On Submodular and Supermodular Functions on Lattices and Related Structures

Dan A. Simovici

University of Massachusetts Boston
Department of Computer Science
Boston, USA
dsim@cs.umb.edu

Abstract

We give single-operations characterizations for submodular and supermodular functions on lattices that have monotonicity properties. We associate to such functions metrics on lattices and we investigate corresponding metrics on the sets of partitions.

***Keywords*-lattice; semilattice; submodularity; entropy;**

I. Introduction

Submodular functions are useful in combinatorial optimization problems [3], [4], [6]. They are defined as functions of the form $f : \mathcal{P}(S) \longrightarrow \mathbb{R}$, where S is a finite set and $\mathcal{P}(S)$ is the set of subsets of S, which satisfy the submodular inequality, that is, $f(X \cap Y) + f(X \cup Y) \leqslant f(X) + f(Y)$ for $X, Y \in \mathcal{P}(S)$. This inequality is equivalent to the "diminishing return property" of these functions which means that for every $X, Y \in \mathcal{P}(S)$ such that $X \subseteq Y$ and $x \in S - Y$, we have

$$f(X \cup \{x\}) - f(X) \geqslant f(Y \cup \{x\}) - f(Y).$$

In this note we study submodular functions and their duals (known as supermodular functions) defined on lattices and we link these functions with a generalization of conditional entropy in lattices and with certain classes of metrics defined on these structures. The characterization of submodular or supermodular functions that have a monotonicity-linked property (obtained in Section II) is formulated using only one of the lattice operations, which opens the possibility of extending the notion of modularity to semilattices.

Section III is dedicated to submodular and supramodular functions defined on partition lattices.

The extension of semimodularity to functions defined on semilattices is of interest for the generalization of the notion of entropy (and of metrics derived from this notion) from partition lattices to other algebraic structures that play a role in designing data mining algorithms.

II. Submodular Functions on Lattices

A *semilattice* is a semigroup $(S, \diamond)$, where $\diamond$ is a commutative and idempotent operation. This is a pervasive algebraic structure, with numerous applications in mathematics and computer science.

A *lattice* is an algebraic structure $(L, \vee, \wedge)$ such that both $(L, \vee)$ and $(L, \wedge)$ are semilattices and the two operations $\vee$ and $\wedge$ satisfy the absorption laws

$$(x \vee y) \wedge y = y \text{ and } (x \wedge y) \vee y = y,$$

for $x, y \in L$.

Every semilattice $(S, \diamond)$ generates a partial order on S defined by $x \leqslant y$ if and only if $x \diamond y = y$. Thus, a $(L, \vee, \wedge)$ generates two partial orders on L, "$\leqslant_1$" and "$\leqslant_2$" defined by

$$x \leqslant_1 y \text{ if } x \vee y = y, \text{ and } x \leqslant_2 y \text{ if } x \wedge y = y$$

for $x, y \in L$. Note that, by the absorption laws, we have $u \leqslant_1 v$ if and only if $v \leqslant_2 u$ for $u, v \in L$. The partial orders $\leqslant_1$ and $\leqslant_2$ are said to be *dual* of each other. Unless stated explicitly otherwise, we shall use the partial order $\leqslant_1$ on lattices and will denote it simply by "$\leqslant$".

If $(P, \leqslant)$ and $(Q, \leqslant)$ are two partially ordered sets (posets), then $f : P \longrightarrow Q$ is a *monotonic* function if $u \leqslant v$ implies $f(u) \leqslant f(v)$ for $u, v \in P$. If $u < v$ implies $f(u) < f(v)$, then f is said to be *strictly monotonic*. The set of monotonic (strictly monotonic) functions from P to Q is denoted by $\mathsf{MON}(P, Q)$ ($\mathsf{sMON}(P, Q)$, respectively).

If $u \leqslant v$ implies $f(u) \geqslant f(v)$ for $u, v \in S$, then f is an *anti-monotonic function*; the function f is strictly anti-monotonic if $u < v$ implies $f(u) > f(v)$ for all $u, v \in P$. The set of anti-monotonic (strictly anti-monotonic) functions is denoted by $\mathsf{A\text{-}MON}(P,Q)$ ($\mathsf{sA\text{-}MON}(P,Q)$, respectively).

Let $(L,\vee,\wedge)$ be a lattice. A function $f : L \longrightarrow \mathbb{R}$ is *submodular* if

$$f(x \vee y) + f(x \wedge y) \leqslant f(x) + f(y),$$

for every $x, y \in L$.

