex envelope is determined by the extreme points of the domain of these functions. Then, it is not even necessarily true that $f_1(x)$ and $f_2(x)$ are simultaneously convexified by restricting attention to the extreme points of the domain, *i.e.*, the property of convex extendability from vertices does not carry over from individual functions to a collection of functions. We illustrate this behavior using a simple example. Let $x \in [0,1]^2$, $f_1(x) = \max\{x_1 - x_2, x_2 - x_1\}$, and $f_2(x) = \max\{1 - x_1 - x_2, x_1 + x_2 - 1\}$. Define $S^P = \{(x,z) \mid z_1 \geq f_1(x), z_2 \geq f_2(x), x \in [0,1]^2\}$ and note that $S^P = \operatorname{conv}(S^P)$. Further, observe that $f_1(x)$ and $f_2(x)$ are the convex envelopes of their restrictions to $\{0,1\}^2$. Let $X = [0,1]^2 \backslash \{0,1\}^2$. Then, unlike in Corollary 3.9,

$$\operatorname{conv}(S^{P\backslash X}) = \{(x,z) \mid z_1 + z_2 \geq 1, z_1 \geq f_1(x), z_2 \geq f_2(x), x \in [0,1]^2\} \subsetneq S^P,$$

since $\left(\frac{1}{2}, \frac{1}{2}, 0, 0\right) \in S^P \backslash \operatorname{conv}(S^{P\backslash X})$. Hence, x cannot be restricted to $P\backslash X$ in the simultaneous convexification of $f_1(x)$ and $f_2(x)$. Let f'_i denote f_i with its domain restricted to $P\backslash X$. Now, we interpret the example using Remark 3.11. The remark asserts by contra-positivity that if there exists an $x \in P$ such $\operatorname{conv}_P(f'_1 + f'_2)(x) \gneq \operatorname{conv}_P(f'_1)(x) + \operatorname{conv}_P(f'_2)(x)$ then $\operatorname{conv}(S^{P\backslash X}) \subsetneq \{(x,z) \mid z_1 \geq \operatorname{conv}_P(f'_1)(x), z_2 \geq \operatorname{conv}_P(f'_2)(x), x \in [0,1]^2\}$. The above argument shows that $(\frac{1}{2}, \frac{1}{2})$ is such a point. For this example, one can refine the original triangulations to make them identical. Consider restricting attention to $S^{X'}$ where $X' = \{0,1\}^2 \cup \left(\frac{1}{2}, \frac{1}{2}\right)$. Now, the conditions of Corollary 3.9 are satisfied and it can be easily verified that $S^P = \operatorname{conv}(S^{X'})$.

Corollary 3.12. *Let $h_{ij}(x_i) : [0,1]^{n_i} \mapsto \mathbb{R}$, $i \in K = \{1,\ldots,k\}$, $j \in J_i = \{1,\ldots,j_i\}$ be submodular (resp. supermodular) when restricted to $\{0,1\}^{n_i}$ and convex-extendable (concave-extendable) from $\{0,1\}^{n_i}$. Let Z_{ij}, $i \in K$, $j \in J_i$ be convex subsets of $\mathbb{R}$ such that $Z_{ij} \supseteq \{h_{ij}(x_i) \mid x_i \in [0,1]^{n_i}\}$. Further, let $z \in \mathbb{R}^{\sum_{i=1}^k j_i}$ and $g(z,x) : \prod_{i=1}^k \prod_{j=1}^{j_i} Z_{ij} \times \{0,1\}^{\sum_{i=1}^k n_i} \mapsto \mathbb{R}^m$ be a vector of submodular functions in (z,x), each of which is non-decreasing (resp. non-increasing) on z_{ij}, and each function is also concave on $\prod_{j=1}^{j_i} Z_{ij} \times [0,1]^{n_i}$ when for each $i' \neq i$ and $j' \in \{1,\ldots,j_{i'}\}$, $x_{i'}$ and $z_{i'j'}$ are fixed. Let $c(x)$ denote the composite function $g\big(h_{11}(x_1),\ldots,h_{kj_k}(x_k),x\big)$ and for any set P, let $T^P = \{(x,w) \mid w \geq c(x), x \in P\})$. Then,*

$$\operatorname{conv}\left(T^{[0,1]^{\sum_{i=1}^k n_i}}\right) = \operatorname{conv}\left(T^{\{0,1\}^{\sum_{i=1}^k n_i}}\right). \tag{7}$$

If, for some i, all $h_{ij}(x_i) = \operatorname{conv}_{[0,1]^{n_i}}(h_{ij})(x_i)$ then g need not be non-decreasing (resp. non-increasing) in z_i. If each h_{ij} is non-decreasing, then $c(x)$ is a vector of submodular functions when restricted to $\{0,1\}^n$. The convex hull of $T^{[0,1]^{\sum_{i=1}^k n_i}}$ is then described by interpolating $c(x)$ using the Kuhn's triangulation of $[0,1]^{\sum_{i=1}^k n_i}$.

