Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Discrete Applied Mathematics 142 (2004) 53–65
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

Disjunctive analogues of submodular and supermodular pseudo-Boolean functions

S. Foldes[a,1], P.L. Hammer[b]

[a] *Department of Mathematics, Tampere University of Technology, PL 553, Tampere-Hervanta 33101, Finland*
[b] *RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway, NJ 08854-8003, USA*

Received 8 February 2002; received in revised form 24 November 2003; accepted 8 December 2003

Abstract

We consider classes of real-valued functions of Boolean variables defined by disjunctive analogues of the submodular and supermodular functional inequalities, obtained by replacing in these inequalities addition by disjunction (max operator). The disjunctive analogues of submodular and supermodular functions are completely characterized by the syntax of their disjunctive normal forms. Classes of functions possessing combinations of these properties are also examined. A disjunctive representation theory based on one of these combination classes exhibits syntactic and algorithmic analogies with classical DNF theory.

Keywords: Submodularity; Supermodularity; Pseudo-Boolean functions; Disjunctive normal form; Implicents; Conseusus; Resolution

1. Preliminaries

We consider *pseudo-Boolean* functions, i.e. functions from some discrete hypercube $\mathbb{B}^n$ (where $\mathbb{B} = \{0, 1\}$, and $n \geqslant 1$) to the real numbers $\mathbb{R}$. To emphasize n, we shall say that a function is n-ary. We recall that $\mathbb{B}^n$ is a distributive and complemented (i.e. Boolean) lattice, whose partial order is defined by componentwise comparison. The lattice join and meet of two vectors V, W in $\mathbb{B}^n$ are denoted by $V \vee W$ and VW.

The set $\mathbb{R}$ being totally ordered, it is trivially a distributive lattice where the join and meet of two numbers is simply their maximum and minimum, denoted by $\vee$ and $\wedge$, respectively. Here in $\mathbb{R}$ we shall avoid using juxtaposition for meet, as the product generally differs from the meet (unlike within $\mathbb{B}$ or in the Boolean ring $\mathbb{B}^n$).

Pseudo-Boolean functions include Boolean functions (that may be identified with those pseudo-Boolean functions whose range is contained in $\{0, 1\}$), as well as various combinatorial set functions such as matroid rank functions (by representing each argument set by its characteristic vector). Several classes of functions have been thoroughly investigated. The class of *monotone non-decreasing* functions consists of those pseudo-Boolean functions f that satisfy the inequality

$$f(V) \leqslant f(V \vee W)$$

for all V, W in their domain. Similarly, f is *monotone non-increasing* if it always satisfies

$$f(V) \geqslant f(V \vee W).$$

A function f is *submodular* if for all V, W in its domain $\mathbb{B}^n$

$$f(V) + f(W) \geqslant f(VW) + f(V \vee W) \tag{1}$$

[1] The results of this paper were obtained while at RUTCOR.

E-mail addresses: stephan.foldes@tut.fi, foldes@cc.tut.fi (S. Foldes), hammer@rutcor.rutgers.edu (P.L. Hammer).

0166-218X/$ - see front matter © 2003 Published by Elsevier B.V.
doi:10.1016/j.dam.2003.12.008

and it is *supermodular* if it always satisfies

$$f(V) + f(W) \leqslant f(VW) + f(V \vee W). \tag{2}$$

The role played by those pseudo-Boolean functions (or set functions, using the alternative language of finite power sets instead of $\mathbb{B}^n$) which satisfy (1) and (2) is well-known (see e.g. Fujishige [8], Lovász [11], Narayanan [12] and Topkis [13].) On the other hand, in the context of Boolean functions, the disjunctive analogue of inequality (1), i.e., where $+$ is interpreted as the $\vee$ operator in $\mathbb{B} = \{0,\ 1\}$, defines a class of Boolean functions in one-to-one correspondence with finite partial pre-orders (Ekin, Hammer and Peled [4]).