The function $f : L \longrightarrow \mathbb{R}$ is *supermodular* if

$$f(x \vee y) + f(x \wedge y) \geqslant f(x) + f(y),$$

for every $x, y \in L$.

The classes of submodular functions and supermodular functions on $(L,\vee,\wedge)$ are denoted by $\mathsf{SUBM}(L,\vee,\wedge)$ and by $\mathsf{SUPM}(L,\vee,\wedge)$, respectively. When the lattice is clear from context we denote these classes by SUBM and SUPM.

If a function $f : L \longrightarrow \mathbb{R}$ is both submodular and supermodular, then f satisfies the equality $f(x \vee y) + f(x \wedge y) = f(x) + f(y)$ for $x, y \in L$. Such functions are known as *modular functions* or as *lattice valuations*. The last term is used in [1], where lattices that have strictly monotonic valuations are referred to as metric lattices. We show here that metrics can be introduced on lattices using submodular or supermodular functions and that the presence on metrics that have certain monotonicity properties imply the existence of submodular or supermodular functions.

EXAMPLE 2.1. Many submodular functions can be naturally associated to finite graphs (see [5], [2]). Let $\mathcal{G} = (V,E)$ be a graph having V as its set of vertices and E as its set of edges.

For a set of edges K of $\mathcal{G}$ let $v(K)$ be the number of vertices incident with an edge in K. It is easy to see that $v : \mathcal{P}(E) \longrightarrow \mathbb{R}$ is both monotonic and submodular.

Let S be a set of vertices of $\mathcal{G}$ and let $t(S)$ be the number of edges whose endpoints are in S. Then, t is monotonic and supermodular. Similarly, if $\ell(S)$ is the number of edges that have at least one end point in S, then ℓ is a monotonic and submodular function. □

The next theorem allows the introduction of two subtypes of submodular functions.

THEOREM 2.1. *Let $(L,\vee,\wedge)$ be a lattice.*

If $f : L \longrightarrow \mathbb{R}$ is a function such that

$$f(t) + f(u \wedge v) \leqslant f(t \wedge u) + f(t \wedge v), \tag{2.1}$$

for $t, u, v \in L$, or

$$f(t) + f(u \vee v) \leqslant f(t \vee u) + f(t \vee v) \tag{2.2}$$

for $t, u, v \in L$, then f is a submodular function on L.

Proof: Suppose that $f(t) + f(u \wedge v) \leqslant f(t \wedge u) + f(t \wedge v)$ for $t, u, v \in L$. Substituting $u \vee v$ for t we obtain

$$\begin{aligned} & f(u \vee v) + f(u \wedge v) \\ \leqslant \quad & f((u \vee v) \wedge u) + f((u \vee v) \wedge v) \\ = \quad & f(u) + f(v), \end{aligned}$$

by the absorption property of L, which shows that f is submodular.

If $f(t)+f(u\vee v) \leqslant f(t\vee u)+f(t\vee v)$ for $t, u, v \in L$, by substituting $u \wedge v$ for t we obtain

$$\begin{aligned} & f(u \wedge v) + f(u \vee v) \\ \leqslant \quad & f((u \wedge v) \vee u) + f((u \wedge v) \vee v) \\ = \quad & f(u) + f(v), \end{aligned}$$

so f is again, submodular. ■

Theorem 2.1 allows us to introduce two classes of functions on a lattice $(L,\vee,\wedge)$. Namely, $\mathsf{SUBM}(L,\vee,\wedge)_\wedge$ consists of those function that satisfy Inquality (2.1) and $\mathsf{SUBM}(L,\vee,\wedge)_\vee$ consists of those function that satisfy Inquality (2.2).

THEOREM 2.2. *If $f : L \longrightarrow \mathbb{R}$ is a function such that $f(t) + f(u \wedge v) \geqslant f(t \wedge u) + f(t \wedge v)$ for $t, u, v \in L$, or $f(t) + f(u \vee v) \geqslant f(t \vee u) + f(t \vee v)$ for $t, u, v \in L$, then f is a supermodular function on L.*

Proof: The proof of this theorem that refers to supermodularity has an entirely similar argument. ■

We denote the classes of functions

$$\mathsf{SUBM}(L,\vee,\wedge)_\wedge, \mathsf{SUBM}(L,\vee,\wedge)_\vee,$$

by $\mathsf{SUBM}_\wedge$, $\mathsf{SUBM}_\vee$, respectively, when the lattice L is clear from context.