Proof. We first show (7). By Proposition 3.2, the convex envelope of each $h_{ij}(x_i)$ is obtained by Kuhn's triangulation of $[0,1]^{n_i}$. Let $\mu_i(x_i)$ be the measure associated with Kuhn's triangulation of $[0,1]^{n_i}$. Then, $\mu_i(x_i)$ is an inclusion certificate that shows that all points in $[0,1]^{n_i} \backslash \{0,1\}^{n_i}$ can be excluded from consideration while constructing the convex hull of $\{(x_i, z_{ij}) \mid z_{ij} \geq h_{ij}(x_i), j \in J_i, x_i \in [0,1]^{n_i}\}$. Therefore, by Theorem 3.6, there exists an inclusion certificate μ such that $E_{\mu(x^0)}[c(x)] \leq c(x^0)$ and if $x^0 \notin \{0,1\}^{\sum_{i=1}^k n_i}$ then $\mu(x^0)$ is not a Dirac measure. Consequently, (7) follows. If $h_{ij}(x_i^0) = \operatorname{conv}_{[0,1]^{n_i}}(h_{ij})(x_i^0)$ then, by definition, $E_{\mu_i(x_i^0)}[h_{ij}(x_i)] = h_{ij}(x_i^0)$. Therefore, by Theorem 3.6, g need not be non-decreasing in z_i, If each h_{ij} is non-decreasing, it follows from Lemma 2.6.4 in [40] that $c(x)$ is a vector of submodular functions when restricted to $\{0,1\}^{\sum_{i=1}^k n_i}$. The last statement then follows from Proposition 3.2 and Corollary 3.9. □

The conditions in Corollary 3.12 are stricter than the conditions in Lemma 2.6.4 in [40]. We additionally require that g is jointly concave on $\prod_{j=1}^{j} Z_{ij} \times [0,1]^{n_i}$. The following example shows that this additional condition cannot be removed. Consider, x and $\log(x+1)$, which are increasing submodular and concave-extendable from $\{0,1\}$. Further, $g(z_1,z_2)=z_1z_2$ satisfies the requirements of Lemma 2.6.4 in [40]. However, $x\log(x+1)$ is strictly convex and therefore not convex-extendable from $\{0,1\}$.

Although Corollary 3.12 was presented for submodular functions defined over the $[0,1]$-hypercube, it can be generalized in a straightforward way to lattice families. This is because a lattice-submodular function can be extended to the whole hypercube while maintaining submodularity and without altering the convex envelope over the lattice family; see Corollary 3.4 in [37] for a proof. We now illustrate some applications of Corollary 3.12 in constructing relaxations for nonlinear programs

Example 3.13. *Let $x_i \in \mathbb{R}$ and $c_r(x) = -\sum_{l=1}^{L}\prod_{i=1}^{k} h_{ilr}(x_i)$ for $r \in \{1,\dots,m\}$. We define $c(x) = \bigl(c_1(x),\dots,c_m(x)\bigr)$ and assume that $h_{i'l'r}$ is a univariate, nonnegative, and concave function for each $i' \in \{1,\dots,k\}$ and $l' \in \{1,\dots,L\}$. Then, it follows from Corollary 3.12 that*

$$\operatorname{conv}\left(\left\{(x,w) \;\middle|\; w \geq c(x),\ x \in [0,1]^k\right\}\right) = \operatorname{conv}\left(\left\{(x,w) \;\middle|\; w \geq c(x),\ x \in \{0,1\}^k\right\}\right).$$

If for all i, l, and r, h_{ilr} is linear then the above result holds even if some h_{ilr} take negative values. This is because $h_{ilr}(x_i) = \operatorname{conv}_{[0,1]^{n_i}}(h_{ilr})(x_i)$. If for each i, all h_{ilr} are non-decreasing or all h_{ilr} are non-increasing then it follows from Corollary 3.12 that

$$\begin{aligned}&\operatorname{conv}\left(\left\{(x,w) \;\middle|\; w \geq c(x), x \in [0,1]^k\right\}\right)\\ &= \left\{(x,w) \;\middle|\; w_r \geq \operatorname{conv}_{[0,1]^k}(c_r)(x), r = 1,\dots,m,\ x \in \{0,1\}^k\right\},\end{aligned}$$

since we may replace x_i with $\bar{x}_i = 1 - x_i$ in the latter case. Here, $\operatorname{conv}_{[0,1]^k}(c_r)(x)$ for $r = 1,\dots,m$ are obtained from the Kuhn's triangulation of $[0,1]^k$. In other words, individual convexification of the epigraph of each $c_r(x)$ yields the simultaneous convex hull of the epigraphs of $c(x)$. □