In this paper we are interested in extending to the full class of pseudo-Boolean functions, the disjunctive analogy of the functional inequalities (1) and (2), i.e. the inequalities obtained by replacing in (1) and (2) addition ($+$) by join (max, $\vee$), and to explore consequences of this formal analogy. The analogue classes so defined will be described in terms of the disjunctive normal forms of their members.

Note that the set of pseudo-Boolean functions on a given domain $\mathbb{B}^n$ is partially ordered by

$$g \leqslant f \quad g(V) \leqslant f(V), \qquad \text{for all } V \in \mathbb{B}^n.$$

This partial order is in fact a distributive lattice, the join operation of which is also called *disjunction*.

A theory of implicants and disjunctive normal forms for pseudo-Boolean functions was proposed in [5], extending the DNF theory of Boolean functions. The definition of variables $x, y, x, \ldots$ and complemented variables $\bar{x}, \bar{y}, \bar{z}, \ldots$ remains unaltered. However, in the pseudo-Boolean context, we define a *literal* as a function of the form $a + bx, b \neq 0$, where $a, b \in \mathbb{R}$ and x is any variable. The literal $a + bx$ is called *positive* or *negative* according to whether b is positive or negative. An *elementary conjunction* is a function of the form

$$a + b\tilde{x}_1 \cdots \tilde{x}_m, \tag{3}$$

where $a, b \in \mathbb{R}, b \geqslant 0$ and each $\tilde{x}_i$ is a Boolean literal (i.e. a variable or a complemented variable). An *elementary disjunction* is of the form

$$a + b(\tilde{x}_1 \vee \cdots \vee \tilde{x}_m).$$

An *implicant* of a pseudo-Boolean function f is an elementary conjunction g defined on the same domain, such that $g \leqslant f$, and g is a *prime implicant* if $g \leqslant h$ implies $g = h$ for all implicants h of f. Similarly, an *implicatum* of f is an elementary disjunction g such that $f \leqslant g$, and g is a *prime implicatum* if $h \leqslant g$ implies $h = g$ for all implicata h of f. A *disjunctive normal form* (DNF) of f is an expression of f as a join of implicants

$$f = g_1 \vee \cdots \vee g_k,$$

where all the g_i have the same minimum value. (The minimum value of an elementary conjunction (3) is a.) The *canonical DNF* of f is obtained if we take the join of all prime implicants of f (which are finite in number). For example, the DNFs

$$(1 + 2xy) \vee (1 + x\bar{y})$$

and

$$(1 + x) \vee (1 + 2xy)$$

represent the same function on $\mathbb{B}^2$, and the latter DNF is canonical.

The canonical DNF can be computed by a pseudo-Boolean *consensus algorithm* starting from any given DNF. Extending the Blake–Quine consensus procedure for Boolean functions, the following pseudo-Boolean algorithm consists in the repeated application (in any order) of the two basic steps below, until none can be applied.

Absorption : If $g_i \leqslant g_j, i \neq j$, in the DNF $g_1 \vee \cdots \vee g_k$, then remove g_i from the join expression.

Consensus : If for some $i \neq j$ in $g_1 \vee \cdots \vee g_k$

$$g_i = a + bx_1\tilde{x}_2 \cdots \tilde{x}_m,$$

$$g_j = a + c\bar{x}_1\tilde{y}_2 \cdots \tilde{y}_t,$$

$$\tilde{x}_2 \cdots \tilde{x}_m\tilde{y}_2 \cdots \tilde{y}_t \neq 0,$$

then adjoin $g = a + (b \wedge c)\tilde{x}_2 \cdots \tilde{x}_m\tilde{y}_2 \cdots \tilde{y}_t$ to $g_1 \vee \cdots \vee g_k$ to form $g_1 \vee \cdots \vee g_k \vee g$, provided that $g \nleqslant g_j$ for $j = 1, \ldots, k$.

