ORACLE TRACTABILITY OF SKEW BISUBMODULAR FUNCTIONS*

ANNA HUBER† AND ANDREI KROKHIN‡

Abstract. In this paper we consider skew bisubmodular functions as recently introduced by the authors and Powell. We construct a convex extension of a skew bisubmodular function which we call Lovász extension in correspondence to the submodular case. We use this extension to show that skew bisubmodular functions given by an oracle can be minimised in polynomial time.

Key words. submodular functions, optimization, computational complexity

AMS subject classifications. 68Q25, 68Q17, 90C60, 90C27

1. Introduction.

A key task in combinatorial optimisation is the minimisation of discrete functions. Important examples are *submodular* functions, see e. g. [6, 14, 15, 19], and *bisubmodular* functions, see e. g. [2, 6, 15, 18]. A finitary function on a set D is any function with domain D^n where $n \in \mathbb{N}$, the number n is called the *arity* of the function. Submodular and bisumbodular functions can be viewed as real-valued finitary functions on D where D is a 2-element set for the submodular case and a 3-element set for the bisubmodular case. Fix a finite set D. One says that a class $\mathcal{C}$ of rational-valued finitary functions on D is *oracle-tractable* if there is an algorithm which, given a function $f \in C$ represented by a value-giving oracle, finds a minimiser of f in time polynomial in the arity of f. The oracle tractability of submodular and bisubmodular functions has been shown in [8, 14] and [18] respectively, with many subsequent improvements (see e. g. [15]). Results about oracle tractability for other classes of discrete functions can be found in [12, 13].

Submodular and bisubmodular functions play an important role for classifying the complexity of optimisation problems known as *valued constraint satisfaction problems* (VCSPs). These problems amount to minimising finitary functions on D represented as sums of bounded-arity functions. In the general-valued VCSP, such functions can also take infinite values, but we consider only the finite-valued case here. In this case, the complexity of VCSPs is now well understood [9, 10, 16, 17]. In particular, submodularity characterises tractable VCSPs on a two-element domain D [4]. In [9, 10] a generalisation of bisubmodularity, *skew bisubmodularity*, is introduced and used to classify the complexity of VCSPs on a three-element domain D: the tractable cases correspond to submodularity and skew bisubmodularity. The tractability of skew bisubmodular function minimisation in the VCSP setting (i.e. represented as sums of bounded-arity skew bisubmodular functions) follows from [16], but the question whether skew bisubmodular functions are also tractable in the oracle model has been left open in [9]. In this paper we construct a convex extension of a skew bisubmodular function, called Lovász extension in correspondence to the submodular case [14], and show the oracle tractability of skew bisubmodular functions.

Very closely related results have recently appeared in [7], where the authors acknowledge this work. They generalise the notion of skew bisubmodular function by

*This work was supported by the UK EPSRC grants EP/H000666/1 and EP/J000078/1.

†Department of Computing and Mathematics, University of Derby, Kedleston Road, Derby DE22 1GB, UK(a.huber@derby.ac.uk).

‡School of Engineering and Computing Sciences, Durham University, Durham DH1 3LE, UK (andrei.krokhin@durham.ac.uk).

allowing each variable in a function to have its own degree of skewness. They also describe a Lovász extension for such functions which leads to an efficient minimisation algorithm, study corresponding polyhedra, and prove a min-max theorem. The problem of finding a combinatorial algorithm for minimising skew bisubmodular functions is left open, both in our work and in [7].

2. Definitions and Main Result. Skew bisubmodularity, also known as *α-bisubmodularity*, is defined for a fixed number $\alpha \in (0,1]$ and functions $f : D^n \to \mathbb{R}$ where $|D| = 3$ and $n \in \mathbb{N}$. In [9, 10], the elements of D are denoted by $-1, 0, 1$. In this paper, we will fix $\alpha \in (0,1]$ throughout and, for convenience of notation, denote the elements of D by $-\alpha, 0, 1$, replacing the name -1 by $-\alpha$. Obviously, there is a direct correspondence between functions over $\{-1,0,1\}$ and functions over $\{-\alpha,0,1\}$. The definition of α-bisubmodularity as in [9, 10] is then as follows. Let $n \in \mathbb{N}$. We write $[n] := \{1,\ldots,n\}$.

Define the order $\prec$ on D through $0 \prec 1$, $0 \prec -\alpha$ and 1 and $-\alpha$ being incomparable. We also denote the corresponding component-wise order on D^n by $\prec$.

Define the binary operation $\wedge_0$ on D as follows.

$1 \wedge_0 -\alpha = -\alpha \wedge_0 1 = 0;$
$x \wedge_0 y = \min(x,y)$ with respect to the above order if $\{x,y\} \neq \{-\alpha,1\}$.

For $a \in D$, define the binary operation $\vee_a$ as follows:

$1 \vee_a -\alpha = -\alpha \vee_a 1 = a;$
$x \vee_a y = \max(x,y)$ with respect to the above order if $\{x,y\} \neq \{-\alpha,1\}$.

We also denote the corresponding component-wise operations on D^n by $\wedge_0$ and $\vee_a$ respectively.

Definition 1. *A function $f : D^n \to \mathbb{R}$ is called* α-bisubmodular *if, for all $\mathbf{a},\mathbf{b} \in D^n$,*

$$f(\mathbf{a} \wedge_0 \mathbf{b}) + \alpha \cdot f(\mathbf{a} \vee_0 \mathbf{b}) + (1-\alpha) \cdot f(\mathbf{a} \vee_1 \mathbf{b}) \leq f(\mathbf{a}) + f(\mathbf{b}). \tag{1}$$

The above inequality defines submodular functions if we restrict D to $\{0,1\}$ and it defines bisubmodular functions if $\alpha = 1$.