THEOREM 2.3. *Let $(L,\vee,\wedge)$ be a lattice. Any function $f \in \mathsf{SUBM}_\wedge$ is anti-monotonic and any function in $\mathsf{SUBM}_\vee$ is monotonic.*

Proof: Let $f : L \longrightarrow \mathbb{R}$ be a function in $\mathsf{SUBM}_\wedge$ and let t, u be two elements of L such that $t \leqslant u$. Choosing $v = u$ in the definition of $\mathsf{SUBM}_\wedge$ yields

$$f(t) + f(u) \leqslant f(t) + f(t \wedge u) = 2f(t),$$

which implies $f(u) \leqslant f(t)$. Thus, f is anti-monotonic.

Similarly, choosing $v = u$ in the definition of $\mathsf{SUBM}_\vee$ we obtain

$$f(t) + f(u) \leqslant 2f(u),$$

so $f(t) \leqslant f(u)$, which shows that f is monotonic. ■

THEOREM 2.4. *If $f \in \mathsf{SUBM}$ and f is anti-monotonic, then $f \in \mathsf{SUBM}_\wedge$; if $f \in \mathsf{SUBM}$ and f is monotonic, then $f \in \mathsf{SUBM}_\vee$.*

Proof: Let f be a submodular and anti-monotonic function. The submodularity implies

$$f((t\wedge u)\wedge(t\wedge v))+f((t\wedge u)\vee(t\wedge v))\leqslant f(t\wedge u)+f(t\wedge v)$$

for every $t,u,v\in L$. Since

$$(t\wedge u)\wedge(t\wedge v)\leqslant t$$

it follows that $f(t)\leqslant f((t\wedge u)\wedge(t\wedge v))$. By the subdistributive inequality that holds in any lattice (see [1] or [7]) we have

$$(t\wedge u)\vee(t\wedge v)\leqslant t\wedge(u\vee v),$$

hence $f(t\wedge(u\vee v))\leqslant f((t\wedge u)\vee(t\wedge v))$ because f is anti-monotonic. Again, by the anti-monotonicity of f,

$$f(u\vee v)\leqslant f(t\wedge(u\vee v)),$$

so $f(u\vee v)\leqslant f((t\wedge u)\vee(t\wedge v))$. Consequently, we have

$$\begin{aligned} f(t)+f(u\wedge v) &\leqslant f((t\wedge u)\wedge(t\wedge v)) \\ &\quad +f((t\wedge u)\vee(t\wedge v)) \\ &\leqslant f(t\wedge u)+f(t\wedge v). \end{aligned}$$

The second statement of the theorem has a similar argument. ■

COROLLARY 2.2. *For any lattice $(L,\vee,\wedge)$ we have*

$$\begin{aligned} \mathit{SUBM}_\vee &= \mathit{SUBM}\cap\mathit{MON}, \\ \mathit{SUBM}_\wedge &= \mathit{SUBM}\cap\mathit{A\text{-}MON}, \\ \mathit{SUPM}_\vee &= \mathit{SUPM}\cap\mathit{MON}, \\ \mathit{SUPM}_\wedge &= \mathit{SUPM}\cap\mathit{A\text{-}MON}. \end{aligned}$$

For a function $f:L\longrightarrow\mathbb{R}$ define the functions $\kappa_f:L^2\longrightarrow\mathbb{R}_{\geqslant 0}$ and $\lambda_f:L^2\longrightarrow\mathbb{R}_{\geqslant 0}$ by

$$\begin{aligned} \kappa_f(x,y) &= f(x\wedge y)-f(y), \\ \lambda_f(x,y) &= f(x\vee y)-f(y) \end{aligned}$$

for $x,y\in L$. Note that f is submodular (supermodular) if and only if $\kappa_f(x,y)+\lambda_f(x,y)\leqslant 0$ ($\kappa_f(x,y)+\lambda_f(x,y)\geqslant 0$). The functions κ_f and λ_f are intended as abstract counterparts of conditional entropy.

The next result shows that the monotonicity properties of κ_f and λ_f in their first argument imply monotonicity properties for f, while monotonicity properties of κ_f and λ_f in their second argument imply modularity properties for f.