In [5] we have defined the *dual* of a pseudo-Boolean function f as

$$f^d(V) = 1 - f(\bar{V})$$

for all $V \in \mathbb{B}^n$, where $\bar{V}$ is the complement of V in the Boolean lattice $\mathbb{B}^n$. Dualization defines a dual order automorphism of $\mathbb{R}^{(\mathbb{B}^n)}$, leaves all variables unchanged, and converts elementary conjunctions to elementary disjunctions and vice versa.

The dual notion to DNF is that of a *conjunctive normal form* (CNF): it is an expression of a pseudo-Boolean function as a meet of implicata,

$$f = g_1 \wedge \cdots \wedge g_k,$$

where all the g_i have the same maximum value. The *canonical CNF* is obtained if we take the meet of *all* prime implicata of f (which are finite in number).

Let us finally recall that the undirected covering (Hasse) diagram of the lattice $\mathbb{B}^n$ endows the domain of each pseudo-Boolean function with a graph structure. In this graph shortest paths (geodesics) between two vertices U, W play a significant role. A vertex V is *geodesically between* U and W if it lies on some geodesic connecting U and W. A set S of vertices is geodesically convex (contains all V that are geodesically between any two vertices in the set) if and only if S is a discrete *subcube*, i.e. the vertex set of some face of the solid cube spanned by $\mathbb{B}^n$ in $\mathbb{R}^n$.

The early references to pseudo-Boolean function theory are [9,10]; a recent survey appears in [1]. For Boolean functions, the DNF syntax for classes defined by certain functional inequalities and equations is studied in [3] where in particular, disjunctive Boolean analogues of submodular and supermodular functions were examined. In the full pseudo-Boolean context, the DNF syntax of monotone and Horn functions was described in [6].

The remainder of the paper is divided into two further sections. Section 2 contains results concerning characterizations of disjunctive submodular and disjunctive supermodular functions. Section 3 discusses the representation of arbitrary pseudo-Boolean functions as disjunctions of a particular type of functions studied in Section 2 (basic conjunction i.e. conjunctive modular functions).

2. Characterizations

A pseudo-Boolean function f is called *disjunctive submodular* if it satisfies the inequality $f(V) \vee f(W) \geqslant f(VW) \vee f(V \vee W)$ for any $V, W \in \mathbb{B}^n$, and it is called *disjunctive supermodular* if it satisfies $f(V) \vee f(W) \leqslant f(VW) \vee f(V \vee W)$ for any $V, W \in \mathbb{B}^n$. For brevity's sake, we shall often omit the qualifier "disjunctive" when referring to these two properties, as in the sequel we do not need to deal with the original, additive meaning of submodularity and supermodularity.

Lemma 1. *The disjunction $f_1 \vee \cdots \vee f_m$ of disjunctive submodular (resp. supermodular) functions is disjunctive submodular (resp. supermodular).*

Proof. Let $f_1, \ldots, f_m$ be submodular pseudo-Boolean functions, and let $V, W \in \mathbb{B}^n$.

$$\left[\bigvee_i f_i(V)\right] \vee \left[\bigvee_i f_i(W)\right] = \bigvee_i [f_i(V) \vee f_i(W)] \geqslant \bigvee_i [f_i(VW) \vee f_i(V \vee W)]$$

$$= \left[\bigvee_i f_i(VW)\right] \vee \left[\bigvee_i f_i(V \vee W)\right].$$

The proof for the case of supermodular functions is obtained in the same way, by simply reversing the inequalities. □

Theorem 1. *A pseudo-Boolean function f is disjunctive submodular if and only if it has a DNF in which every elementary conjunction has at most one complemented and at most one uncomplemented variable.*

Proof. A simple case analysis shows that every elementary conjunction of the stated form is submodular. Therefore any DNF of the stated form, being a disjunction of submodular functions, is submodular (by Lemma 1).

For the converse, assume that the canonical DNF of a disjunctive submodular pseudo-Boolean function contains a prime implicant of the form $a + bx_1x_2P$ or $a + b\bar{x}_1\bar{x}_2P$.