The following is the main result of this paper.

Theorem 2. *There exists an algorithm that finds a minimum of a given α-bisubmodular function $f : D^n \to \mathbb{Q}$ in time polynomial in n if f is given by an oracle.*

Proof. In the remainder of the paper we will construct for any α-bisubmodular function $f : D^n \to \mathbb{Q}$ a convex extension $f^L : [-\alpha,1]^n \to \mathbb{R}$ which takes its minimal value on D^n and which can be efficiently computed on every rational vector in $[-\alpha,1]^n$. The theorem then follows from convex optimisation techniques, in the same way that sub- and bisubmodular minimisation are achieved through convex optimisation, see [14] and [18] respectively. □

3. Lovász Extension for Skew Bisubmodular Functions. For $\mathbf{x} \in [-\alpha,1]^n$, let $\mathcal{P}(\mathbf{x})$ be the set of all probability distributions on D^n with marginals $\mathbf{x}$, i. e.

$$\mathcal{P}(\mathbf{x}) := \left\{ \lambda : D^n \to [0,1] \mid \sum_{\mathbf{a}\in D^n} \lambda(\mathbf{a}) = 1, \sum_{\mathbf{a}\in D^n} \lambda(\mathbf{a})\mathbf{a} = \mathbf{x} \right\}.$$

Definition 3 (Lovász Extension). *For a function $f : D^n \to \mathbb{R}$, define the* Lovász extension $f^L : [-\alpha, 1]^n \to \mathbb{R}$ *through*

$$f^L(\mathbf{x}) := \sum_{\mathbf{a} \in D^n} \lambda_{\mathbf{x}}(\mathbf{a}) f(\mathbf{a}),$$

where $\lambda_{\mathbf{x}}$ is the unique element of $\mathcal{P}(\mathbf{x})$ such that its support forms a chain in D^n with respect to the order $\prec$. (The existence and the uniqueness of this element are proved below in Lemma 4).

Note that, for any $\mathbf{a} \in D^n$, we have $\lambda_{\mathbf{a}}(\mathbf{a}) = 1$ and thus $f^L(\mathbf{a}) = f(\mathbf{a})$, i.e. f^L is indeed an extension of f. It also follows directly from the definition that

$$\min \left\{ f(\mathbf{a}) \;\middle|\; \mathbf{a} \in D^n \right\} = \min \left\{ f^L(\mathbf{x}) \;\middle|\; \mathbf{x} \in [-\alpha, 1]^n \right\}.$$

The restriction of f^L to $[0,1]^n$ is the ordinary Lovász extension for $f|_{\{0,1\}^n}$, as in [14]. In the case $\alpha = 1$, the function f^L is the Lovász extension for bisubmodular functions as in [18].

Lemma 4. *For every $\mathbf{x} \in [-\alpha, 1]^n$, there is a unique element $\lambda_{\mathbf{x}}$ of $\mathcal{P}(\mathbf{x})$ such that its support forms a chain in D^n with respect to the order $\prec$.*

Proof. Let $\mathbf{x} \in [-\alpha, 1]^n$ and write $\mathbf{x} = (x_1, \ldots, x_n)$.

Construction: We will construct an element $\lambda_{\mathbf{x}} \in \mathbb{R}^{D^n}$ and show that it has the required properties. For this, we will recursively construct two sequences, $(\mathbf{u_i})_{i \in \mathbb{N}}$ in D^n and $(\mathbf{x_i})_{i \in \mathbb{N}}$ in $[-\alpha, 1]^n$. For every $i \in \mathbb{N}$ we write $\mathbf{u_i} = (u_{i1}, \ldots, u_{in})$ and $\mathbf{x_i} = (x_{i1}, \ldots, x_{in})$.

Let $\mathbf{x_1} := \mathbf{x}$. Assuming that $\mathbf{x_i}$ is already constructed for some $i \in \mathbb{N}$, we will construct $\mathbf{u_i}$ and $\mathbf{x_{i+1}}$ as follows.

Let N_i, Z_i, and P_i denote the subsets of $[n]$ consisting of all $j \in [n]$ such that $x_{ij} < 0$, $x_{ij} = 0$, and $x_{ij} > 0$, respectively. Define

$$u_{ij} := \begin{cases} -\alpha & \text{for } \; j \in N_i \\ 0 & \text{for } \; j \in Z_i \\ 1 & \text{for } \; j \in P_i, \end{cases}$$

$$\lambda_{\mathbf{x}}(\mathbf{u_i}) := \begin{cases} \min \left\{ \min \left\{ -\frac{x_{ij}}{\alpha} \;\middle|\; j \in N_i \right\}, \min \left\{ x_{ij} \;\middle|\; j \in P_i \right\} \right\} & \text{if } \; \mathbf{u_i} \neq \mathbf{0} \\ 1 - \lambda_{\mathbf{x}}(\mathbf{u_1}) - \cdots - \lambda_{\mathbf{x}}(\mathbf{u_{i-1}}) & \text{if } \; \mathbf{u_i} = \mathbf{0} \end{cases}$$

and let

$$\mathbf{x_{i+1}} := \mathbf{x_i} - \lambda_{\mathbf{x}}(\mathbf{u_i}) \mathbf{u_i}. \tag{2}$$

