Online Appendix

EC.1. Lattice Theory and Supermodularity

The proofs in the current paper use several concepts from the theory of lattice programming and supermodular functions, which we formally define here. The presentation follows closely Milgrom and Shannon (1994) and Topkis (1998), to which we direct the interested reader for proofs and a detailed treatment of the subject.

Let X be any set equipped with a transitive, reflexive, antisymmetric order relation $\geq$. For elements $\boldsymbol{x}, \boldsymbol{y} \in X$, let $\boldsymbol{x} \vee \boldsymbol{y}$ denote the least upper bound (or the join) of $\boldsymbol{x}$ and $\boldsymbol{y}$ (if it exists), and let $\boldsymbol{x} \wedge \boldsymbol{y}$ denote the greatest lower bound (or the meet) of $\boldsymbol{x}$ and $\boldsymbol{y}$ (if it exists).

Definition EC.1. The set X is a *lattice* if for every pair of elements $\boldsymbol{x}, \boldsymbol{y} \in X$, the join and the meet exist and are elements of X.

Similarly, $S \subset X$ is a *sublattice* if it is closed under the join and meet operations. In our treatment, the typical lattices under consideration are subsets of the hypercube $\mathcal{H}_n = [0,\, 1]^n$. Therefore, the operations $\geq$ and $\leq$ are understood in component-wise fashion, and $\wedge$ ($\vee$) are given by component-wise minimum (maximum).

Our analysis requires stating when the sets of maximizers (or minimizers) of a function is increasing or decreasing in particular state variables. To compare two such sets, we use the *strong set order* introduced by Veinott (1989). If X is a lattice with the relation $\geq$, and Y, Z are elements of the power set of X, we say that $Y \geq Z$ if, for every $\boldsymbol{y} \in Y$ and $\boldsymbol{z} \in Z$, $\boldsymbol{y} \vee \boldsymbol{z} \in Y$ and $\boldsymbol{y} \wedge \boldsymbol{z} \in Z$. For instance, $[2,4] \geq [1,3]$, but $[1,5] \not\geq [2,4]$ and $[2,4] \not\geq [1,5]$. Analogous definitions hold for the $\leq$ relation.

Definition EC.2. For a lattice $S \subseteq \mathbb{R}^n$, a function $f : S \to \mathbb{R}$ is said to be supermodular if $f(\boldsymbol{x}' \wedge \boldsymbol{x}'') + f(\boldsymbol{x}' \vee \boldsymbol{x}'') \geq f(\boldsymbol{x}') + f(\boldsymbol{x}'')$, for all $\boldsymbol{x}'$ and $\boldsymbol{x}'' \in S$.

Similarly, a function f is called *submodular* if $-f$ is supermodular. Supermodular and submodular functions have been studied extensively in various fields, such as physics (Choquet 1954), economics (Schmeidler (1986), Topkis (1998), Milgrom and Shannon (1994)), combinatorial optimization (Lovász (1982), Schrijver (2003), Fujishige (2005)), or mathematical finance (Föllmer and Schied 2004), to name only a few. They also play a central role in our treatment, since they admit a compact characterization for their concave envelopes.

Apart from the definition, several methods are known for testing whether a function is supermodular. One such test, applicable to functions $f : \mathbb{R}^n \to \mathbb{R}$ that are twice continuously differentiable, is $\frac{\partial^2 f}{\partial x_i \partial x_j} \geq 0, \forall\, i \neq j \in \{1, \ldots, n\}$. Two particular examples that occur often throughout our analysis are the following.

EXAMPLE EC.1 (**Lemma 2.6.2 in Topkis (1998)**). Suppose $Y \subseteq \mathbb{R}$ is a convex set, X is a sublattice of $\mathbb{R}^n$, $\boldsymbol{a} \in \mathbb{R}^n$ is a vector that satisfies $\boldsymbol{a}^T\boldsymbol{x} \in Y, \forall \boldsymbol{x} \in X$, $g: Y \to \mathbb{R}$, and $f(\boldsymbol{x}) = g(\boldsymbol{a}^T\boldsymbol{x})$. Then, f is supermodular in $\boldsymbol{x}$ on X if one of the following conditions holds:

- $\boldsymbol{a} \geq 0$ and g is convex on Y.
- $n = 2$, $\operatorname{sign}(a_1) = -\operatorname{sign}(a_2)$, and g is concave on Y.

We note that the results above hold even when g is not twice continuously differentiable. For an overview of many other relevant classes of supermodular functions, we direct the interested reader to Topkis (1998) and Fujishige (2005).

As suggested earlier, we are interested in characterizing conditions when the set of maximizers (or minimizers) of a function is increasing (decreasing) with particular problem parameters. The following result provides a fairly general set of such conditions.

THEOREM EC.1 (**Theorem 2.8.2 in Topkis (1998)**). *If X and T are lattices, S is a sublattice of $X \times T$, S_t is the section of S at $\boldsymbol{t} \in T$, and $f(\boldsymbol{x}, \boldsymbol{t})$ is supermodular in $(\boldsymbol{x}, \boldsymbol{t})$ on S, then $\arg\max_{\boldsymbol{x} \in S_t} f(\boldsymbol{x}, \boldsymbol{t})$ is increasing in $\boldsymbol{t}$ on $\{\boldsymbol{t} : \arg\max_{\boldsymbol{x} \in S_t} f(\boldsymbol{x}, \boldsymbol{t}) \neq \emptyset\}$.*

We note that more general conditions are known in the literature, based on concepts such as quasisupermodular functions (see, e.g., Milgrom and Shannon (1994)). However, the result above suffices for our purposes in the present paper. To see how it can be used in a concrete setting, we include the following example, which is a well-known result in operations research (see, e.g., Example 8-15 in Heyman and Sobel (1984), Proposition 3.1 in Bensoussan et al. (1983) or Theorem 3.10.2 in Topkis (1998)), which is very useful in our analysis of Problem 1. We include its derivation here for completeness.

LEMMA EC.1. *Let $f(x,u) = c(u) + g(x+u)$, where $c, g : \mathbb{R} \to \bar{\mathbb{R}}$ are arbitrary proper convex functions. Then, $\arg\min_u f(x,u)$ is decreasing in x, and $x + \arg\min_u f(x,u)$ is increasing in x.*

Proof. Assume first that c, g are real-valued. Note that

$$\min_u \left[c(u) + g(x+u)\right] = -\max_u \left[-c(u) - g(x+u)\right] \overset{(r \overset{\text{def}}{=} -u)}{=} -\max_r \left[-c(-r) - g(x-r)\right].$$

Since $-g$ is concave, the function $-c(-r) - g(x-r)$ is supermodular in (x,r) on the lattice $\mathbb{R} \times \mathbb{R}$, by the second condition of Example EC.1. Therefore, by Theorem EC.1, $\arg\max_r \left[-c(-r) - g(x-r)\right]$ is increasing in x, which implies that $\arg\min_u f(x,u)$ is decreasing in x. Similarly, letting $y \overset{\text{def}}{=} x + u$, it can be argued that the set $\arg\min_y f(x,y)$ is increasing in x, which concludes the proof.

The monotonicity conditions above would hold even if c, g were extended-real, since this would be equivalent to adding constraints of the form $L \leq u \leq H$ and/or $\tilde{L} \leq x + u \leq \tilde{H}$, and considering real-valued functions on a modified lattice (Rockafellar 1970), which would still fall in the realm of Theorem EC.1. $\square$

EC.2. Convex and Concave Envelopes

Our proofs make use of several known results concerning concave envelopes of functions, which are summarized below. The notation and statements follow quite closely those of Tardella (2008) and Tawarmalani et al. (2010), to which we refer the interested reader for a more comprehensive overview and references.

DEFINITION EC.3. Consider a function $f : S \to \mathbb{R}$, where S is a non-empty convex subset of $\mathbb{R}^n$. The function $\text{conc}_S(f) : S \to \mathbb{R}$ is said to be the *concave envelope* of f over S if and only if

(i) $\text{conc}_S(f)$ is concave over S

(ii) $\text{conc}_S(f)(\boldsymbol{x}) \geq f(\boldsymbol{x}), \forall \boldsymbol{x} \in S$

(iii) $\text{conc}_S(f)(\boldsymbol{x}) \leq h(\boldsymbol{x})$, for any concave $h(\boldsymbol{x})$ satisfying $h(\boldsymbol{x}) \geq f(\boldsymbol{x})$.