THEOREM 2.5. *Let $(L,\vee,\wedge)$ be a lattice. If $f\in\mathit{SUBM}_\wedge$, then κ_f is anti-monotonic in its first argument and monotonic in the second. If $f\in\mathit{SUBM}_\vee$, then λ_f is monotonic in its first argument and anti-monotonic in the second.*

Conversely, if κ_f is anti-monotonic in its first argument, then f is anti-monotonic; if κ_f is monotonic in its second argument, then f is submodular.

Also, if λ_f is monotonic in its first argument, then f is monotonic; if λ_f is anti-monotonic in its second argument, then f is supermodular.

Proof: The anti-monotonicity of κ_f in its first argument follows directly from the anti-monotonicity of f. To prove the monotonicity of κ_f in its second argument let $y,z\in L$ be such that $y\leqslant z$. The definition of $\mathsf{SUBM}_\wedge$ allows us to write

$$\begin{aligned} f(z)+f(x\wedge y) &\leqslant f(z\wedge x)+f(z\wedge y) \\ &= f(z\wedge x)+f(y) \text{ (because } y\leqslant z), \end{aligned}$$

which translates into $\kappa_f(x,y)\leqslant\kappa_f(x,z)$.

Similarly, the monotonicity of λ_f in its first argument follows from the monotonicity of f. To prove the anti-monotonicity of λ_f in its second argument, let $y,z\in L$ be such that $y\leqslant z$. Since $f\in\mathsf{SUBM}_\vee$ we have

$$\begin{aligned} f(y)+f(x\vee z) &\leqslant f(y\vee x)+f(y\vee z) \\ &= f(y\vee x)+f(z) \text{ (because } y\leqslant z), \end{aligned}$$

which amounts to $\lambda_f(x,z)\leqslant\lambda_f(x,y)$.

For the converse implications suppose that κ_f is anti-monotonic in its first argument, so $x_1\leqslant x_2$ implies $\kappa_f(x_1,y)\geqslant\kappa_f(x_2,y)$, that is $f(x_1\wedge y)\geqslant f(x_2\wedge y)$ for every $y\in L$. Choosing $y=x_2$ it follows that $f(x_1)\geqslant f(x_2)$, that is, f is anti-monotonic.

Suppose now that κ_f is monotonic in its second argument, that is $y_1\leqslant y_2$ implies $f(x\wedge y_1)-f(y_1)\geqslant f(x\wedge y_2)-f(y_2)$. Choosing $y_2=x\vee y_1$ yields

$$f(x\wedge y_1)+f(x\vee y_1)\leqslant f(y_1)+f(x),$$

which shows the submodularity of f.

Similar arguments can be made for the last part of the theorem involving the function δ_f. ■

THEOREM 2.6. *Let $(L,\vee,\wedge)$ be a lattice.*

If $f\in\mathit{SUBM}_\wedge$, then $\kappa_f(u,v)+\kappa_f(v,w)\geqslant\kappa_f(u,w)$ for $u,v,w\in L$.

If $f\in\mathit{SUBM}_\vee$, then $\lambda_f(u,v)+\lambda_f(v,w)\geqslant\lambda_f(u,w)$ for $u,v,w\in L$.

Proof: It is easy to see that by expressing κ_f in terms of f the inequality that we need to prove is equivalent to

$$f(u\wedge v)+f(v\wedge w)\geqslant f(u\wedge w)+f(v),$$

which holds by the definition of $\mathsf{SUBM}_\wedge$. The proof of the second part is similar. ■

Note that $\kappa_f(x,x)=0$ for $f\in\mathsf{SUBM}_\wedge$ and $x\in L$. Similarly, $\lambda_f(x,x)=0$ for $f\in\mathsf{SUBM}_\vee$ and $x\in L$.

For $f\in\mathsf{SUBM}_\wedge$ define the mapping $d_f:L^2\longrightarrow\mathbb{R}$ as

$$d_f(x,y)=\kappa_f(x,y)+\kappa_f(y,x)=2f(x\wedge y)-f(x)-f(y)$$

for $x, y \in L$. Similarly, for $g \in \mathsf{SUBM}_\vee$, let $\delta_f : L^2 \longrightarrow \mathbb{R}$ be given by

$$\delta_g(x,y) = \lambda_g(x,y) + \lambda_g(y,x) = 2g(x \vee y) - f(x) - f(y)$$

for $x, y \in L$.