From this construction we have for every $j \in [n]$ that

$$\begin{array}{llll} u_{ij} = 0 & & \Rightarrow x_{i+1,j} = 0 & \Rightarrow u_{i+1,j} = 0 \\ u_{ij} = 1 & \Rightarrow \lambda_{\mathbf{x}}(\mathbf{u_i}) \le x_{ij} & \Rightarrow x_{i+1,j} \ge 0 & \Rightarrow u_{i+1,j} \in \{0, 1\} \\ u_{ij} = -\alpha & \Rightarrow \lambda_{\mathbf{x}}(\mathbf{u_i}) \le -\frac{x_{ij}}{\alpha} & \Rightarrow x_{i+1,j} \le 0 & \Rightarrow u_{i+1,j} \in \{0, -\alpha\}, \end{array}$$

so $u_{i+1,j} \preceq u_{ij}$ and thus $\mathbf{u_{i+1}} \preceq \mathbf{u_i}$. Furthermore, if $\mathbf{u_i} \neq \mathbf{0}$ and $m \in [n]$ is such that

$$\text{either } m \in N_i \;\text{ and }\; -\frac{x_{im}}{\alpha} = \min \left\{ -\frac{x_{ij}}{\alpha} \;\middle|\; j \in N_i \right\} = \lambda_{\mathbf{x}}(\mathbf{u_i})$$

$$\text{or}\quad m \in P_i \ \text{ and } \ x_{im} = \min\left\{x_{ij} \;\middle|\; j \in P_i\right\} = \lambda_{\mathbf{x}}(\mathbf{u_i}),$$

then $x_{i+1,m} = 0$ and thus $u_{i+1,m} = 0$, whereas $u_{im} \neq 0$. Thus $\mathbf{u_{i+1}} \prec \mathbf{u_i}$.

Clearly, this recursive construction yields $\mathbf{u_{n+1}} = 0$. Let $k \in \mathbb{N}$ be such that $\mathbf{u_{k-1}} \neq \mathbf{0}$ and $\mathbf{u_k} = \mathbf{0}$ and let $\lambda_{\mathbf{x}}(\mathbf{v}) := 0$ for all $\mathbf{v} \in D^n \setminus \{\mathbf{u_1}, \dots, \mathbf{u_k}\}$. The construction yields that the support of $\lambda_{\mathbf{x}}$ forms a chain in D^n with respect to the order $\prec$. We will now prove that $\lambda_{\mathbf{x}} \in \mathcal{P}(\mathbf{x})$.

The choice of k yields $\lambda_{\mathbf{x}}(\mathbf{u_1}), \dots, \lambda_{\mathbf{x}}(\mathbf{u_{k-1}}) \neq 0$. Equation (2) yields

$$\sum_{i=1}^{k-1} \lambda_{\mathbf{x}}(\mathbf{u_i})\mathbf{u_i} = \mathbf{x}. \tag{3}$$

Let $j \in [n]$ be such that $u_{k-1,j} \neq 0$. As $\mathbf{0} \prec \mathbf{u_{k-1}} \prec \ldots \prec \mathbf{u_1}$, one has $u_{k-1,j} = \cdots = u_{1j}$ and thus

$$\sum_{i=1}^{k-1} \lambda_{\mathbf{x}}(\mathbf{u_i})u_{ij} = x_j$$

from (3) yields

$$\sum_{i=1}^{k-1} \lambda_{\mathbf{x}}(\mathbf{u_i}) = \frac{x_j}{u_{1j}} \leq 1.$$

If

$$\sum_{i=1}^{k-1} \lambda_{\mathbf{x}}(\mathbf{u_i}) = 1,$$

then $\lambda_{\mathbf{x}}(\mathbf{u_k}) = 0$ by definition and $\lambda_{\mathbf{x}}$ is supported by the chain $\{\mathbf{u_1}, \dots, \mathbf{u_{k-1}}\}$. If

$$\sum_{i=1}^{k-1} \lambda_{\mathbf{x}}(\mathbf{u_i}) < 1,$$

then $\lambda_{\mathbf{x}}(\mathbf{u_k}) > 0$ by definition and $\lambda_{\mathbf{x}}$ is supported by the chain $\{\mathbf{u_1}, \dots, \mathbf{u_k}\}$. One has

$$\sum_{\mathbf{a} \in D^n} \lambda_{\mathbf{x}}(\mathbf{a}) = \sum_{i=1}^{k} \lambda_{\mathbf{x}}(\mathbf{u_i}) = 1$$

by definition and

$$\sum_{\mathbf{a} \in D^n} \lambda_{\mathbf{x}}(\mathbf{a})\mathbf{a} = \sum_{i=1}^{k} \lambda_{\mathbf{x}}(\mathbf{u_i})\mathbf{u_i} = \sum_{i=1}^{k-1} \lambda_{\mathbf{x}}(\mathbf{u_i})\mathbf{u_i} \overset{(3)}{=} \mathbf{x},$$

so $\lambda_{\mathbf{x}} \in \mathcal{P}(\mathbf{x})$.