Suppose $a + bx_1x_2P$ is a prime implicant. Then neither $a + bx_1P$ nor $a + bx_2P$ is an implicant. There are then vectors V, W such that

$$v_1 = 1 \quad v_2 = 0 \quad P(V) = 1 \quad f(V) < a + b,$$

$$w_1 = 0 \quad w_2 = 1 \quad P(W) = 1 \quad f(W) < a + b.$$

But clearly $f(V \vee W) \geqslant a + b$, contradicting submodularity.

For $a + b\bar{x}_1\bar{x}_2P$ the argument is similar with $a + b\bar{x}_1P$ and $a + b\bar{x}_2P$ as non-implicants, and $f(VW) \geqslant a + b$ contradicting submodularity. □

Lemma 2. *If $h : \mathbb{R} \to \mathbb{R}$ is monotone non-decreasing, and the pseudo-Boolean function $f : \mathbb{B}^n \to \mathbb{R}$ is disjunctive supermodular, then the composition $h \circ f$ is also disjunctive supermodular.*

Proof. The result follows from the fact that $\mathbb{R}$ being a chain, h is a lattice endomorphism, i.e. for any real numbers y, z,

$$h(y \vee z) = h(y) \vee h(z). \qquad \square$$

Theorem 2. *A pseudo-Boolean function f is disjunctive supermodular if and only if it is the disjunction of a monotone non-decreasing and of a monotone non-increasing function.*

Proof. It is easy to verify that every monotone non-decreasing, as well as every monotone non-increasing pseudo-Boolean function is disjunctive supermodular. Their disjunction is then supermodular by virtue of Lemma 1.

Conversely, let us prove, by induction on the cardinality of the range of a supermodular function f, that f is the disjunction of a monotone non-increasing and of a monotone non-decreasing function.

First, the claim is obviously true if f is constant.

Second, also, if the range of f consists of two numbers a and b, $a < b$, consider the sets

$$P = \{V \in \mathbb{B}^n : \forall W \geqslant V,\ f(W) = b\},$$

$$N = \{V \in \mathbb{B}^n : \forall W \leqslant V,\ f(W) = b\}.$$

Let the n-ary pseudo-Boolean functions f_1 and f_2 be defined by

$$f_1(V) = b \text{ for } V \in P, \qquad f_1(V) = a \text{ for } V \in P,$$

$$f_2(V) = b \text{ for } V \in N, \qquad f_1(V) = a \text{ for } V \in N.$$

Clearly f_1 is monotone non-decreasing and f_2 is monotone non-increasing. Obviously $f_1 \vee f_2 \leqslant f$.

If there is a point V with $f(V) = b$ and $f_1(V) = f_2(V) = a$, then $V \in P \cup N$. It follows that there are points U, W such that $U \leqslant V \leqslant W$, and $f(U) = f(W) = a$. Let us denote by

$$V' = (U \vee \bar{V})W,$$

the relative complement of V in the interval $[U, W]$. Then $U = VV'$ and $W = V \vee V'$. Therefore, $f(VV') = f(U) = a$ and $f(V \vee V') = f(W) = a$, while $f(V) = b$, implying that $f(V) \vee f(V') = b > a$, in contradiction with the assumed supermodularity. Hence $f_1 \vee f_2 \geqslant f$.

From the conclusions of the last two paragraphs it follows that $f = f_1 \vee f_2$.

Third, if the range of f has at least 3 elements, let t be any element of the range of f such that $\min f < t < \max f$. Let h and k be functions $\mathbb{R} \to \mathbb{R}$ defined by

$$h(x) = \begin{cases} \min f & \text{if } x \leqslant t, \\ x & \text{if } t < x \end{cases}$$

and

$$k(x) = \begin{cases} x & \text{if } x \leqslant t, \\ t & \text{if } t < x. \end{cases}$$

We have $f = (h \circ f) \vee (k \circ f)$.