It follows immediately from Theorem 2.6 that d_f is a semimetric on L, when f belongs to $\mathsf{SUBM}_\wedge$ and that δ_g has the same property if $g \in \mathsf{SUBM}_\vee$. If the functions involved are in sMON (in the first case) or in sA-MON (in the second), then d_f or δ_g are metrics.

Conversely, if d_f is a semimetric on L, where $d_f(x,y) = 2f(x \wedge y) - f(x) - f(y)$, then $f \in \mathsf{SUBM}_\wedge$. Indeed, in this case, the triangular inequality of d_f amounts to

$$\begin{aligned} & 2f(x \wedge y) - f(x) - f(y) \\ & \quad +2f(y \wedge z) - f(y) - f(z) \\ \geqslant & \quad 2f(x \wedge z) - f(x) - f(z), \end{aligned}$$

for $x, y, z \in L$, which is clearly equivalent to the defining equality of $\mathsf{SUBM}_\wedge$. Similarly, if d_g is a semimetric on L, g belongs to $\mathsf{SUBM}_\vee$.

III. Submodular Functions on the Lattice of Partitions

A *partition* of a non-empty set S is a collection of non-empty subsets of S, $\pi = \{B_i \mid i \in I\}$ such that $i, j \in I$ and $i \neq j$ implies $B_i \cap B_j = \emptyset$ and $\bigcup_{i \in I} B_i = S$. The subsets B_i are referred as *blocks* of π. The set of partitions of a set S is denoted by $\mathsf{PART}(S)$.

If $\pi, \tau \in \mathsf{PART}(S)$ we write $\pi \leqslant \tau$ if every block of π is included in a block of τ. This relation between partitions is a partial order. The largest element of the partially ordered set $(\mathsf{PART}(S), \leqslant)$ is the one-block partition $\omega_S = \{S\}$, while the smallest element is the partition $\alpha_S = \{\{x\} \mid x \in S\}$.

We assume from now on that all partitions are considered over finite sets.

The partial ordered set $(\mathsf{PART}(S), \leqslant)$ is actually a lattice. The meet of two partitions $\pi \wedge \tau$ is the partition of S whose blocks are the non-empty intersections of the form $B \cap C$, where $B \in \pi$ and $C \in \sigma$.

Consider the bipartite graph $\mathcal{G}_{\pi,\sigma}$ having $\{B_1, \ldots, B_m, C_1, \ldots, C_n\}$ as its vertices, where

$$\pi = \{B_1, \ldots, B_m\} \text{ and } \sigma = \{C_1, \ldots, C_n\}.$$

An edge exists between B_i and C_j if and only if $B_i \cap C_j \neq \emptyset$. The blocks of the partition $\pi \wedge \sigma$ consist of non-empty sets of the form $B_i \cap C_j$ and correspond to the edges of $\mathcal{G}_{\pi,\sigma}$.

Let $\mathcal{C}_1, \ldots, \mathcal{C}_k$ be the connected components of $\mathcal{G}_{\pi,\sigma}$. For every connected component $\mathcal{C}$ we have

$$\bigcup \{B_i \mid B_i \in \mathcal{C}\} = \bigcup \{C_j \mid C_j \in \mathcal{C}\}$$

and that the blocks of the partition $\pi \vee \sigma$ have the form $\bigcup \mathcal{C}$.

A partition π covers a partition μ in $(\mathsf{PART}(S), \leqslant)$ if π can be obtained from μ by fusing two blocks of μ. Partition lattices are prototypical for the so called upper semimodular lattices [1], characterized by the following property: if $\pi_1 \neq \pi_2$ and both π_1, π_2 cover a partition σ, the $\pi_1 \vee \pi_2$ covers both π_1 and π_2.

If $\pi \in \mathsf{PART}(S)$ and $\emptyset \neq C \subseteq S$, we denote by π_C the partition $\pi_C = \{B \cap C \mid B \in \pi\}$. This is the *trace* of π on C.

Note that if $\pi, \sigma \in \mathsf{PART}(S)$, $\pi = \{B_1, \ldots, B_n\}$, and $\sigma = \{C_1, \ldots, C_n\}$, then we have

$$\pi \wedge \sigma = \pi_{C_1} + \cdots + \pi_{C_n} = \sigma_{B_1} + \cdots + \sigma_{B_m}. \quad (3.3)$$

Let S be a finite set such that $|S| \geqslant 2$ and let β be a number, $\beta > 1$.