Uniqueness: Let $(\mathbf{u_i})_{i \in \mathbb{N}}$, $(\mathbf{x_i})_{i \in \mathbb{N}}$ and $\lambda_{\mathbf{x}}$ be as constructed above, let $\mathbf{v_1} \succ \ldots \succ \mathbf{v_\ell}$ be a chain in D^n and let $\mu \in \mathcal{P}(\mathbf{x})$ have support $\{\mathbf{v_1}, \dots, \mathbf{v_\ell}\}$. We will show that $\mu = \lambda_{\mathbf{x}}$. We have

$$\sum_{i=1}^{\ell} \mu(\mathbf{v_i})\mathbf{v_i} = \mathbf{x}. \tag{4}$$

Let $j \in [n]$. As $\mathbf{v_1} \succ \ldots \succ \mathbf{v}_\ell$, unless $v_{1j} = 0$, there is a $h \in [\ell]$ such that $v_{1j} = \cdots = v_{hj} \neq 0$ and either $h = \ell$ or $v_{hj} \succ v_{h+1,j} = \cdots = v_{\ell j} = 0$. If $v_{1j} = 0$, Equation (4) yields $x_{1j} = 0$ and thus $u_{1j} = 0$ by definition of u_{1j}. Otherwise, we have

$$v_{1j} \sum_{i=1}^{h} \mu(\mathbf{v_i}) = \sum_{i=1}^{h} \mu(\mathbf{v_i}) v_{ij} = \sum_{i=1}^{\ell} \mu(\mathbf{v_i}) v_{ij} \overset{(4)}{=} x_j. \tag{5}$$

As $\sum_{i=1}^{h} \mu(\mathbf{v_i}) > 0$, the numbers v_{1j}, u_{1j} and x_j all have the same sign. Since $v_{1j}, u_{1j} \in \{-\alpha, 0, 1\}$, it must hold that $v_{1j} = u_{1j}$. This yields $\mathbf{v_1} = \mathbf{u_1}$.

If $\ell = 1$, we are done, as μ and $\lambda_\mathbf{x}$ both take the value 1 on $\mathbf{v_1} = \mathbf{u_1}$ and 0 otherwise, so $\mu = \lambda_\mathbf{x}$. If $\ell > 1$, let $m \in [\ell - 1]$ be such that $\mathbf{v_h} = \mathbf{u_h}$ holds for all $h \leq m$ and $\mu(\mathbf{v_h}) = \lambda_\mathbf{x}(\mathbf{u_h})$ holds for all $h < m$. We will show that $\mu(\mathbf{v_m}) = \lambda_\mathbf{x}(\mathbf{u_m})$ and $\mathbf{v_{m+1}} = \mathbf{u_{m+1}}$.

As $\mathbf{v_m} \succ \mathbf{v_{m+1}}$ there is a $j \in [n]$ such that $v_{m+1,j} = 0$ but $v_{mj} \neq 0$.

As $\mathbf{v_1} \succ \ldots \succ \mathbf{v}_\ell$, one has $v_{1j} = \cdots = v_{mj} \succ v_{m+1,j} = \cdots = v_{\ell j} = 0$, and thus

$$\begin{aligned} \mu(\mathbf{v_m}) v_{mj} &= \sum_{i=1}^{m} \mu(\mathbf{v_i}) v_{ij} - \sum_{i=1}^{m-1} \mu(\mathbf{v_i}) v_{ij} \\ &= \sum_{i=1}^{\ell} \mu(\mathbf{v_i}) v_{ij} - \sum_{i=1}^{m-1} \lambda_\mathbf{x}(\mathbf{u_i}) u_{ij} \\ &\overset{(4),(2)}{=} x_j - (x_j - x_{mj}) \\ &= x_{mj} \end{aligned}$$

So if $v_{mj} = 1$ we must have $\mu(\mathbf{v_m}) = x_{mj}$ and if $v_{mj} = -\alpha$ we must have $\mu(\mathbf{v_m}) = -\frac{x_{mj}}{\alpha}$.

If $\mu(\mathbf{v_m}) \neq \min\left\{\min\left\{-\frac{x_{mp}}{\alpha} \;\middle|\; p \in N_i\right\}, \min\left\{x_{mp} \;\middle|\; p \in P_i\right\}\right\} = \lambda_\mathbf{x}(\mathbf{u_m})$ we get a contradiction to (4) as then $\mu(\mathbf{v_m}) > \lambda_\mathbf{x}(\mathbf{u_m})$, and so, for $j' \in [n]$ such that $u_{(m+1)j'} = 0$ but $u_{mj'} \neq 0$ we get the following. As $\mathbf{u_1} \succ \ldots \succ \mathbf{u_k}$, one has $u_{1j'} = \cdots = u_{mj'} \succ u_{m+1,j'} = \cdots = u_{kj'} = 0$.

If $u_{mj'} = 1$, then $v_{1j'} = \cdots = v_{mj'} = u_{1j'} = \cdots = u_{mj'} = 1$ and $v_{(m+1)j'}, \ldots, v_{\ell j'} \in \{0, 1\}$, and so we have

$$\begin{aligned} \sum_{i=1}^{\ell} \mu(\mathbf{v_i}) v_{ij'} \geq \sum_{i=1}^{m} \mu(\mathbf{v_i}) v_{ij'} = \sum_{i=1}^{m} \mu(\mathbf{v_i}) \\ > \sum_{i=1}^{m} \lambda_\mathbf{x}(\mathbf{u_i}) = \sum_{i=1}^{m} \lambda_\mathbf{x}(\mathbf{u_i}) u_{ij'} = \sum_{i=1}^{k} \lambda_\mathbf{x}(\mathbf{u_i}) u_{ij'} = x_{j'}, \end{aligned}$$

contradiction to (4).