Both h and k are monotone non-decreasing, therefore by Lemma 2 both $h \circ f$ and $k \circ f$ are supermodular. Since both $h \circ f$ and $k \circ f$ have smaller ranges than f, the proof is completed by induction. □

An elementary disjunction was defined earlier as a disjunction of literals that are required to have the same maximum. Disjunctions of literals without this restrictive requirement are characterized by the next lemma.

Let us call a pseudo-Boolean function f *disjunctive* or *conjunctive modular* if and only if it satisfies

$$f(VW) \vee f(V \vee W) = f(V) \vee f(W) \tag{4}$$

or

$$f(VW) \wedge f(V \vee W) = f(V) \wedge f(W), \tag{5}$$

respectively, for every V, W in the domain of f. The replacement in (5) of the equality "=" by "$\leqslant$" or by "$\geqslant$" would define properties essentially dual to disjunctive submodularity and disjunctive supermodularity, respectively. Conjunctive modularity can therefore be viewed as the combination of the duals of the two main properties first considered in this section. Also, as shown by Lemmas 3, 4 and Theorems 3, 4, conjunctive modularity is the dual of the property of disjunctive modularity.

Lemma 3. *A pseudo-Boolean function f is disjunctive modular if and only if it is a disjunction of literals.*

Proof. The "if" part follows from two observations:

(i) Every literal satisfies (4).
(ii) The set of pseudo-Boolean functions satisfying (4) is closed under disjunction.

For the "only if" part, assume the validity of (4). By disjunctive submodularity, the canonical DNF of f is of the form

$$f = c_0 + \bigvee_i c_i P_i = \bigvee_i (c_0 + c_i P_i)$$

with every $c_i > 0$ and P_i of the form x, $\bar{y}$, or $x\bar{y}$. But by disjunctive supermodularity

$$f = \bigvee (c_0 + c_i P_i : P_i \text{ positive or negative})$$

and no $c_0 + c_i x\bar{y}$ can be a prime implicant. This proves the lemma. □

Lemma 4. *A pseudo-Boolean function f is conjunctive modular if and only if it is a conjunction of literals.*

Proof. The following relations are equivalent for any given vectors V, W:

$$f(VW) \wedge f(V \vee W) = f(V) \wedge f(W),$$

$$[-f(VW)] \vee [-f(V \vee W)] = [-f(V)] \vee [-f(W)],$$

$$[1 - f(VW)] \vee [1 - f(V \vee W)] = [1 - f(V)] \vee [1 - f(W)],$$

$$f^d(\overline{VW}) \vee f^d(\overline{V \vee W}) = f^d(\bar{V}) \vee f^d(\bar{W}),$$

$$f^d(\bar{V} \vee \bar{W}) \vee f^d(\bar{V}\bar{W}) = f^d(\bar{V}) \vee f^d(\bar{W}).$$

By Lemma 3 this last equation is true for all V, W if and only if f^d is a disjunction of literals. This is the case if and only if f is a conjunction of literals. □

Theorem 3 (Disjunctive modularity theorem). *For any n-ary pseudo-Boolean function f the following are equivalent*:

(a) *f satisfies $f(VW) \vee f(V \vee W) = f(V) \vee f(W)$ for all $V, W \in \mathbb{B}^n$*;
(b) *f is a disjunction of literals*;
(c) *the prime implicata of f are $a_1 + b_1 D_1, \ldots, a_m + b_m D_m$ where $a_1 < \cdots < a_m$ and $D_1 > \cdots > D_m$*;
(d) *if a vector V is geodesically between U, W (i.e. $d(U,V) + d(V,W) = d(U,W)$ in the Hamming metric) then*
$$f(V) \leqslant f(U) \vee f(W);$$
(e) *for each $t \in \mathbb{R}$ the inverse image $f^{-1}(-\infty, t]$ is a subcube of $\mathbb{B}^n$.*

Proof. The equivalence of (a) and (b) was seen in Lemma 3. The equivalence of (d) and (e) is obvious. From (d) it is easy to deduce (a). Literals obviously satisfy condition (d). Also, the join of any two functions satisfying (d) satisfies (d). Thus (b) implies (d). It follows that (a), (b), (d) and (e) are equivalent.