As mentioned before, Theorem 3.6 generalizes Corollary 2.7 by allowing more general functions, sets, and measures. Perhaps the most useful case of Corollary 2.7 is when each P_i is a line-segment. This is also a special case of Example 3.13 and hence of Corollary 3.12. In addition, Example 3.13 shows that individual concave envelopes suffice to simultaneously convexify multilinear functions with negative coefficients. This is not in general true if some of the multilinear functions have positive coefficients as is shown below:

$$\begin{aligned}&\operatorname{conv}(\{(x_1,x_2,z_1,z_2) \mid z_1 \geq x_1x_2,\ z_2 \geq -x_1x_2,\ x \in [0,1]^2\})\\ &\neq \{(x_1,x_2,z_1,z_2) \mid z_1 \geq \max\{0, x_1+x_2-1\},\ z_2 \geq -\min\{x_1,x_2\},\ x \in [0,1]^2\},\end{aligned}$$

since $(0.5, 0.5, 0, -0.5)$ is feasible to the latter set whereas $z_1 + z_2 \geq 0$ for the former. This is once again because the convex envelopes of x_1x_2 and $-x_1x_2$ do not share the same triangulation.

4 Conclusions

We studied the problem of simultaneously convexifying the epigraphs/hypographs of a collection of functions. In particular, we extended the results of [39] to collections of functions. As a consequence, we proved that in order to convexify a system of multilinear equations over a cartesian product of compact convex sets, it is sufficient to restrict attention to the extreme points of the set. We interpreted orthogonal disjunctions theory from the viewpoint of convex extensions and highlighted its connection to the general case of disjunctive programming. We introduced the notion of an inclusion certificate to study function approximations. In particular, we showed that if the convex envelopes of a collection of functions share the same inclusion certificate then individual convexification leads to simultaneous convexification. This result was shown to be true when the convex envelopes share the same polyhedral subdivision, for example in the case of submodular convex-extendable functions. Other function classes that share the same polyhedral subdivision are discussed in [37] and our results extends to these functions as well. We proved that the two main results of [37] admit a common generalization that reveals interesting properties of their convex envelopes. We showed that our techniques also work when the shared inclusion certificate yields underestimators that are not necessarily convex. In this case, we showed that a direct use of disjunctive programming has the potential of improving relaxations beyond individual convexification. Future work will concentrate on implementing techniques to exploit the new relaxations proposed here in global optimization software.

References

[1] F. A. Al-Khayyal and J. E. Falk. Jointly constrained biconvex programming. *Mathematics of Operations Research*, 8:273–286, 1983.

[2] K. M. Anstreicher and S. A. Burer. Computable representations for convex hulls of low-dimensional quadratic forms. *Mathematical Programming*, pages 33–43, 2010.

[3] E. Balas. Disjunctive programming: Properties of the convex hull of feasible points. *Discrete Applied Mathematics*, 89(1-3):3–44, 1998. Original manuscript was published as a technical report in 1974.

[4] X. Bao, N. V. Sahinidis, and M. Tawarmalani. Multiterm polyhedral relaxations for nonconvex, quadratically constrained quadratic programs. *Optimization Methods & Software*, 24:485–504, 2009.

[5] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization*. MPS-SIAM Series on Optimization, Philadelphia, first edition, 2001. 488 pages.

[6] H. P. Benson. Concave envelopes of monomial functions over rectangles. *Naval Research Logistics*, 51:467–476, 2004.

[7] H. P. Benson. On the construction of convex and concave envelope formulas for bilinear and fractional functions on quadrilaterals. *Computational Optimization and Applications*, 27:5–22, 2007.

[8] S. Burer and A. N. Letchford. On nonconvex quadratic programming with box constraints. *Siam Journal on Optimization*, 20:1073–1089, 2009.

[9] S. Ceria and J. Soares. Convex programming for disjunctive convex optimization. *Mathematical Programming*, 86A:595–614, 1999.

[10] D. Coppersmith, O. Günlük, J. Lee, and J. Leung. A polytope for a product of real linear functions in 0/1 variables. *IBM Research Report*, 2003.