In words, $\text{conc}_S(f)$ is the point-wise smallest concave function defined on S that over-estimates f. An example is included in Figure EC.1. In a similar fashion, one can define the *convex envelope* of f, denoted by $\text{conv}_S(f)$, as the point-wise largest convex under-estimator of f on S. For the rest of the exposition, we focus attention on concave envelopes, but all the concepts and results can be translated in a straightforward manner to convex envelopes, by recognizing that $\text{conv}_S(f) = -\text{conc}_S(-f)$.

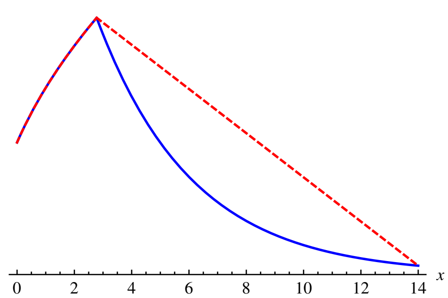


Figure EC.1 Example of a function $f : [0,14] \to \mathbb{R}$ (solid line) and its concave envelope $\text{conc}_{[0,14]}(f)$ (dashed line).

One of the main reasons for the interest in concave envelopes is the fact that the set of global maxima of f is contained in the set of global maxima of $\text{conc}_S(f)$, and the two maximum values coincide. Expressing the concave envelope of a function is a difficult task in general, and even evaluating $\text{conc}_S(f)$ at a particular point $\boldsymbol{x}$ can be as hard as minimizing the function f (Tardella 2008). In some cases, however, concave envelopes can be constructed by restricting attention to a subset of the points in the domain S. One such instance, particularly relevant to the treatment in our paper, is summarized in the following definition.

Definition EC.4. A function $f: P \to \mathbb{R}$, where P is a non-empty polytope, is said to be *concave-extendable from the set* $S \subset P$ if the concave envelope of f over P is the same as the concave envelope of the function $f|_S$ over P, where

$$f|_S(\boldsymbol{x}) \stackrel{\text{def}}{=} \begin{cases} f(\boldsymbol{x}), & \boldsymbol{x} \in S \\ -\infty, & \text{otherwise.} \end{cases}$$

When $S = \text{ext}(P)$, we say that f is concave-extendable from the vertices of P. Such functions are known to admit piece-wise affine concave envelopes, which further generate a relevant partition of the polytope P (this connection and other relevant results are included in Section EC.2.1). A natural question, in this context, is how to recognize a function that is concave-extendable from vertices. To the best of our knowledge, the most general characterization in the literature seems to be the following result from Tardella (2008).

Lemma EC.2 **(Corollary 3 in Tardella (2008))**. *Let $\mathscr{D}$ be a set of vectors in $\mathbb{R}^n$ parallel to some edges of the polytope P. Let f be a function that is convex5. on P along all directions in $\mathscr{D}$, and let S denote the union of the faces of P (including the zero-dimensional faces* $\text{ext}(P)$*) that do not have any edge parallel to a direction in $\mathscr{D}$. Then,* $\text{conc}_P(f) = \text{conc}_S(f)$*. In particular, if f is edge-convex on P (i.e., $\mathscr{D}$ is maximal), then f is concave-extendable from* $\text{ext}(P)$*.*

This characterization yields several interesting functions. For instance, any f that is convex on P is concave-extendable from $\text{ext}(P)$; when P is a hypercube, any f that is component-wise convex is also concave-extendable from $\text{ext}(P)$ (an example in the latter category often studied in the literature is the case of monomials). For more examples and references, the interested reader can check Tardella (2008) and Tawarmalani et al. (2010).

EC.2.1. Concave Envelopes of Concave-Extendable Functions

Functions that are concave-extendable from vertices are known to admit polyhedral concave envelopes, i.e., concave envelopes that are given by the minimum of a finite collection of affine functions. The resulting concave envelopes also induce a polyhedral subdivision6. of the domain P, which is relevant for several results in our treatment. To illustrate this connection, following Tawarmalani et al. (2010), let $V \in \mathbb{R}^{n \cdot |\text{ext}\, P|}$ denote a matrix with columns V_i given by the vertices of P, let $f(V) \stackrel{\text{def}}{=} \big(f(\boldsymbol{V}_1), f(\boldsymbol{V}_2), \ldots, f(\boldsymbol{V}_{|\text{ext}\, P|})\big)$, and consider the following primal-dual pair of linear programs

$$\begin{aligned} P(\boldsymbol{x}) \stackrel{\text{def}}{=} \min_{\boldsymbol{a},b}\ & \boldsymbol{a}^T\boldsymbol{x} + b \\ \text{s.t. } & \boldsymbol{a}^T V + \boldsymbol{e}^T b \geq f(V) \\ & \boldsymbol{a} \in \mathbb{R}^n,\, b \in \mathbb{R} \end{aligned} \qquad \begin{aligned} D(\boldsymbol{x}) \stackrel{\text{def}}{=} \max_{\boldsymbol{\lambda}}\ & f(V)^T\boldsymbol{\lambda} \\ \text{s.t. } & V\boldsymbol{\lambda} = \boldsymbol{x} \\ & \boldsymbol{e}^T\boldsymbol{\lambda} = 1 \\ & \boldsymbol{\lambda} \geq 0 \end{aligned} \tag{EC.1}$$

It can be shown (see Rockafellar (1970) or Tawarmalani et al. (2010)) that the optimal values in both programs are finite, and equal to $\operatorname{conc}_P(f)(\boldsymbol{x})$. Moreover, let $\mathcal{D}_{f,P}$ denote the feasible region of the primal program (which only depends on f and P, and is independent of $\boldsymbol{x}$), and, for a given $(\boldsymbol{a},b) \in \mathcal{D}_{f,P}$, let $J(\boldsymbol{a},b)$ be the index set of constraints of $\mathcal{D}_{f,P}$ that are tight at $(\boldsymbol{a},b)$, let $V\big(J(\boldsymbol{a},b)\big)$ be the matrix obtained from V by keeping the columns in $J(\boldsymbol{a},b)$, and let $R(\boldsymbol{a},b) \stackrel{\text{def}}{=} \operatorname{conv}\big(V\big(J(\boldsymbol{a},b)\big)\big)$. Then, the following theorem summarizes several relevant properties of the linear programs above, and their connection with $\operatorname{conc}_P(f)$.

Theorem EC.2 **(Theorem 2.4 in Tawarmalani et al. (2010))**. *Consider a function $f: P \to \mathbb{R}$ which is concave-extendable from the vertices of P, where P is a full-dimensional polytope in $\mathbb{R}^n$. The following results hold:*

1. *The optimal values in $P(\boldsymbol{x})$ and $D(\boldsymbol{x})$ are the same, and equal to $\operatorname{conc}_P(f)(\boldsymbol{x})$.*
2. *Let $(\bar{\boldsymbol{a}},\bar{b}) \in \operatorname{ext}(\mathcal{D}_{f,P})$. Then, $(\bar{\boldsymbol{a}},\bar{b})$ is optimal for $P(\boldsymbol{x})$ if and only if $\boldsymbol{x} \in R(\bar{\boldsymbol{a}},\bar{b})$. Further, the extreme points of $\mathcal{D}_{f,P}$ are in one-to-one correspondence with the non-vertical facets of $\operatorname{conc}_P(f)$.*
3. *For any $(\bar{\boldsymbol{a}},\bar{b}) \in \operatorname{ext}(\mathcal{D}_{f,P})$, the inequality $\bar{\boldsymbol{a}}^T\boldsymbol{x} + \bar{b} \geq f(\boldsymbol{x})$ defines a facet of $\operatorname{conc}_P(f)$ over $R(\bar{\boldsymbol{a}},\bar{b})$.*
4. *$\mathcal{R}_{f,P} \stackrel{\text{def}}{=} \big\{R(\bar{\boldsymbol{a}},\bar{b}) : (\bar{\boldsymbol{a}},\bar{b}) \in \operatorname{ext}(\mathcal{D}_{f,P})\big\}$ is a polyhedral subdivision of $\operatorname{conv}(V)$, and $\operatorname{conc}_P(f)$ can be computed by interpolating f affinely over each element of $\mathcal{R}_{f,P}$.*

Proof. This is a direct adaptation of Theorem 2.4 in Tawarmalani et al. (2010), to which we direct the reader for a proof and discussion. □

The previous theorem essentially states that $\operatorname{conc}_P(f)$ is given by affine interpolations of f over a particular polyhedral subdivision of P, given by the polytopes $R(\bar{\boldsymbol{a}},\bar{b})$, for $(\bar{\boldsymbol{a}},\bar{b}) \in \operatorname{ext}(\mathcal{D}_{f,P})$ (also known as the *linearity domains* of $\operatorname{conc}_P(f)$ (Tardella 2008)). From this result, utilizing the same notation as before, one can derive the following characterization concerning the problem of maximizing f over P.