Let us prove now the validity of the implication of (e) by (c). First, if $t < a_1$, the set $f^{-1}(\leftarrow, t]$ being empty, the implication holds. Second, if $t \geqslant \max f$, then $f^{-1}(\leftarrow, t]$ is the whole cube $\mathbb{B}^n$. Finally, let $a_1 \leqslant t < \max f$, let i be the largest index with $a_i \leqslant t$ and let us suppose that there is a V in $f^{-1}(\leftarrow, t]$ such that $D_i(V) = 1$. In that case $D_1(V) = \cdots = D_i(V) = 1$ and

$$f(V) = \left[\bigwedge_{j \leqslant i} (a_j + b_j)\right] \wedge \left[\bigwedge_{j \geqslant i+1} (a_j + b_j D_j(V))\right]$$

and since $a_j + b_j = \max f$, it follows that $f(V) \geqslant (\max f) \wedge (\bigwedge_{j \geqslant i+1} a_j) > a_i$, in contradiction with the assumption on V. Hence, $V \in f^{-1}(\leftarrow, t]$ implies that V does not belong to the subcube of points where $D_i = 1$. On the other hand, if V is a point with $D_i(V) = 0$, as $a_i + b_i D_i$ is an implicatum of f, we have $V \in f^{-1}(\leftarrow, t]$. Hence, $f^{-1}(\leftarrow, t]$ is the subcube of points for which $D_i = 0$.

Conversely, assume (e). Let $t_1 < \cdots < t_m$ be the range of f. Let $P_i = \{V \in \mathbb{B}^n : f(V) \leqslant t_i\}$, let D_i be the Boolean elementary disjunction such that $D_i^{-1}(0) = P_i$, and let $a_i = t_i$, $b_i = t_m - t_i$. Obviously each $a_i + b_i D_i$ is an implicatum of f and $D_1 > \cdots > D_m$. To show (c), it is enough to establish that every implicatum $h = a + bD$ of f (where $b > 0$ and D is a Boolean elementary disjunction) is an implicatum of one of the $a_i + b_i D_i$. For such an implicatum $a + bD$, obviously $t_m \leqslant a + b$ and we can suppose that $a < t_m$ for otherwise the claim is obvious. From (e) it follows that $f^{-1}(-\infty, a]$ is a subcube, and in fact it must coincide with one of the P_i, namely with the P_i with the largest index i such that $t_i \leqslant a$. For this i, also $D_i \leqslant D$ and thus $a_i + b_i D_i \leqslant a + bD$. □

Remark. Condition (a) of Theorem 3 states that the dual of f is both disjunctive submodular and disjunctive supermodular.

Theorem 4 (Conjunctive modularity). *For any n-ary pseudo-Boolean function f the following are equivalent*:

(a) *f satisfies $f(VW) \wedge f(V \vee W) = f(V) \wedge f(W)$ for all $V, W \in \mathbb{B}^n$*;
(b) *f is a conjunction of literals*;
(c) *the prime implicants of f are $c_0 + c_1 P_1, \ldots, c_0 + c_m P_m$ where $c_1 < \cdots < c_m$* and $P_1 > \cdots > P_m$;
(d) *if a vector V is geodesically between U, W then*

$$f(V) \geqslant f(U) \wedge f(W);$$

(e) *for each $t \in \mathbb{R}$ the inverse image $f^{-1}[t, \infty)$ is a subcube.*

Proof. By duality from Theorem 3. □

We shall call a pseudo-Boolean function a *basic conjunction* (respectively, *disjunction*) if it is a conjunction of literals (respectively, disjunction of literals), and we shall say that it is *positive*, or *negative*, if the literals involved are all positive, or negative.