[11] Y. Crama. Recognition problems in polynomials in $0-1$ programming. *Mathematical Programming*, 44:139–155, 1989.

[12] Y. Crama. Concave extensions for non-linear $0-1$ maximization problems. *Mathematical Programming*, 61:53–60, 1993.

[13] C. A. Floudas and C. E. Gounaris. A review of recent advances in global optimization. *Journal of Global Optimization*, 45:3–38, 2009.

[14] J-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I*. Springer-Verlag, Berlin, 1993.

[15] J-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Springer, 2004.

[16] M. Jach, D. Michaels, and R. Weismantel. The convex envelope of $(n-1)$-convex functions. *SIAM Journal on Optimization*, 19:1451–1466, 2008.

[17] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11:796–817, 2001.

[18] C. W. Lee. Subdivisions and triangulations of polytopes. In J. E. Goodman and J. O'Rourke, editors, *Handbook of discrete and computational geometry*, chapter 14. CRC Press, 1997.

[19] J. Linderoth. A simplicial branch-and-bound algorithm for solving quadratically constrained quadratic programs. *Mathematical Programming*, 103:251–282, 2005.

[20] M. Locatelli and F. Schoen. On convex envelopes and underestimators for bivariate functions. available at `http://www.optimization-online.org/DB_FILE/2009/11/2462.pdf`, 2009.

[21] L. Lovász. Submodular functions and convexity. In M. Grötschel and B. Korte, editors, *Mathematical Programming: The State of the Art*, pages 235–257. Springer, 1982.

[22] C. A. Meyer and C. A. Floudas. Convex envelopes of edge-concave functions. *Mathematical programming*, 103:207–224, 2005.

[23] M. W. Padberg. The boolean quadric polytope: Some characteristics, facets and relatives. *Mathematical programming*, 45:139–172, 1989.

[24] P. A. Parillo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96:307–321, 2003.

[25] A. D. Rikun. A convex envelope formula for multilinear functions. *Journal of Global Optimization*, 10:425–437, 1997.

[26] R. T. Rockafellar. *Convex Analysis*. Princeton Mathematical Series. Princeton University Press, 1970.

[27] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. A Series of Comprehensive Studies in Mathematics. Springer, Berlin, 1998.

[28] C.-D. Rodrigues, D. Quadri, P. Michelon, and S. Gueye. A t-linearization scheme to exactly solve 0-1 quadratic knapsack problems. In *Proceedings of the European Workshop on Mixed Integer Programming*, pages 251–260. CIRM, Marseille, France, 2010.

[29] H. S. Ryoo and N. V. Sahinidis. Analysis of bounds for multilinear functions. *J. Global Optimization*, 19:403–424, 2001.

[30] H. D. Sherali. Convex envelopes of multilinear functions over a unit hypercube and over special discrete sets. *Acta Mathematica Vietnamica*, 22:245–270, 1997.

[31] H. D. Sherali and W. P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal of Discrete Mathematics*, 3:411–430, 1990.

[32] C. De Simeone. The cut polytope and the boolean quadric polytope. *Discrete Mathematics*, 79:71–75, 1990.

[33] R. Stubbs and S. Mehrotra. A branch and cut method for $0-1$ mixed convex programming. *Mathematical Programming*, 86:515–532, 1999.

[34] F. Tardella. Existence and sum decomposition of vertex polyhedral convex envelopes. *Optimization Letters*, 2:363–375, 2008.

[35] M. Tawarmalani. Convexification and global optimization of nonlinear programs. Presented at Workshop on Integer Programming and Continuous Optimization, Chemnitz, 2004.

[36] M. Tawarmalani, J.-P. P. Richard, and K. Chung. Strong valid inequalities for orthogonal disjunctions and bilinear covering sets. *Mathematical Programming*, 124:481–512, 2010.

[37] M. Tawarmalani, J.-P. P. Richard, and C. Xiong. Deriving polyhedral envelopes through polyhedral subdivisions. Krannert Working Paper, 1234 (available at `http://www.optimization-online.org/DB_HTML/2010/06/2640.html`, 2010.

[38] M. Tawarmalani and N. V. Sahinidis. Semidefinite relaxations of fractional programs via novel techniques for constructing convex envelopes of nonlinear functions. *Journal of Global Optimization*, 20:137–158, 2001.

[39] M. Tawarmalani and N. V. Sahinidis. Convex extensions and convex envelopes of l.s.c. functions. *Mathematical Programming*, 93:247–263, 2002.

[40] D. M. Topkis. *Supermodularity and Complementarity*. Princeton University Press, Princeton, NJ, 1998.