Corollary EC.1. *For any full-dimensional polytope P and any function $f: P \to \mathbb{R}$ that is concave-extendable from $\operatorname{ext}(P)$, we have*

$$\begin{aligned} \max_{\boldsymbol{x}\in P} f(\boldsymbol{x}) = \max_{\boldsymbol{x},t}\ & t \\ \text{s.t. }\ & t \leq \boldsymbol{a}^T\boldsymbol{x} + b, \forall\, (\boldsymbol{a},b) \in \operatorname{ext}(\mathcal{D}_{f,P}). \\ & \boldsymbol{x} \in P. \end{aligned}$$

Proof. For any function f, we have $\max_{\boldsymbol{x}\in P} f(\boldsymbol{x}) = \max_{\boldsymbol{x}\in P} \operatorname{conc}_P(f)(\boldsymbol{x})$. By Theorem EC.2, the latter function is exactly given by

$$\operatorname{conc}_P(f)(\boldsymbol{x}) = \min_{(\boldsymbol{a},b)\in \operatorname{ext}(\mathcal{D}_{f,P})} \boldsymbol{a}^T\boldsymbol{x} + b,$$

which immediately leads to the conclusion of the corollary. □

One particular case that is very relevant in our analysis is that of concave-extendable functions f defined on the unit hypercube, i.e., $P = \mathcal{H}_n = [0,1]^n$, which are also *supermodular*. It turns out that the concave envelope of any such function can be compactly described by the *Lovász extension* of the function f.

Definition EC.5 (**Lovász (1982)**). Given any $\boldsymbol{x} \in \mathcal{H}_n$, find a permutation $\pi \in \Pi(\{1,\dots,n\})$ such that $x_{\pi(1)} \geq x_{\pi(2)} \geq \cdots \geq x_{\pi(n)}$. Then, the *Lovász extension* of the function $f : \mathcal{H}_n \to \mathbb{R}$ at the point $\boldsymbol{x}$ is given by

$$\begin{aligned} f^{\mathcal{L}}(\boldsymbol{x}) &\stackrel{\text{def}}{=} \left(1 - x_{\pi(1)}\right) f(\mathbf{0}) + \sum_{j=1}^{n-1} (x_{\pi(j)} - x_{\pi(j+1)}) f\left(\sum_{r=1}^{j} \mathbf{1}_{\pi(r)}\right) + x_{\pi(n)} f\left(\sum_{r=1}^{n} \mathbf{1}_{\pi(r)}\right) \\ &= f(\mathbf{0}) + \sum_{i=1}^{n} \left[f\left(\sum_{j=1}^{i} \mathbf{1}_{\pi(j)}\right) - f\left(\sum_{j=1}^{i-1} \mathbf{1}_{\pi(j)}\right)\right] x_{\pi(i)}. \end{aligned} \tag{EC.2}$$

It can be seen from the definition that the Lovász extension of f is given by an affine interpolation of f on simplicies of the form $\Delta_\pi \stackrel{\text{def}}{=} \operatorname{conv}\left(\left\{\mathbf{0} + \sum_{j=1}^{k} \mathbf{1}_{\pi(j)} \,:\, k = 0,\dots,n\right\}\right)$. The collection of corresponding simplicies $\{\Delta_\pi\}_{\pi\in\Pi(\{1,\dots,n\})}$ is known as the *Kuhn triangulation* of the hypercube. Using a result by Lovász (1982), one can show the following remarkable fact.

Theorem EC.3 (**Theorem 3.3 in Tawarmalani et al. (2010)**). *Consider a function $f : \mathcal{H}_n \to \mathbb{R}$. The concave envelope of f over $\mathcal{H}_n$ is given by $f^{\mathcal{L}}$ if and only if f is supermodular when restricted to $\{0,1\}^n$ and concave-extendable from $\{0,1\}^n$.*

In the context of Theorem EC.2, this result immediately yields the following corollary, which provides a full characterization of the concave envelope of supermodular and concave-extendable functions on hypercubes.

Corollary EC.2. *Consider a function $f : \mathcal{H}_n \to \mathbb{R}$ that is supermodular on $\{0,1\}^n$ and concave-extendable from $\{0,1\}^n$. Then the following results hold:*

1. *The concave envelope of f on $\mathcal{H}_n$ is given by*

$$\operatorname{conc}_{\mathcal{H}_n}(f)(\boldsymbol{x}) = f(\mathbf{0}) + \min_{\pi\in\Pi(\{1,\dots,n\})} \sum_{i=1}^{n} \left[f\left(\sum_{j=1}^{i} \mathbf{1}_{\pi(j)}\right) - f\left(\sum_{j=1}^{i-1} \mathbf{1}_{\pi(j)}\right)\right] x_{\pi(i)}.$$

2. *The set of inequalities $\boldsymbol{a}^T\boldsymbol{x} + b \geq f(\boldsymbol{x})$ defining non-vertical facets of* $\operatorname{conc}_{\mathcal{H}_n}(f)$ *is given by*

$$\begin{aligned} \operatorname{ext}(\mathcal{D}_{f,P}) = \Bigg\{ (\boldsymbol{a}, b) \in \mathbb{R}^{n+1} \,:\, b = f(\mathbf{0}),\ \boldsymbol{a} = \sum_{i=1}^{n} \left[f\left(\sum_{j=1}^{i} \mathbf{1}_{\pi(j)}\right) - f\left(\sum_{j=1}^{i-1} \mathbf{1}_{\pi(j)}\right)\right] \mathbf{1}_{\pi(i)}, \\ \text{for } \pi\in\Pi(\{1,\dots,n\}) \Bigg\}. \end{aligned}$$

3. *The polyhedral subdivision $\mathcal{R}_{f,\mathcal{H}_n}$ of $\mathcal{H}_n$ yielding the concave envelope is exactly the Kuhn triangulation.*

Proof. The proof is a direct application of Theorem EC.2 and Theorem EC.3 and is omitted (see Tawarmalani et al. (2010) for complete details). □

In fact, the above results hold for the more general case of polytopes whose extreme points are *integer sublattices of* $\{0,1\}^n$. Any such polytope P is given by

$$P = \mathcal{H}_n \cap \{\boldsymbol{x} : x_i \geq x_j, \forall (i,j) \in E\} \cap \{\boldsymbol{x} : x_i = 0, \forall i \in \mathcal{I}_0\} \cap \{\boldsymbol{x} : x_i = 1, \forall i \in \mathcal{I}_1\},$$

for some $E \subseteq \{1,\ldots,n\}^2$ and $\mathcal{I}_{0,1} \subseteq \{1,\ldots,n\}$ (see Tawarmalani et al. (2010) and the original reference Grötschel et al. (1988) for details). For any such lattice, the Definition EC.5 of the Lovász extension is modified, by only including permutations that are compatible with the pre-order on P, i.e.,

$$\Pi^P = \{\pi \in \Pi(\{1,\ldots,n\}) : \pi^{-1}(i) \leq \pi^{-1}(j), \forall (i,j) \in E\}.$$

In other words, if $(i,j) \in E$, then i always appears before j in the permutations in Π^P.

All the results in Theorem EC.3 and Corollary EC.2 then hold with the same modification for the set of permutations (see Tawarmalani et al. (2010) for details).

EC.2.2. Summability of Concave Envelopes

For any two functions f, g defined on a polytope $P \subseteq \mathbb{R}^n$, it is always true that $\text{conc}_P(f+g) \leq \text{conc}_P(f) + \text{conc}_P(g)$, and equality holds if one of the two functions is affine (Tardella 2008). In practice, it is relevant to seek sufficient conditions on f, g and P that guarantee equality, since these would allow constructing the concave envelope of a (complex) sum of functions by characterizing the envelopes of individual components. The following result provides a general characterization of such conditions.