Theorem 5. *For any pseudo-Boolean function f the following are equivalent*:

(i) *f is disjunctive supermodular and f^d is disjunctive submodular*;
(ii) *f is the disjunction of a positive basic conjunction and a negative basic conjunction.*

Proof. Observe first that the submodularity of the dual means that the following equivalent inequalities are valid:

$$f^d(V \vee W) \vee f^d(VW) \leqslant f^d(V) \vee f^d(W),$$

$$-[(-f^d(V \vee W)) \wedge (-f^d(VW))] \leqslant -[(-f^d(V)) \wedge (-f^d(W))],$$

$$[(1 - f^d(V \vee W)) \wedge (1 - f^d(VW))] \geqslant [1 - f^d(V)] \wedge [1 - f^d(W)],$$

$$[1 - f^d(\bar{V} \vee \bar{W})] \wedge [1 - f^d(\bar{V}\bar{W})] \geqslant [1 - f^d(\bar{V})] \wedge [1 - f^d(\bar{W})],$$

$$[1 - f^d(\overline{VW})] \wedge [1 - f^d(\overline{V \vee W})] \geqslant [1 - f^d(\bar{V})] \wedge [1 - f^d(\bar{W})],$$

$$f(VW) \wedge f(V \vee W) \geqslant f(V) \wedge f(W). \tag{6}$$

Assume (ii). Let $f = p \vee n$, where p and n are a positive and a negative basic conjunction. Then f is supermodular by Theorem 2. For the submodularity of the dual, we need to establish (6).

Case 1: If $f(V) = p(V)$, $f(W) = p(W)$ then by Theorem 4

$$f(VW) \geqslant p(VW) \geqslant p(VW) \wedge p(V \vee W) = p(V) \wedge p(W) = f(V) \wedge f(W),$$

$$f(V \vee W) \geqslant p(V \vee W) \geqslant p(VW) \wedge p(V \vee W) = p(V) \wedge p(W) = f(V) \wedge f(W)$$

and the two inequalities imply (6).

Case 2: If $f(V) = n(V)$, $f(W) = n(W)$, Theorem 4 applied to n instead of p shows that (6) holds.

Case 3: If $f(V) = p(V)$, $f(W) = n(W)$ then

$$f(V \vee W) \geqslant p(V \vee W) \geqslant p(V) = f(V),$$

$$f(VW) \geqslant n(VW) \geqslant n(W) = f(W),$$

implying (6).

Case 4: If $f(V) = n(V)$, $f(W) = p(W)$, an argument similar to that given in Case 3 shows that (6) holds.

Conversely, assume (i). From Theorem 2 it follows that

$$f = c + (p \vee n), \tag{7}$$

where $c = \min f$, p is monotone non-decreasing with minimum 0, and n is monotone non-increasing with minimum 0. We may choose p and n maximal, in the sense that if $p' > p$ is also monotone non-decreasing with minimum 0, we can not substitute p' for p in (7), and similarly for n.

We need to show only that both p and n are basic conjunctions, positivity and negativity follow from the fact that p and n are nondecreasing and non-increasing, respectively. We do this for p, the proof is similar for n. In view of condition (c) of Theorem 4, it suffices to show that for any two prime implicants $c_0 + c_i P_i$ and $c_0 + c_j P_j$ of p, either $P_i < P_j$ or $P_j < P_i$. Note that $c_0 = 0$ because $\min p = 0$.