Equally, if $u_{mj'} = -\alpha$, we have $v_{1j'} = \cdots = v_{mj'} = u_{1j'} = \cdots = u_{mj'} = -\alpha$ and $v_{(m+1)j'}, \ldots, v_{\ell j'} \in \{0, -\alpha\}$, and so

$$\begin{aligned} \sum_{i=1}^{\ell} \mu(\mathbf{v_i}) v_{ij'} \leq \sum_{i=1}^{m} \mu(\mathbf{v_i}) v_{ij'} = -\alpha \sum_{i=1}^{m} \mu(\mathbf{v_i}) \\ < -\alpha \sum_{i=1}^{m} \lambda_\mathbf{x}(\mathbf{u_i}) = \sum_{i=1}^{m} \lambda_\mathbf{x}(\mathbf{u_i}) u_{ij'} = \sum_{i=1}^{k} \lambda_\mathbf{x}(\mathbf{u_i}) u_{ij'} = x_{j'}, \end{aligned}$$

contradiction to (4). We thus have $\mu(\mathbf{v_m}) = \lambda_{\mathbf{x}}(\mathbf{u_m})$. The fact that $\mathbf{v_h} = \mathbf{u_h}$ and $\mu(\mathbf{v_h}) = \lambda_{\mathbf{x}}(\mathbf{u_h})$ holds for all $h \leq m$ implies $\mathbf{v_{m+1}} = \mathbf{u_{m+1}}$ by a similar argument as used to show $\mathbf{v_1} = \mathbf{u_1}$ in (5). This finishes the inductive proof that $\mathbf{v_h} = \mathbf{u_h}$ for all $h \in [\ell]$ and that $\mu = \lambda_{\mathbf{x}}$. ∎

3.1. Convex Closure. As, for every $\mathbf{x} \in [-\alpha, 1]^n$, the set $\mathcal{P}(\mathbf{x})$ is a compact and non-empty subset of $\mathbb{R}^{D^n}$, the set

$$\left\{ \sum_{\mathbf{a} \in D^n} \lambda(\mathbf{a}) f(\mathbf{a}) \;\middle|\; \lambda \in \mathcal{P}(\mathbf{x}) \right\}$$

is a compact and non-empty subset of $\mathbb{R}$, and so contains its infimum.

Definition 5 (Convex Closure). *For a function $f : D^n \to \mathbb{R}$, its* convex closure *$f^- : [-\alpha, 1]^n \to \mathbb{R}$ is defined by*

$$f^-(\mathbf{x}) := \min \left\{ \sum_{\mathbf{a} \in D^n} \lambda(\mathbf{a}) f(\mathbf{a}) \;\middle|\; \lambda \in \mathcal{P}(\mathbf{x}) \right\}.$$

Proposition 6. *f^- is convex.*

Proof. Let $\beta \in (0, 1)$ and $\mathbf{x}, \mathbf{y} \in [-\alpha, 1]^n$. Let $\mu \in \mathcal{P}(\mathbf{x})$ be such that

$$f^-(\mathbf{x}) = \sum_{\mathbf{a} \in D^n} \mu(\mathbf{a}) f(\mathbf{a})$$

and let $\nu \in \mathcal{P}(\mathbf{y})$ be such that

$$f^-(\mathbf{y}) = \sum_{\mathbf{a} \in D^n} \nu(\mathbf{a}) f(\mathbf{a}).$$

Then $\beta\mu + (1 - \beta)\nu \in \mathcal{P}(\beta\mathbf{x} + (1 - \beta)\mathbf{y})$, and so

$$\begin{aligned} f^-(\beta\mathbf{x} + (1-\beta)\mathbf{y}) &= \min \left\{ \sum_{\mathbf{a} \in D^n} \lambda(\mathbf{a}) f(\mathbf{a}) \;\middle|\; \lambda \in \mathcal{P}(\beta\mathbf{x} + (1-\beta)\mathbf{y}) \right\} \\ &\leq \sum_{\mathbf{a} \in D^n} (\beta\mu + (1-\beta)\nu)(\mathbf{a}) f(\mathbf{a}) \\ &= \beta \sum_{\mathbf{a} \in D^n} \mu(\mathbf{a}) f(\mathbf{a}) + (1-\beta) \sum_{\mathbf{a} \in D^n} \nu(\mathbf{a}) f(\mathbf{a}) \\ &= \beta f^-(\mathbf{x}) + (1-\beta) f^-(\mathbf{y}). \end{aligned}$$

∎

3.2. Convexity of the Lovász Extension. The following lemma generalises the corresponding results for submodular and bisubmodular functions, see [14] and [18].

Lemma 7. *The Lovász extension f^L is convex if and only if f is α-bisubmodular.*

Proof. Let $\mathbf{a}, \mathbf{b} \in D^n$. If f^L is convex, it holds that

$$f^L\left(\tfrac{\mathbf{a}+\mathbf{b}}{2}\right) \leq \tfrac{f^L(\mathbf{a}) + f^L(\mathbf{b})}{2} = \tfrac{f(\mathbf{a}) + f(\mathbf{b})}{2}. \quad (6)$$

It is easy to check that

$$(\mathbf{a} \wedge_0 \mathbf{b}) + \alpha(\mathbf{a} \vee_0 \mathbf{b}) + (1 - \alpha)(\mathbf{a} \vee_1 \mathbf{b}) = \mathbf{a} + \mathbf{b}, \quad (7)$$

and so the probability distribution λ with $\lambda(\mathbf{a}\wedge_0\mathbf{b}) = \frac{1}{2}$, $\lambda(\mathbf{a}\vee_0\mathbf{b}) = \frac{\alpha}{2}$ and $\lambda(\mathbf{a}\vee_1\mathbf{b}) = \frac{(1-\alpha)}{2}$ is in $\mathcal{P}(\frac{\mathbf{a}+\mathbf{b}}{2})$. Furthermore, we have

$$\mathbf{a} \wedge_0 \mathbf{b} \preceq \mathbf{a} \vee_0 \mathbf{b} \preceq \mathbf{a} \vee_1 \mathbf{b},$$

which means that $\lambda = \lambda_{\frac{\mathbf{a}+\mathbf{b}}{2}}$ and thus the value of the Lovász extension at $\frac{\mathbf{a}+\mathbf{b}}{2}$ is

$$f^L\left(\tfrac{\mathbf{a}+\mathbf{b}}{2}\right) = \tfrac{1}{2} f(\mathbf{a} \wedge_0 \mathbf{b}) + \tfrac{\alpha}{2} f(\mathbf{a} \vee_0 \mathbf{b}) + \tfrac{(1-\alpha)}{2} f(\mathbf{a} \vee_1 \mathbf{b}). \tag{8}$$

Equations (6) and (8) imply (1), so f is α-bisubmodular.