THEOREM EC.4 **(Theorem 3 in Tardella (2008))**. *For a polytope $P \subset \mathbb{R}^n$, let f, g be two functions that are concave-extendable from* $\text{ext}(P)$*, and let $\mathcal{R}_{f,P} = \{F_i : i \in \mathcal{I}\}$, and $\mathcal{R}_{g,P} = \{G_j : j \in \mathcal{J}\}$ denote the polyhedral subdivisions of P that yield the linearity domains of* $\text{conc}_P(f)(\boldsymbol{w})$ *and* $\text{conc}_P(g)(\boldsymbol{w})$*, respectively. Then, the following conditions are equivalent:*

(i) $\text{conc}_P(f) + \text{conc}_P(g)$ *is concave-extendable from vertices*

(ii) $\text{conc}_P(f) + \text{conc}_P(g) = \text{conc}_P(f+g)$

(iii) $F_i \cap G_j$ *has all vertices in* $\text{ext}(P), \forall i \in \mathcal{I}, \forall j \in \mathcal{J}$.

This theorem provides sufficient conditions for the concave envelope of a sum of two functions to be exactly given by the sum of the two separate concave envelopes. In view of the discussion in Section EC.2.1, the following corollary summarizes a particularly relevant class of functions that satisfy these requirements.

COROLLARY EC.3. *For any polytope $P \subseteq \mathcal{H}_n$ such that $\text{ext}(P)$ is a sublattice of $\{0,1\}^n$, and any finite collection of functions $h_i : P \to \mathbb{R}$, $i \in \mathcal{I}$, that are convex and supermodular on $\text{ext}(P)$,*

$$\text{conc}_P\left(\sum_{i\in\mathcal{I}} h_i\right) = \sum_{i\in\mathcal{I}} \text{conc}_P\left(h_i\right).$$

Proof. A sum of convex and supermodular functions is also convex and supermodular. By Theorem EC.3, the concave envelope of any convex and supermodular function is given by the Lovász extension, which is an affine interpolation of the function on the simplicies Δ_π in the Kuhn triangulation. Applying this result for each h_i and for the sum immediately yields the result. □

EC.3. Technical Results

LEMMA 2. *Suppose $f^\star : P \to \mathbb{R}$ is convex on P and supermodular on $\text{ext}(P)$. Consider an arbitrary $\hat{\boldsymbol{w}} \in \text{ext}(P) \cap \arg\max_{\boldsymbol{w}\in P} f^\star(\boldsymbol{w})$, and let $\boldsymbol{g}^\pi$ be given by (11). Then,*

1. *For any $\boldsymbol{w} \in P$, we have*

$$f^\star(\boldsymbol{w}) \le f^\star(\hat{\boldsymbol{w}}) + (\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g}^\pi, \ \forall \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}.$$

2. *There exists a set of convex weights $\{\lambda_\pi\}_{\pi\in\mathscr{S}_{\hat{\boldsymbol{w}}}}$ such that $\boldsymbol{g} = \sum_{\pi\in\mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \boldsymbol{g}^\pi$ satisfies*

$$(\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g} \le 0, \forall \boldsymbol{w} \in P.$$

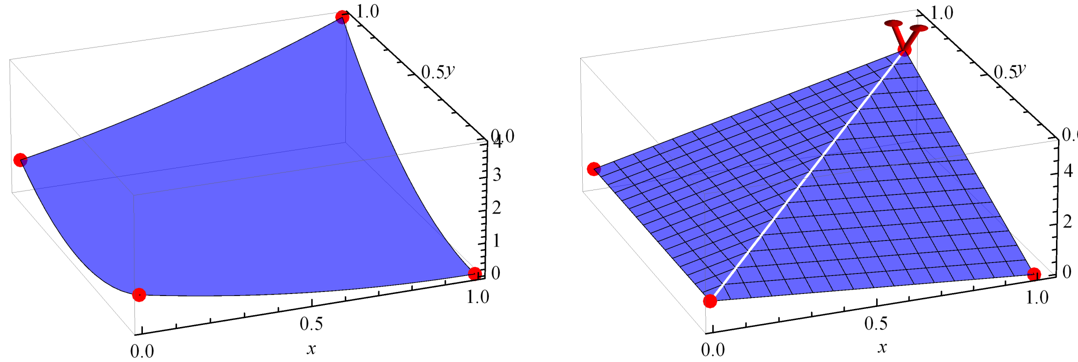
Proof. Note first that the set $\text{ext}(P) \cap \arg\max_{\boldsymbol{w}\in P} f^\star(\boldsymbol{w})$ is nonempty, since $f^\star$ is convex. Therefore, since the vertices of $\text{ext}(P)$ are integral, $\hat{\boldsymbol{w}} = \mathbf{1}_S$ for some $S \subseteq \{1,\dots,n\}$, and $\hat{\boldsymbol{w}}$ belongs to the intersection of all simplices Δ_π that correspond to permutations π in the set7. $\mathscr{S}_{\hat{\boldsymbol{w}}} \stackrel{\text{def}}{=} \Pi^P(S) \times \Pi^P(S^C)$. Here, $\Pi^P(S)$ is any permutation of the elements in S that is consistent with the pre-order on P. For instance, if $\{i,j\} \subseteq S$ for some $(i,j) \in E$, then $\Pi^P(S)$ contains only permutations of S such that i appears before j.

[**1**] To argue the first claim, note that (by (10) and (11) in Lemma 1), the set $\{\boldsymbol{g}^\pi : \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}\}$ contains valid supergradients of the concave function $\text{conc}_{\mathcal{W}}(f^\star)$ at $\hat{\boldsymbol{w}}$. As such, the supergradient inequality applied to the concave function $\text{conc}_{\mathcal{W}}(f^\star)$ at $\hat{\boldsymbol{w}}$ yields

$$\text{conc}_{\mathcal{W}}(f^\star)(\boldsymbol{w}) \le \text{conc}_{\mathcal{W}}(f^\star)(\hat{\boldsymbol{w}}) + (\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g}^\pi, \ \forall \boldsymbol{w} \in P, \ \forall \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}.$$

The desired inequality follows since $f^\star(\hat{\boldsymbol{w}}) = \text{conc}_{\mathcal{W}}(f^\star)(\hat{\boldsymbol{w}})$, and $f^\star(\boldsymbol{w}) \le \text{conc}_{\mathcal{W}}(f^\star)(\boldsymbol{w}), \forall \boldsymbol{w} \in P$. For an example illustrating the relation, please refer to Figure EC.2.

[**2**] The intuition behind the second claim is geometric. We essentially seek to show that, at any vertex $\hat{\boldsymbol{w}}$ maximizing $f^\star$, there exists a supergradient of $f^\star$ that: (i) is obtained as a convex combination of the supergradients corresponding to non-vertical facets of the concave envelope



(a) $f^\star : \mathcal{H}_2 \to \mathbb{R}$ (b) $\mathrm{conc}_{\mathcal{H}_2}(f^\star)$

Figure EC.2 A convex and supermodular function (a) and its concave envelope (b). Here, $\mathcal{W} = \mathcal{H}_2$, $\Pi^{\mathcal{W}} = \{(1,2),(2,1)\}$, and $\mathcal{K}^{\mathcal{W}} = \{\Delta_{(1,2)}, \Delta_{(2,1)}\}$, where $\Delta_{(1,2)} = \mathrm{conv}(\{(0,0),(1,0),(1,1)\}$ and $\Delta_{(2,1)} = \mathrm{conv}(\{(0,0),(0,1),(1,1)\})\}$. The plot in Figure (b) also shows the two normals of non-vertical facets of $\mathrm{conc}_{\mathcal{W}}(f^\star)$, corresponding to $\boldsymbol{g}^{(1,2)}$ and $\boldsymbol{g}^{(2,1)}$.

that attain the value $f^\star(\hat{\boldsymbol{w}})$ at $\hat{\boldsymbol{w}}$, and (ii) is a direction of decrease. In the example of Figure EC.2, this means that there is a convex combination $\boldsymbol{g}$ of $\boldsymbol{g}^{(1,2)}$ and $\boldsymbol{g}^{(2,1)}$ at $\hat{\boldsymbol{w}} = (1,1)$, such that $\boldsymbol{g} \geq 0$.