For a partition $\pi = \{B_1, \ldots, B_m\} \in \mathsf{PART}(S)$ define the function $f_S : \mathsf{PART}(S) \longrightarrow \mathbb{R}$ as

$$f_S(\pi) = b \left(1 - \sum_{j=1}^{m} \left(\frac{|B_j|}{|S|} \right)^\beta \right),$$

where $\beta > 1$. Then, $f_S(\omega_S) = 0$ and $f_S(\alpha_S) = b(1 - |S|^{\beta-1})$.

Let $S_1, \ldots, S_\ell$ be ℓ non-empty and pairwise disjoint sets and let $S = \bigcup_{k=1}^{\ell} S_k$. Assume that $\pi_k = \{B_{k1}, \ldots, B_{km_k}\}$ is a partition on S_k for $1 \leqslant k \leqslant \ell$. Then, the collection of sets $\{B_{kj} \mid 1 \leqslant k \leqslant \ell, 1 \leqslant j \leqslant m_k\}$ is a partition of the set S denoted by $\pi_1 + \cdots + \pi_\ell$. We have

$$f_S(\pi_1 + \cdots + \pi_\ell) = \sum_{k=1}^{\ell} \left(\frac{|S_k|}{|S|} \right)^\beta f_{S_k}(\pi_k) + f_S(\{S_1, \ldots, S_\ell\}). \quad (3.4)$$

Therefore, taking into account Equalities (3.3) we can write

$$f_S(\pi \wedge \sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|} \right)^\beta f_{C_j}(\pi_{C_j}) + f_S(\sigma).$$

By the definition of κ_{f_S} we have

$$\kappa_{f_S}(\pi, \sigma) = f_S(\pi \wedge \sigma) - f_S(\sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|} \right)^\beta f_{C_j}(\pi_{C_j}). \quad (3.5)$$

LEMMA 3.1. *Let $\phi : [0,1] \longrightarrow \mathbb{R}$ be a convex function such that $\phi(x) \leqslant x$ for $x \in [0,1]$, $w_1, \ldots, w_n$ be n positive numbers such that $\sum_{i=1}^{n} w_i = 1$, and let*

$a_1, \ldots, a_n \in [0,1]$. *We have*

$$1 - \phi\left(\sum_{i=1}^n w_i a_i\right) - \phi\left(\sum_{i=1}^n w_i(1-a_i)\right) \geqslant \sum_{i=1}^n \phi(w_i)(1-\phi(a_i)-\phi(1-a_i)).$$

Proof: By Jensen's inequality applied to ϕ we have

$$\begin{aligned} \phi\left(\sum_{i=1}^n w_i a_i\right) &\leqslant \sum_{i=1}^n w_i \phi(a_i), \\ \phi\left(\sum_{i=1}^n w_i (1-a_i)\right) &\leqslant \sum_{i=1}^n w_i \phi(1-a_i). \end{aligned}$$

Taking into account that $\sum_{i=1}^n w_i = 1$ we have

$$\begin{aligned} 1 - \phi\left(\sum_{i=1}^n w_i a_i\right) - \phi\left(\sum_{i=1}^n w_i(1-a_i)\right) &\geqslant \sum_{i=1}^n w_i(1-\phi(a_i)-\phi(1-a_i)) \\ &\geqslant \sum_{i=1}^n \phi(w_i)(1-\phi(a_i)-\phi(1-a_i)) \end{aligned}$$

because $w_i \geqslant \phi(w_i)$ for $1 \leqslant i \leqslant n$. ■

LEMMA 3.2. *Let $\pi \in \mathsf{PART}(S)$ and let C, D be two non-empty disjoint subsets of S. We have*

$$|C|^\beta f_C(\pi_C) + |D|^\beta f_D(\pi_D) \leqslant (|C|+|D|)^\beta f_{C\cup D}(\pi_{C\cup D}).$$

Proof: Let $\pi = \{B_1, \ldots, B_n\}$. Define

$$w_i = \frac{|B_i \cap (C \cup D)|}{|C \cup D|}, a_i = \frac{|B_i \cap C|}{|B_i \cap (C \cup D)|}$$

for $1 \leqslant i \leqslant n$, so $1 - a_i = \frac{|B_i \cap D|}{|B_i \cap (C\cup D)|}$.