On the other hand, let f be α-bisubmodular. We will show $f^L = f^-$, as then f^L is convex by Proposition 6.

Let $\mathbf{x} \in [-\alpha, 1]^n$. We will show $f^L(\mathbf{x}) = f^-(\mathbf{x})$.

Let

$$\mathcal{M}(\mathbf{x}) := \left\{ \lambda \in \mathcal{P}(\mathbf{x}) \;\middle|\; \sum_{\mathbf{a} \in D^n} \lambda(\mathbf{a}) f(\mathbf{a}) = f^-(\mathbf{x}) \right\}.$$

For every $\mathbf{a} = (a_1, \ldots, a_n) \in D^n$ denote $z(\mathbf{a}) := \left| \left\{ i \in [n] \;\middle|\; a_i = 0 \right\} \right|$. As $\mathcal{M}(\mathbf{x})$ is a compact and non-empty subset of $\mathbb{R}^{D^n}$, the set

$$\left\{ \sum_{\mathbf{a} \in D^n} \lambda(\mathbf{a}) z^2(\mathbf{a}) \;\middle|\; \lambda \in \mathcal{M}(\mathbf{x}) \right\}$$

is a compact and non-empty subset of $\mathbb{R}$ and so contains its supremum. Let $\mu \in \mathcal{M}(\mathbf{x})$ be such that

$$\sum_{\mathbf{a} \in D^n} \mu(\mathbf{a}) z^2(\mathbf{a}) = \max \left\{ \sum_{\mathbf{a} \in D^n} \lambda(\mathbf{a}) z^2(\mathbf{a}) \;\middle|\; \lambda \in \mathcal{M}(\mathbf{x}) \right\}.$$

To show $f^L(\mathbf{x}) = f^-(\mathbf{x})$, it is left to show that $\mu = \lambda_{\mathbf{x}}$. By Lemma 4 it suffices to show that μ is supported by a chain.

Assume that $\operatorname{supp}(\mu)$ is not a chain, and let $\mathbf{a}, \mathbf{b} \in \operatorname{supp}(\mu)$ be incomparable. We will define a function $\nu \in \mathcal{M}(\mathbf{x})$ to contradict the choice of μ. As f is α-bisubmodular, we have

$$f(\mathbf{a} \wedge_0 \mathbf{b}) + \alpha \cdot f(\mathbf{a} \vee_0 \mathbf{b}) + (1-\alpha) \cdot f(\mathbf{a} \vee_1 \mathbf{b}) \leq f(\mathbf{a}) + f(\mathbf{b}). \tag{9}$$

Let $r := \min \left\{ \mu(\mathbf{a}),\ \mu(\mathbf{b}),\ \frac{1-\mu(\mathbf{a} \wedge_0 \mathbf{b})}{1+\alpha},\ 1 - \mu(\mathbf{a} \vee_0 \mathbf{b}),\ 1 - \mu(\mathbf{a} \vee_1 \mathbf{b}) \right\}$. Then $r > 0$ by the choice of $\mathbf{a}$ and $\mathbf{b}$.

Define the function ν on D^n as follows. Case (i): If all $\mathbf{a}, \mathbf{b}, \mathbf{a} \wedge_0 \mathbf{b}, \mathbf{a} \vee_0 \mathbf{b}$ and $\mathbf{a} \vee_1 \mathbf{b}$ are distinct, define

$$\begin{aligned}
\nu(\mathbf{a}) &:= \mu(\mathbf{a}) - r, \\
\nu(\mathbf{b}) &:= \mu(\mathbf{b}) - r, \\
\nu(\mathbf{a} \wedge_0 \mathbf{b}) &:= \mu(\mathbf{a} \wedge_0 \mathbf{b}) + r, \\
\nu(\mathbf{a} \vee_0 \mathbf{b}) &:= \mu(\mathbf{a} \vee_0 \mathbf{b}) + r \cdot \alpha, \\
\nu(\mathbf{a} \vee_1 \mathbf{b}) &:= \mu(\mathbf{a} \vee_1 \mathbf{b}) + r \cdot (1-\alpha), \\
\text{and} \quad \nu(\mathbf{c}) &:= \mu(\mathbf{c}) \quad \text{otherwise.}
\end{aligned} \tag{10}$$

If any of the five elements $\mathbf{a}, \mathbf{b}, \mathbf{a} \wedge_0 \mathbf{b}, \mathbf{a} \vee_0 \mathbf{b}$ and $\mathbf{a} \vee_1 \mathbf{b}$ coincide, we have to make the corresponding adjustments as follows. Firstly note that, as $\mathbf{a}$ and $\mathbf{b}$ are incomparable, it is easy to check that at most one pair of the elements can coincide, and that there are only the following four possibilities for these two coinciding elements: (ii) $\mathbf{a} \wedge_0 \mathbf{b} = \mathbf{a} \vee_0 \mathbf{b}$, (iii) $\mathbf{a} \vee_0 \mathbf{b} = \mathbf{a} \vee_1 \mathbf{b}$, (iv) $\mathbf{a} \vee_1 \mathbf{b} = \mathbf{a}$ and (v) $\mathbf{a} \vee_1 \mathbf{b} = \mathbf{b}$.