In order to construct the candidate vector $\boldsymbol{g}$, let us first consider the problem of maximizing $f^\star$ on P. By Corollary EC.1 in the Online Appendix (and also from Lemma 1), we have

$$
\begin{aligned}
\max_{\boldsymbol{w}\in P} f^\star(\boldsymbol{w}) = \max_{t,\,\boldsymbol{w}} \;& t \\
\text{s.t. } & t \leq (\boldsymbol{g}^\pi)^T \boldsymbol{w} + g_0,\ \forall\, (\boldsymbol{g}^\pi, g_0) \in \mathrm{ext}(\mathcal{D}_{f^\star,P}). \quad (*) \\
& \boldsymbol{w} \in P.
\end{aligned}
$$

If we denote the optimal value by $J^\star$, then $t = J^\star$ and $\boldsymbol{w} = \hat{\boldsymbol{w}}$ are optimal in the program on the right. Furthermore, the only constraints $(*)$ that are tight at $\hat{\boldsymbol{w}}$ are those corresponding to $\pi \in \mathscr{S}_{\hat{\boldsymbol{w}}}$. As such, by adding and subtracting terms $(\boldsymbol{g}^\pi)^T \hat{\boldsymbol{w}}$, we have that the left program (in the following primal-dual pair) is equivalent to the problem above:

$$
\begin{array}{rll|l}
& \max\limits_{t,\,\boldsymbol{w}} \; t & = & \min\limits_{\lambda_\pi,\boldsymbol{\eta},\mu_{i,j}} \; \mathbf{1}'\boldsymbol{\eta} + \sum\limits_{\pi\in\mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \left(J^\star - \hat{\boldsymbol{w}}'\boldsymbol{g}^\pi\right) \\
\lambda_\pi \to & t \leq J^\star + (\boldsymbol{w}-\hat{\boldsymbol{w}})'\,\boldsymbol{g}^\pi,\ \forall\, \pi \in \mathscr{S}_{\hat{\boldsymbol{w}}} & & -\sum\limits_{\pi\in\mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi \boldsymbol{g}^\pi + \boldsymbol{\eta} + \tilde{\boldsymbol{\mu}} \geq 0 \\
\boldsymbol{\eta} \to & \boldsymbol{w} \leq \mathbf{1} & & \sum\limits_{\pi\in\mathscr{S}_{\hat{\boldsymbol{w}}}} \lambda_\pi = 1 \\
& \boldsymbol{w} \geq 0 & & \\
\mu_{i,j} \to & w_j - w_i \leq 0 & & \lambda_\pi, \boldsymbol{\eta}, \mu_{i,j} \geq 0
\end{array}
\tag{EC.3}
$$

where $\tilde{\mu}_i \stackrel{\text{def}}{=} -\sum_{j\,:(i,j)\in E} \mu_{i,j} + \sum_{j\,:(j,i)\in E} \mu_{j,i}$. The primal and dual programs above have an optimal value $J^\star$, and, in any dual optimal solution, $\boldsymbol{\eta}^\star = \sum_\pi \lambda^\star_\pi \boldsymbol{g}^\pi - \tilde{\boldsymbol{\mu}}^\star$. Furthermore, by complementary slackness, there exists an optimal dual solution (corresponding to the primal optimal solution $J^\star$, $\hat{\boldsymbol{w}}$) satisfying $\eta_i^\star = 0, \forall i \in S^C$. This implies that $\boldsymbol{\eta}^\star$ satisfies $\boldsymbol{\eta}^\star_{S^C} = 0$, $\boldsymbol{\eta}^\star_S \geq 0$.

The candidate vector $\boldsymbol{g}$ we would like to consider is exactly $\boldsymbol{g} = \sum_\pi \lambda^\star_\pi \boldsymbol{g}^\pi = \boldsymbol{\eta}^\star + \tilde{\boldsymbol{\mu}}^\star$. To complete part **[2]**, we need to check that

$$(\boldsymbol{w} - \hat{\boldsymbol{w}})^T \boldsymbol{g} \leq 0, \forall \boldsymbol{w} \in P.$$

However, note that $\boldsymbol{w} - \hat{\boldsymbol{w}}$ can be written (for any $\boldsymbol{w} \in P$) as a conic combination of the vectors $\boldsymbol{w}_a - \hat{\boldsymbol{w}}$, where $\boldsymbol{w}_a$ are vertices of P adjacent8. to $\hat{\boldsymbol{w}}$. Therefore, the required condition holds at an arbitrary $\boldsymbol{w}$ if and only if it holds at all vertices of P adjacent to $\hat{\boldsymbol{w}}$.

To characterize the latter set, denoted by $\mathcal{A}(\hat{\boldsymbol{w}})$, we introduce the following sets of nodes:

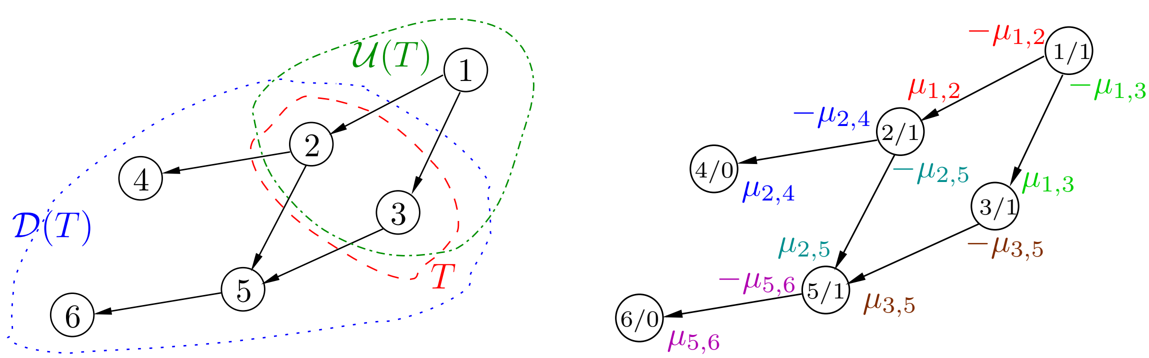
$$\mathcal{D}(T) \stackrel{\text{def}}{=} \{k \in \{1, \dots, n\} : \exists\, i \in T \text{ and a directed path in } G \text{ from } i \text{ to } k\} \tag{EC.4a}$$

$$\mathcal{U}(T) \stackrel{\text{def}}{=} \{k \in \{1, \dots, n\} : \exists\, i \in T \text{ and a directed path in } G \text{ from } k \text{ to } i\}. \tag{EC.4b}$$

$\mathcal{D}(T)$ contains all the nodes "in the downstream" of nodes $i \in T$ (by definition, we automatically include in $\mathcal{D}(T)$ the set T itself). In particular, in any feasible $\boldsymbol{w} \in P$, we have $w_k \leq w_i, \forall k \in \mathcal{D}(T), \forall i \in T$. Similarly, $\mathcal{U}(T)$ has all the nodes "in the upstream" of nodes $i \in T$, and any feasible $\boldsymbol{w} \in P$ satisfies $w_k \geq w_i, \forall k \in \mathcal{U}(T), \forall i \in T$. For an example, please refer to Figure EC.3.

With $\mathcal{D}(T)$ and $\mathcal{U}(T)$ as above, Lemma EC.3 in the Online Appendix provides the following inclusion relation for $\mathcal{A}(\hat{\boldsymbol{w}})$:

$$\mathcal{A}(\hat{\boldsymbol{w}}) \subseteq \left\{\hat{\boldsymbol{w}} - \mathbf{1}_{\mathcal{D}(T) \cap S} : T \subseteq S\right\} \cup \left\{\hat{\boldsymbol{w}} + \mathbf{1}_{\mathcal{U}(T) \cap S^C} : T \subseteq S^C\right\}.$$



(a) (b)

Figure EC.3 Example of a preorder graph with downstream and upstream nodes (a), and a vertex in the corresponding uncertainty set $\mathcal{W}$, with the dual variables (b). Here, $G = (V, E)$, where $V = \{1, \dots, 5\}$ and $E = \{(1,2), (1,3), (2,4), (2,5), (3,5), (5,6)\}$. In (a), $T = \{2,3\}$, so that $\mathcal{D}(T) = \{2,3,4,5,6\}$, and $\mathcal{U}(T) = \{1,2,3\}$. In (b), the relevant vertex in $\mathcal{W}$ is $\hat{\boldsymbol{w}} = (1,1,1,0,1,0)$, so that $S_y = \{1,2,3,5\}$, and $\mu_{24} = \mu_{56} = 0$.