By Lemma 3.1 applied to the function $\phi(x) = x^\beta$, that is convex on $[0,1]$ when $\beta > 1$ we have:

$$\begin{aligned} 1 &- \left(\sum_{i=1}^n \frac{|B_i \cap C|}{|C \cup D|}\right)^\beta - \left(\sum_{i=1}^n \frac{|B_i \cap D|}{|C \cup D|}\right)^\beta \\ &\geqslant \sum_{i=1}^n \left(\frac{|B_i \cap (C\cup D)|}{|C \cup D|}\right)^\beta \left(1 - \left(\frac{|B_i \cap C|}{|B_i \cap (C \cup D)|}\right)^\beta \right. \\ &\quad \left. - \left(\frac{|B_i \cap D|}{|B_i \cap (C \cup D)|}\right)^\beta\right), \end{aligned}$$

which is equivalent to the inequality of the lemma. ■

THEOREM 3.1. *The function $f_S : \mathsf{PART}(S) \longrightarrow \mathbb{R}$ is anti-monotonic and submodular.*

Proof: To prove that f_S is anti-monotonic it suffices to show that if $\pi \leqslant \tau$ such that τ covers π, then $f(\pi) \geqslant f(\tau)$.

Suppose that $\pi = \{B_1, \ldots, B_m\}$; without loss of generality we may assume that τ results from π by fusing the blocks B_{m-1} and B_m. Since B_{m-1} and B_m are non-empty sets we have $|B_{m-1}| \geqslant 1$ and $|B_m| \geqslant 1$ which implies $|B_{m-1}|^\beta + |B_m|^\beta \leqslant |B_{m-1} \cup B_m|^\beta$. Therefore,

$$\begin{aligned} f_S(\pi) &= b\left(1 - \sum_{j=1}^m \left(\frac{|B_j|}{|S|}\right)^\beta\right) \\ &\geqslant b\left(1 - \sum_{j=1}^{m-2} \left(\frac{|B_j|}{|S|}\right)^\beta - \left(\frac{|B_{m-1} \cup B_m|}{|S|}\right)^\beta\right) \\ &= f_S(\tau), \end{aligned}$$

which allows us to conclude that f_S is indeed anti-monotonic.

To prove that f_S is submodular we shall use the second part of Theorem 2.5 and show that the function $\kappa_{f_S}(\pi, \sigma)$ is monotonic in its second argument.

Let $\pi, \sigma, \tau \in \mathsf{PART}(S)$ such that $\sigma \leqslant \tau$ and τ covers σ. Again, we may assume without loss of generality that $\sigma = \{C_1, \ldots, C_n\}$ and τ is obtained from σ by fusing C_{n-1} and C_n. By Equality (3.5) it suffices to show that

$$\begin{aligned} &|C_{n-1}|^\beta f_{C_{n-1}}(\pi_{C_{n-1}}) + |C_n|^\beta f_{C_n}(\pi_{C_n}) \\ &\quad \leqslant |C_{n-1} \cup C_n|^\beta f_{C_{n-1}\cup C_n}(\pi_{C_{n-1}\cup C_n}), \end{aligned}$$

which holds by Lemma 3.2. ■

IV. Further Work

The characterization of submodular (or supermodular) monotonic and anti-monotonic functions provided by Corollary 2.2 makes use only of one of the operations of the lattice. This makes it possible to extend the notions of submodularity and supermodularity to functions defined on semilattices. This extension is relevant to defining entropies for set covers and metrics on the space of covers of a set. In turn, metrics on set covers can help extending well-known data mining algorithms that make use of the metric space of partitions in feature selection and classification to multi-valued attributes.

Acknowledgement

This research was supported by a S&T grant from the University of Massachusetts.

References

[1] G. Birkhoff. *Lattice Theory*. American Mathematical Society, Providence, Rhode Island, third edition, 1995.

[2] A. Frank. Submodular functions in graph theory. *Discrete Mathematics*, 111:231–243, 1993.

[3] B. Korte and J. Vygen. *Combinatorial Optimization: Theory and Algorithms*. Springer, Heidelberg, fifth edition, 2012.

[4] J. Lee. *A First Course in Combinatorial Optimization*. Cambridge University Press, Cambridge, 2004.
[5] H. Narayanan. *Submodular Functions and Electrical Networks*. North Holland, 1997.
[6] A. Schrijver. *Combinatorial Optimization*. Springer, New York, 2002.
[7] D. A. Simovici and C. Djeraba. *Mathematical Tools for Data Mining*. Springer, London, second edition, 2014.