In case (ii), we define $\nu(\mathbf{a} \wedge_0 \mathbf{b}) := \mu(\mathbf{a} \wedge_0 \mathbf{b}) + r \cdot (1+\alpha)$ and all other function values as in (10), in case (iii), we define $\nu(\mathbf{a} \vee_0 \mathbf{b}) := \mu(\mathbf{a} \vee_0 \mathbf{b}) + r$ and all other function values as in (10), and in cases (iv) and (v), we define $\nu(\mathbf{a} \vee_1 \mathbf{b}) := \mu(\mathbf{a} \vee_1 \mathbf{b}) - r \cdot \alpha$ and all other function values as in (10).

The image of ν is in $[0,1]$ by the choice of r, and it is easy to check that in all five cases we have

$$\sum_{\mathbf{c} \in \{\mathbf{a},\mathbf{b},\mathbf{a}\wedge_0\mathbf{b},\mathbf{a}\vee_0\mathbf{b},\mathbf{a}\vee_1\mathbf{b}\}} \nu(\mathbf{c}) = \sum_{\mathbf{c} \in \{\mathbf{a},\mathbf{b},\mathbf{a}\wedge_0\mathbf{b},\mathbf{a}\vee_0\mathbf{b},\mathbf{a}\vee_1\mathbf{b}\}} \mu(\mathbf{c}).$$

This yields

$$\sum_{\mathbf{c} \in D^n} \nu(\mathbf{c}) = \sum_{\mathbf{c} \in D^n} \mu(\mathbf{c}) = 1,$$

so ν is a probability distribution. Furthermore, an easy calculation using Equation (7) yields

$$\sum_{\mathbf{c} \in \{\mathbf{a},\mathbf{b},\mathbf{a}\wedge_0\mathbf{b},\mathbf{a}\vee_0\mathbf{b},\mathbf{a}\vee_1\mathbf{b}\}} \nu(\mathbf{c})\mathbf{c} = \sum_{\mathbf{c} \in \{\mathbf{a},\mathbf{b},\mathbf{a}\wedge_0\mathbf{b},\mathbf{a}\vee_0\mathbf{b},\mathbf{a}\vee_1\mathbf{b}\}} \mu(\mathbf{c})\mathbf{c}$$

in all five cases, and so

$$\sum_{\mathbf{c} \in D^n} \nu(\mathbf{c})\mathbf{c} = \sum_{\mathbf{c} \in D^n} \mu(\mathbf{c})\mathbf{c} = \mathbf{x},$$

so $\nu \in \mathcal{P}(\mathbf{x})$. The α-bisubmodularity inequality (9) yields

$$\begin{aligned} \sum_{\mathbf{c} \in \{\mathbf{a},\mathbf{b},\mathbf{a}\wedge_0\mathbf{b},\mathbf{a}\vee_0\mathbf{b},\mathbf{a}\vee_1\mathbf{b}\}} \mu(\mathbf{c})f(\mathbf{c}) - \sum_{\mathbf{c} \in \{\mathbf{a},\mathbf{b},\mathbf{a}\wedge_0\mathbf{b},\mathbf{a}\vee_0\mathbf{b},\mathbf{a}\vee_1\mathbf{b}\}} \nu(\mathbf{c})f(\mathbf{c}) &= \\ r \cdot (f(\mathbf{a}) + f(\mathbf{b}) - f(\mathbf{a} \wedge_0 \mathbf{b}) - \alpha f(\mathbf{a} \vee_0 \mathbf{b}) - (1-\alpha) f(\mathbf{a} \vee_1 \mathbf{b})) &\overset{(9)}{\geq} 0 \end{aligned}$$

and so

$$\sum_{\mathbf{c} \in D^n} \nu(\mathbf{c})f(\mathbf{c}) \leq \sum_{\mathbf{c} \in D^n} \mu(\mathbf{c})f(\mathbf{c}),$$

so $\nu \in \mathcal{M}(\mathbf{x})$. Finally, we will show that

$$\sum_{\mathbf{c} \in D^n} \nu(\mathbf{c})z^2(\mathbf{c}) > \sum_{\mathbf{c} \in D^n} \mu(\mathbf{c})z^2(\mathbf{c}), \tag{11}$$

which is a contradiction to the choice of μ. Let

$$\begin{aligned} A &:= \left|\left\{i \in [n] \;\middle|\; a_i = 0,\ b_i \neq 0\right\}\right|, \\ B &:= \left|\left\{i \in [n] \;\middle|\; b_i = 0,\ a_i \neq 0\right\}\right|, \\ C &:= \left|\left\{i \in [n] \;\middle|\; a_i = b_i = 0\right\}\right| \quad \text{and} \\ N &:= \left|\left\{i \in [n] \;\middle|\; 0 \neq a_i \neq b_i \neq 0\right\}\right|. \end{aligned}$$