To argue that $(\boldsymbol{w}-\hat{\boldsymbol{w}})^T\boldsymbol{g}\leq 0, \forall\, \boldsymbol{w}\in\mathcal{A}(\hat{\boldsymbol{w}})$, it suffices to check the relation for the larger set on the right. Consider the following separate cases.

[C1] $\boldsymbol{w}-\hat{\boldsymbol{w}}=-\mathbf{1}_{\mathcal{D}(T)\cap S}$, for some $T\subseteq S$. Then,

$$\begin{aligned}(\boldsymbol{w}-\hat{\boldsymbol{w}})^T\boldsymbol{g} &= (\boldsymbol{w}-\hat{\boldsymbol{w}})^T(\boldsymbol{\eta}^\star+\tilde{\boldsymbol{\mu}}^\star)\\ &= -\sum_{i\in\mathcal{D}(T)\cap S}\eta_i^\star+\sum_{i\in\mathcal{D}(T)\cap S}\Big(\sum_{j\,:(i,j)\in E}\mu_{i,j}^\star-\sum_{j\,:(j,i)\in E}\mu_{j,i}^\star\Big)\end{aligned}$$

Consider an arbitrary node i in the summation above. For any $j\in S^C$ such that $(i,j)\in E$, we must have $\mu_{i,j}^\star=0$, by complementary slackness. For any $j\in S$ such that $(i,j)\in E$, we must have $j\in\mathcal{D}(T)\cap S$. Therefore, the dual variable $\mu_{i,j}^\star$ appears in the expression above twice, once with a "+" sign (for the edge (i,j) going out of node i), and once with a "-" sign (for the edge (i,j) going into node j). Since the two terms cancel out, the final expression above contains only terms in $\eta_i^\star$ or $\mu_{i,j}^\star$ with negative signs, hence must be non-positive. To better understand the relation, please refer to Figure EC.3 for an example.

[C2] $\boldsymbol{w}-\hat{\boldsymbol{w}}=\mathbf{1}_{\mathcal{U}(T)\cap S^C}$, for some $T\subseteq S^C$. Then, the complementary slackness conditions at $\hat{\boldsymbol{w}}$ imply that $\eta_i=0, \forall\, i\in\mathcal{U}(T)\cap S^C$. We have

$$\begin{aligned}(\boldsymbol{w}-\hat{\boldsymbol{w}})^T\boldsymbol{g} &= (\boldsymbol{w}-\hat{\boldsymbol{w}})^T(\boldsymbol{\eta}^\star+\tilde{\boldsymbol{\mu}}^\star)\\ &= \sum_{i\in\mathcal{U}(T)\cap S^C}\Big(-\sum_{j\,:(i,j)\in E}\mu_{i,j}^\star+\sum_{j\,:(j,i)\in E}\mu_{j,i}^\star\Big).\end{aligned}$$

By a similar argument as above, consider an arbitrary i in the summation. For any $j\in S$ such that $(j,i)\in E$, we must have $\mu_{j,i}^\star=0$. For any $j\in S^C$ such that $(j,i)\in E$, we must have $j\in\mathcal{U}(T)\cap S^C$. Therefore, $\mu_{i,j}^\star$ again appears twice, once with a "+" sign (for the edge (j,i) going into i), and once with a "-" sign (for the edge (j,i) going out of j). Since the two terms cancel out, the final expression again contains only terms with negative signs. $\square$

The following lemma provides a (partial) characterization for the set of adjacent points in a sublattice polytope of the form (3).

Lemma EC.3. *Consider a polytope $P=\{\boldsymbol{w}\in\mathcal{H}_n : x_i\geq x_j, \forall\,(i,j)\in E\}$, where $E\subseteq\{1,\dots,n\}^2$ is any set of directed edges. Let $\boldsymbol{y}\equiv\mathbf{1}_{S_y}$ denote any vertex of P, where $S_y\subseteq\{1,\dots,n\}$. Then, all the vertices of P adjacent to $\boldsymbol{y}$ are contained in the set*

$$\big\{\boldsymbol{y}-\mathbf{1}_{\mathcal{D}(T)\cap S_y} : T\subseteq S_y\big\}\cup\big\{\boldsymbol{y}+\mathbf{1}_{\mathcal{U}(T)\cap S_y^c} : T\subseteq S_y^c\big\}, \tag{EC.5}$$

where $\mathcal{D}(T)$ and $\mathcal{U}(T)$ are given by:

$$\begin{aligned}\mathcal{D}(T) &\stackrel{\text{def}}{=} \{k\in\{1,\dots,n\} : \exists\; i\in T \text{ and a directed path in } G \text{ from } i \text{ to } k\}\\ \mathcal{U}(T) &\stackrel{\text{def}}{=} \{k\in\{1,\dots,n\} : \exists\; i\in T \text{ and a directed path in } G \text{ from } k \text{ to } i\}.\end{aligned}$$

Proof. Consider any vertex $\boldsymbol{x}$ adjacent to $\boldsymbol{y}$, and let $\boldsymbol{x} = \mathbf{1}_{S_x}$ for some $S_x \subseteq \{1,\dots,n\}$. We claim that $\boldsymbol{x} \le \boldsymbol{y}$ or $\boldsymbol{x} \ge \boldsymbol{y}$. Otherwise, since $\mathbf{1}_{S_x \cup S_y}$ and $\mathbf{1}_{S_x \cap S_y}$ are also valid vertices of P satisfying $\mathbf{1}_{S_x \cup S_y} + \mathbf{1}_{S_x \cap S_y} = \mathbf{1}_{S_x} + \mathbf{1}_{S_y}$, we would obtain two distinct convex representations for $(\boldsymbol{x}+\boldsymbol{y})/2$ in terms of vertices of P, which can never be the case if $\boldsymbol{x}$ and $\boldsymbol{y}$ are adjacent (see, e.g., Lemma 1 in Gurgel and Wakabayashi (1997)).

To complete the proof, we claim that

$$\big\{\boldsymbol{y} - \mathbf{1}_{\mathcal{D}(T)\cap S_y},\, T \subseteq S_y\big\} = \big\{\boldsymbol{x} \in \operatorname{ext}(P) : \boldsymbol{x} \le \boldsymbol{y}\big\}, \tag{EC.6a}$$

$$\big\{\boldsymbol{y} + \mathbf{1}_{\mathcal{U}(T)\cap S_y^c},\, T \subseteq S_y^c\big\} = \big\{\boldsymbol{x} \in \operatorname{ext}(P) : \boldsymbol{x} \ge \boldsymbol{y}\big\}. \tag{EC.6b}$$

We argue (EC.6a) by double inclusion, and (EC.6b) follows by an analogous argument.

Note that "$\subseteq$" follows trivially, since all the points in the set on the left of (EC.6a) are valid extreme points of P and are $\le \boldsymbol{y}$.

To argue "$\supseteq$", consider any $\boldsymbol{x} \le \boldsymbol{y}$ and note that $\boldsymbol{x} = \boldsymbol{y} - \mathbf{1}_{S_y \setminus S_x}$. By definition, $S_y \setminus S_x \subseteq \mathcal{D}(S_y \setminus S_x) \cap S_y$, and we claim the reverse inclusion also holds. To this end, note that S_x cannot contain any elements in $\mathcal{D}(S_y \setminus S_x)$, since the components corresponding to the latter indices are always set to zero when the components corresponding to $S_y \setminus S_x$ are set to zero. Therefore, $\mathcal{D}(S_y \setminus S_x) \cap S_y = S_y \setminus S_x$, so that $\boldsymbol{x} = \boldsymbol{y} - \mathbf{1}_{\mathcal{D}(S_y \setminus S_x)\cap S_y}$, which completes the reverse inclusion. $\square$

We note that the set of adjacent vertices in a binary sublattice does not seem to have a trivial characterization. In particular, it is easy to construct examples showing that the inclusion of the former set in the set in (EC.5) can be strict. For instance, when $P = \mathcal{H}_n$, and $\boldsymbol{y} = \mathbf{1}$, the set in (EC.5) is actually $\operatorname{ext}(P)$, i.e., *all* the extreme points of P.

A natural conjecture would be that the former set can be reached by changing *a single* coordinate i at a time, together with all the relevant corresponding coordinates in $\mathcal{U}(\{i\})$ or $\mathcal{D}(\{i\})$ (in other words, that we can restrict (EC.5) to sets T with $|T| = 1$). Unfortunately, this characterization turns out to be incomplete. To see this, consider the simple example in Figure 1, where $P = \{\boldsymbol{x} \in \mathbb{R}^3 : x_1 \ge x_2,\, x_1 \ge x_3\}$. Here, vertex $(0,0,0)$ is adjacent to *all* the vertices of P – in particular, $(1,1,1)$ – which cannot be reached by changing only one coordinate at a time.

The next result is a complete proof of the main theorem of Section 4 in the paper.

THEOREM 3. *Consider an optimization problem of the form*

$$\max_{\boldsymbol{w}\in P}\Big[\boldsymbol{a}^T\boldsymbol{w} + \sum_{i\in\mathcal{I}} h_i(\boldsymbol{w})\Big], \tag{EC.7}$$

where $P \subset \mathbb{R}^k$ is any polytope, $\boldsymbol{a} \in \mathbb{R}^n$ is an arbitrary vector, $\mathcal{I}$ is a finite index set, and $h_i : \mathbb{R}^n \to \mathbb{R}$ are functions satisfying the following properties

[P1] *h_i are concave extendable from* $\operatorname{ext}(P)$, $\forall\, i \in \mathcal{I}$,

[P2] $\text{conc}_P(h_i + h_j) = \text{conc}_P(h_i) + \text{conc}_P(h_j)$, *for any* $i \neq j \in \mathcal{I}$.

Then there exists a set of affine functions $z_i(\boldsymbol{w})$, $i \in \mathcal{I}$, *satisfying* $z_i(\boldsymbol{w}) \geq h_i(\boldsymbol{w}), \forall \boldsymbol{w} \in P, \forall i \in \mathcal{I}$, *such that*

$$\max_{\boldsymbol{w} \in P} \Big[\boldsymbol{a}^T \boldsymbol{w} + \sum_{i \in \mathcal{I}} z_i(\boldsymbol{w}) \Big] = \max_{\boldsymbol{w} \in P} \Big[\boldsymbol{a}^T \boldsymbol{w} + \sum_{i \in \mathcal{I}} h_i(\boldsymbol{w}) \Big]. \tag{EC.8}$$

Proof. We prove the result for a case with $|\mathcal{I}| = 2$. The general result follows by induction on $|\mathcal{I}|$, and by noting that properties **[P1]** and **[P2]** are preserved under addition of functions. Furthermore, to avoid technicalities, we consider the case when the optimal value in (EC.7), denoted by $J^\star$, is finite9. .

When $|\mathcal{I}| = 2$, note that the affine function $z(\boldsymbol{w}) = J^\star - \boldsymbol{a}^T \boldsymbol{w}$ trivially satisfies the constraints

$$z(\boldsymbol{w}) \geq h_1(\boldsymbol{w}) + h_2(\boldsymbol{w}), \forall \boldsymbol{w} \in P,$$
$$J^\star = \max_{\boldsymbol{w} \in P} \big[\boldsymbol{a}^T \boldsymbol{w} + z(\boldsymbol{w}) \big].$$

Therefore, to prove our claim, it suffices to find two affine functions $z_1(\boldsymbol{w})$ and $z_2(\boldsymbol{w})$, satisfying

$$z_1(\boldsymbol{w}) + z_2(\boldsymbol{w}) = z(\boldsymbol{w})$$
$$z_i(\boldsymbol{w}) \geq h_i(\boldsymbol{w}), \forall \boldsymbol{w} \in P, \forall i \in \mathcal{I}.$$

With $z_2 = z - z_1$, this is equivalent to finding a single affine function z_1 satisfying

$$h_1(\boldsymbol{w}) \leq z_1(\boldsymbol{w}) \leq z(\boldsymbol{w}) - h_2(\boldsymbol{w}), \forall \boldsymbol{w} \in P. \tag{EC.9}$$

To this end, let us consider the functions $f \stackrel{\text{def}}{=} h_1$ and $g \stackrel{\text{def}}{=} z - h_2$. By Property **[P1]**, since z is affine, both f and $-g$ are concave-extendable from $\text{ext}(P)$ (see Section EC.2 of the Online Appendix or Proposition 2 in Tardella (2008)). Also, $f \leq g$ on P. We claim that

$$\text{conc}_P(f)(\boldsymbol{w}) \leq \text{conv}_P(g)(\boldsymbol{w}), \forall \boldsymbol{w} \in P. \tag{EC.10}$$

To see this, consider the function $f - g$. Since both f and $-g$ are concave-extendable from $\text{ext}(P)$, so is $f - g$. By Property **[P2]**, we also have that $\text{conc}_P(f - g) = \text{conc}_P(f) + \text{conc}_P(-g) = \text{conc}_P(f) - \text{conv}_P(g)$. Therefore,

$$\begin{aligned} \max_{\boldsymbol{w} \in P} \big[\text{conc}_P(f)(\boldsymbol{w}) - \text{conv}_P(g)(\boldsymbol{w}) \big] &= \max_{\boldsymbol{w} \in P} \text{conc}_P(f - g)(\boldsymbol{w}) \\ &= \max_{\boldsymbol{w} \in P} (f - g)(\boldsymbol{w}) \\ &\leq 0. \end{aligned}$$

If the maximum is actually 0 in the above expression, then $\text{conc}_P(f) = \text{conv}_P(g)$, so that both are affine functions on P, and $z_1 = \text{conc}_P(f) = \text{conv}_P(g)$ would satisfy the requirement in (EC.9).

Therefore, we assume throughout that there exists $\boldsymbol{w} \in P : \mathrm{conc}_P(f)(w) < \mathrm{conv}_P(g)(\boldsymbol{w})$. We can now introduce the following two sets:

$$\mathcal{H}_f \equiv \mathrm{hypo}(\mathrm{conc}_P(f)) \stackrel{\text{def}}{=} \left\{(\boldsymbol{w}, t) \in \mathbb{R}^{k+1} : \boldsymbol{w} \in P, t \leq \mathrm{conc}_P(f)(\boldsymbol{w})\right\} \tag{EC.11a}$$

$$\mathcal{E}_g \equiv \mathrm{epi}(\mathrm{conv}_P(g)) \stackrel{\text{def}}{=} \left\{(\boldsymbol{w}, t) \in \mathbb{R}^{k+1} : \boldsymbol{w} \in P, t \geq \mathrm{conv}_P(g)(\boldsymbol{w})\right\} \tag{EC.11b}$$

Note that $\mathcal{H}_f$ and $\mathcal{E}_g$ represent the hypograph of $\mathrm{conc}_P(f)$ and the epigraph of $\mathrm{conv}_P(g)$, respectively. As such, they are convex, closed sets. Furthermore, both $\mathcal{H}_f$ and $\mathcal{E}_g$ are *polyhedral* sets, since the concave envelopes of concave-extendable functions are polyhedral (see Section EC.2.1 of the Online Appendix). With $\mathrm{ri}(K)$ denoting the relative interior of a convex set K, we claim that10.

$$\mathcal{H}_f \cap \mathrm{ri}(\mathcal{E}_g) = \emptyset.$$

This follows because any $(\boldsymbol{w}, t) \in \mathcal{H}_f \cap \mathrm{ri}(\mathcal{E}_g)$ would satisfy $\mathrm{conv}_P(g)(\boldsymbol{w}) < t \leq \mathrm{conc}_P(f)(\boldsymbol{w})$, in direct contradiction with (EC.10). Therefore, we have two polyhedral sets, $\mathcal{H}_f$ and $\mathcal{E}_g$, such that $\mathcal{H}_f \cap \mathrm{ri}(\mathcal{E}_g) = \emptyset$. By Theorem 20.2 in Rockafellar (1970), there exists a hyperplane separating $\mathcal{H}_f$ and $\mathcal{E}_g$ *properly* (i.e., not both sets belonging to the hyperplane). In particular, there exist $\boldsymbol{z}_1 \in \mathbb{R}^k$, $z_{1,0}$, $\beta \in \mathbb{R}$ such that $(\boldsymbol{z}_1, z_{1,0}) \neq \boldsymbol{0}$, and

$$\forall\, \boldsymbol{w} \in P,\ \boldsymbol{z}_1' \boldsymbol{w} + z_{1,0}\, f(\boldsymbol{w}) \leq \beta \leq \boldsymbol{z}_1' \boldsymbol{w} + z_{1,0}\, g(\boldsymbol{w}),$$

By proper separability, $z_{1,0} \neq 0$. If $z_{1,0} < 0$, then $f \geq g$ on P, which would contradict our standing assumption that $\exists\, \boldsymbol{w} \in P : \mathrm{conc}_P(f)(\boldsymbol{w}) < \mathrm{conv}_P(g)(\boldsymbol{w})$. Therefore, we are left with $z_{1,0} > 0$, which implies that $z_1(\boldsymbol{w}) \stackrel{\text{def}}{=} (\beta - \boldsymbol{z}_1'\boldsymbol{w})/z_{1,0}$ satisfies equation (EC.9), and hence completes the construction and the proof. $\square$

An example outlining the role of requirement **[P2]** is presented in Figure EC.4.