The incomparability of $\mathbf{a}$ and $\mathbf{b}$ implies that we have either $N > 0$ or, if $N = 0$, we have both $A > 0$ and $B > 0$. It is easy to check that

$$\begin{aligned} z(\mathbf{a} \wedge_0 \mathbf{b}) &= A + B + C + N, \\ z(\mathbf{a} \vee_0 \mathbf{b}) &= C + N, \\ z(\mathbf{a} \vee_1 \mathbf{b}) &= C, \\ z(\mathbf{a}) &= A + C, \\ z(\mathbf{b}) &= B + C, \end{aligned}$$

and so

$$\begin{aligned} & z(\mathbf{a} \wedge_0 \mathbf{b})^2 + \alpha \cdot z(\mathbf{a} \vee_0 \mathbf{b})^2 + (1-\alpha) \cdot z(\mathbf{a} \vee_1 \mathbf{b})^2 - z(\mathbf{a})^2 - z(\mathbf{b})^2 \\ &= (A+B+C+N)^2 + \alpha(C+N)^2 + (1-\alpha)C^2 - (A+C)^2 - (B+C)^2 \\ &= 2(AB + AN + BN + CN) + N^2 + 2\alpha CN + \alpha N^2 \\ &= 2(AB + AN + BN + (1+\alpha)CN) + (1+\alpha)N^2 \ > \ 0, \end{aligned}$$

as $N > 0$ or $AB > 0$. As $r > 0$ this implies

$$r(z(\mathbf{a} \wedge_0 \mathbf{b})^2 + \alpha \cdot z(\mathbf{a} \vee_0 \mathbf{b})^2 + (1-\alpha) \cdot z(\mathbf{a} \vee_1 \mathbf{b})^2 - z(\mathbf{a})^2 - z(\mathbf{b})^2) > 0.$$

An easy calculation yields

$$\sum_{\mathbf{c} \in \{\mathbf{a}, \mathbf{b}, \mathbf{a} \wedge_0 \mathbf{b}, \mathbf{a} \vee_0 \mathbf{b}, \mathbf{a} \vee_1 \mathbf{b}\}} \nu(\mathbf{c}) z^2(\mathbf{c}) > \sum_{\mathbf{c} \in \{\mathbf{a}, \mathbf{b}, \mathbf{a} \wedge_0 \mathbf{b}, \mathbf{a} \vee_0 \mathbf{b}, \mathbf{a} \vee_1 \mathbf{b}\}} \mu(\mathbf{c}) z^2(\mathbf{c})$$

in all five cases for the definition of ν.

From this, the contradicting inequality (11) follows. So μ is supported by a chain, and this implies $\mu = \lambda_{\mathbf{x}}$, which means that $f^L(\mathbf{x}) = f^-(\mathbf{x})$.

Thus $f^L = f^-$ holds and f^L is convex. □

REFERENCES

[1] Bouchet, A.: Greedy algorithm and symmetric matroids. Mathematical Programming, 38, 147–159, 1987

[2] Bouchet, A. and Cunningham, W.H.: Delta-matroids, jump systems and bisubmodular polyhedra. SIAM J. Discrete Math., 8, 17–32, 1995

[3] Chandrasekaran, R. and Kabadi, S.N.: Pseudomatroids. Discrete Math., 71, 205–217, 1988

[4] Cohen, D., Cooper, M., Jeavons, P., and Krokhin, A.: The complexity of soft constraint satisfaction. *Artificial Intelligence*, 170(11):983–1016, 2006.

[5] Edmonds, J.: Submodular functions, matroids, and certain polyhedra. In: R. Guy, H. Hanani, N. Sauer, J. Schönheim (eds.) Combinatorial Structures and Their Applications, pp. 69–87. Gordon and Breach, 1970

[6] Fujishige, S.: Submodular Functions and Optimization. Elsevier, 2005.

[7] Fujishige, S., Tanigawa, S., and Yoshida, Y.: Generalized skew bisubmodularity: A characterization and a min-max theorem. Discrete Optimization, 12, 1-9, 2014.

[8] Grötschel, M., Lovász, L., and Schrijver, A.: The ellipsoid method and its consequences in combinatorial optimization. Combinatorica, 1, 169-197, 1981.

[9] Huber, A., Krokhin, A., and Powell, R.: Skew Bisubmodularity and Valued CSPs. In *Proceedings of SODA'13*, pages 1296-1305, 2013.

[10] Huber, A., Krokhin, A., and Powell, R.: Skew Bisubmodularity and Valued CSPs. SIAM Journal on Computing, 43(3), 10641084, 2014.

[11] Kabadi, S.N. and Chandrasekaran, R.: On totally dual integral systems. Discrete Appl. Math., 26, 87–104, 1990
[12] Krokhin, A. and Larose, B.: Maximizing supermodular functions on product lattices, with application to maximum constraint satisfaction. SIAM Journal on Discrete Mathematics, 22(1), 312-328, 2008.
[13] Kuivinen, F.: On the complexity of submodular function minimisation on diamonds. Discrete Optimization, 8(3), 459-477, 2011.
[14] Lovász, L.: Submodular functions and convexity. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical Programming: The State of the Art*, pages 235–257. Springer, 1983.
[15] McCormick, S.T.: Submodular function minimization. In: K. Aardal, G. Nemhauser, and R. Weismantel, editors, *Handbook on Discrete Optimization*, pages 321–391. Elsevier, 2006.
[16] Thapper, J. and Živný, S.: The power of linear programming for valued CSPs. In *Proceedings of FOCS'12*, pages 669-678, 2012.
[17] Thapper, J. and Živný, S.: The complexiy of finite-valued CSPs. In *Proceedings of STOC'13*, pages 695-704, 2013.
[18] Qi, L.: Directed submodularity, ditroids and directed submodular flows. *Mathematical Programming*, 42(1–3):579–599, 1988.
[19] Schrijver, A.: Combinatorial Optimization: Polyhedra and Efficiency. Springer, 2004