Endnotes

5. Tardella deals with convex envelopes, and his definitions are in terms of edge-concave functions. All of his results can be ported here by essentially switching *convex* with *concave*.

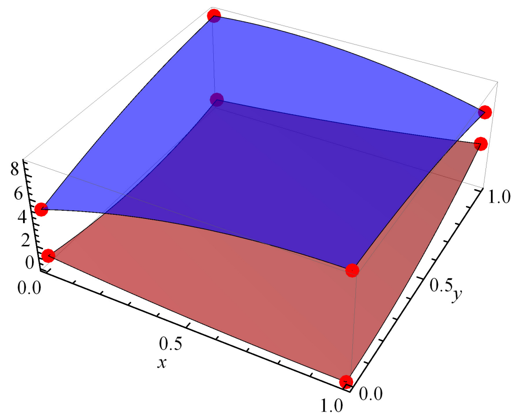
6. For a polytope P, a set of n-dimensional polyhedra $P_1, \ldots, P_m \subseteq P$ is said to be a *polyhedral subdivision* of P if $P = \cup_{i=1}^m P_i$ and $P_i \cap P_j$ is a (possibly empty) face of both P_i and P_j.

7. In other words, any such permutation π has in the first $|S|$ positions the elements $\{i : w_i = 1\}$, and in the remaining $|S^C|$ the elements $\{i : w_i = 0\}$.

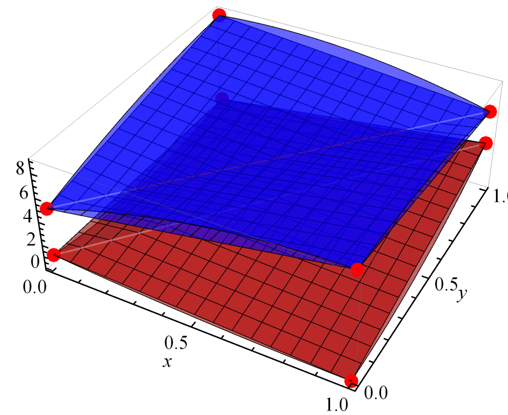
8. The notion of adjacency used here is well established in polyhedral theory - two vertices of a polytope are said to be adjacent if there is an edge (i.e., a face of dimension 1) connecting them. We refer the interested reader to Schrijver (2000) for definitions and details.

9. The arguments are extendable to a case when $J^\star = +\infty$, by allowing extended-real convex functions h_i to be used.

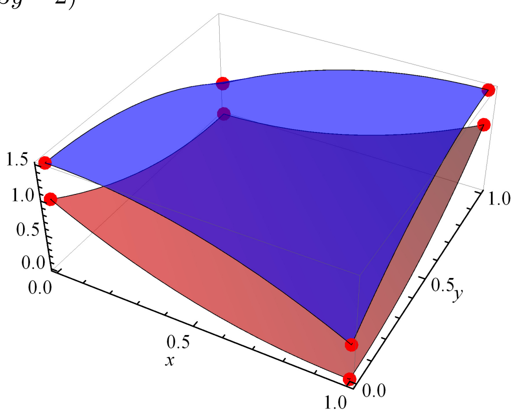
10.We can equivalently prove that $\mathcal{E}_g \cap \mathrm{ri}(\mathcal{H}_f) = \emptyset$.



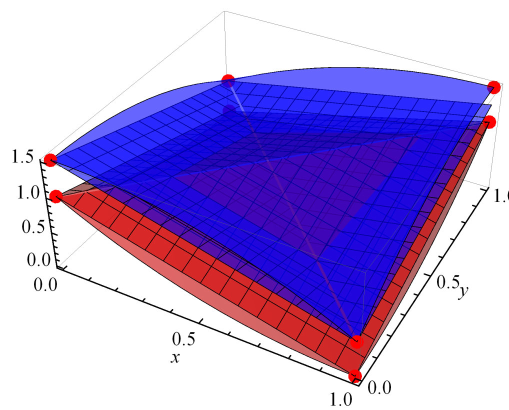
(a) $f(x,y) \stackrel{\text{def}}{=} (x+2y-1)^2 \le g(x,y) \stackrel{\text{def}}{=} 9-(2x+1.5y-2)^2$



(b) f and $\mathrm{conc}_P(f)$; g and $\mathrm{conv}_P(g)$



(c) $f(x,y) \stackrel{\text{def}}{=} (x+y-1)^2 \le g(x,y) \stackrel{\text{def}}{=} 1.5-(x-y)^2$



(d) f and $\mathrm{conc}_P(f)$; g and $\mathrm{conv}_P(g)$

Figure EC.4 The role of requirement **[P2]**. In both Figure (a) and Figure (c), h_1 and h_2 are convex (cvx). In Figure (a), h_1 and h_2 are also supermodular (spm), so that f and $-g$ are cvx, spm, $\mathrm{conc}_P(f-g) = \mathrm{conc}_P(f) - \mathrm{conc}_P(g)$, and $\mathrm{conc}_P(f) \le \mathrm{conv}_P(g)$ in Figure (b). In Figure (c), h_1 is spm, but h_2 is not, so that $-g$ is not spm, and $\mathrm{conc}_P(f-g) \le \mathrm{conc}_P(f) - \mathrm{conc}_P(g)$. Note that $\mathrm{conc}_P(f) \not\le \mathrm{conv}_P(g)$ in Figure (d).

References

Bensoussan, A., M. Crourty, J. M. Proth. 1983. *Mathematical Theory of Production Planning*. Elsevier Science Ltd.

Choquet, G. 1954. Theory of capacities. *Annales de l'Institut Fourier* **5** 131–295.

Fujishige, S. 2005. *Submodular Functions and Optimization*, *Annals of Discrete Mathematics*, vol. 58. 2nd ed. Elsevier.

Grötschel, M., L. Lovácz, A. Schrijver. 1988. *Geometric Algorithms and Combinatorial Optimization*. Springer Verlag, Berlin, New York.

Gurgel, M. A., Y. Wakabayashi. 1997. Adjacency of vertices of the complete pre-order polytope. *Discrete Mathematics* **175**(1-3) 163 – 172.

Heyman, D. P., M. J. Sobel. 1984. *Stochastic Models in Operations Research*. Dover Publications, Inc.

Lovász, L. 1982. Submodular functions and convexity. *Mathematical Programming - The State of the Art*. Springer, 235–257.

Milgrom, P., C. Shannon. 1994. Monotone comparative statics. *Econometrica* **62**(1) pp. 157–180.

Pardalos, P. M., J. B. Rosen. 1986. Methods for global concave minimization: A bibliographic survey. *SIAM Review* **28**(3) pp. 367–379.

Rockafellar, T. 1970. *Convex Analysis*. Princeton University Press.

Schmeidler, D. 1986. Integral representation without additivity. *Proceedings of the Americal Mathematical Society* **97** 255–261.

Schrijver, A. 2000. *Theory of Linear and Integer Programming*. 2nd ed. John Wiley & Sons.

Schrijver, A. 2003. *Combinatorial Optimization: Polyhedra and Efficiency*. 1st ed. Springer.

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

Tawarmalani, M., J.P. Richard, C. Xiong. 2010. Explicit convex and concave envelopes through polyhedral subdivisions. Submitted for publication.

Topkis, D. M. 1998. *Supermodularity and Complementarity*. Princeton University Press.

Veinott, Jr., A. F. 1966. The status of mathematical inventory theory. *Management Science* **12**(11) 745–777.

Veinott, Jr., A. F. 1989. Lattice programming. Unpublished notes from lectures delivered at Johns Hopkins University.