Suppose we had neither $P_i < P_j$ nor $P_j < P_i$. We shall derive a contradiction. We may assume that $c_i \leqslant c_j$, and since p is nondecreasing, we know that no complemented variables occur in P_i or P_j. Let P be the product of the Boolean variables occurring both in P_i and in P_j. Let V_i be the smallest element of the lattice $\mathbb{B}^n$ such that $P_i(V_i) = 1$, and let V_j be defined similarly for P_j. Then $V = V_i \wedge V_j$ is the smallest element of $\mathbb{B}^n$ for which $P(V) = 1$. By the submodularity of f^d,

$$f(V) \geqslant f(V_i) \wedge f(V_j) \geqslant c_i + c.$$

Further, observe that every W with $P(W) = 1$ can be written as $W = W_i \wedge W_j$ where $P_i(W_i) = P_j(W_j) = 1$. Therefore, by the submodularity of f^d,

$$f(W) \geqslant f(W_i) \wedge f(W_j) \geqslant c_i + c.$$

Thus $c + c_i P \leqslant f$ and since $c_i P$ is not an implicant of p ($c_i P_i$ being a prime implicant), $p' = p \vee c_i P$ would contradict the maximality of p. □

In the same vein, the following theorems characterize pseudo-Boolean functions f for which f and f^d possess other combinations of the supermodular-submodular pair of properties.

Theorem 6. *For any pseudo-Boolean function f the following are equivalent:*

(i) *both f and f^d are disjunctive supermodular*;
(ii) *f is monotone non-decreasing or monotone non-increasing.*

Proof. Assume (ii). Since the dual of any monotone function is monotone and all monotone functions are supermodular, (i) follows easily.

Conversely, assume (i). Suppose that f is not monotone. We shall derive a contradiction. Non-monotonicity implies the existence of u, v, w, t in $\mathbb{B}^n$ such that

$$U < V \quad f(U) > f(V),$$

$$W < T \quad f(W) < f(T).$$

By the supermodularity of f

$$f(0) \vee f(V) = f(U \wedge \bar{U} \wedge V) \vee f(U \vee (\bar{U} \wedge V)) \geqslant f(U) \vee f(\bar{U} \wedge V) \geqslant f(U)$$

and thus $f(0) \geqslant f(U)$, and for a similar reason $f(T) \leqslant f(1)$.

On the other hand, the supermodularity of f^d means, by an argument analogous to the one used at the beginning of the proof of Theorem 5 to characterize functions with a submodular dual, that the inequality

$$f(V \vee W) \wedge f(VW) \leqslant f(V) \wedge f(W)$$

holds for all X, Y in $\mathbb{B}^n$. Thus in particular

$$f(T) \wedge f(0) = f(W \vee \bar{W}T) \wedge f(W\bar{W}T) \leqslant f(W) \wedge f(\bar{W}T) \leqslant f(W)$$

and therefore $f(0) \leqslant f(W)$, and for a similar reason $f(V) \geqslant f(1)$.

We would now have simultaneously

$$f(0) \geqslant f(U) > f(V) \geqslant f(1),$$

$$f(0) \leqslant f(W) < f(T) \leqslant f(1).$$

and this is impossible. □

The next two theorems will characterize those submodular functions for which the dual function is supermodular or submodular, respectively.

Theorem 7. *For a pseudo-Boolean function f the following conditions are equivalent*:

(i) *f is disjunctive submodular and f^d is disjunctive supermodular*;
(ii) *f has a DNF*

$$(a + bq^+q^-) \vee (a + cr^+r^-) \vee \ldots,$$

where $q^+, r^+, \ldots$ are positive literals or the constant 1, *$q^-, r^-, \ldots$ are negative literals or the constant* 1, *and for any two terms* $a + bq^+q^-$ and $a + cr^+r^-$ in the DNF, $a + \min(b, c)q^+r^-$ *is an implicant of* f.

Proof. Note first that the supermodularity of f^d means that for all $U, W \in \mathbb{B}^n$

$$f(UW) \wedge f(U \vee W) \leqslant f(U) \wedge f(W)$$

which is true if and only if for all $U, V \in \mathbb{B}^n$

$$f(UW) \wedge f(U \vee W) \leqslant f(U) \tag{8}$$

or, equivalently, if $U \leqslant V \leqslant W$ in $\mathbb{B}^n$ implies

$$f(U) \wedge f(W) \leqslant f(V). \tag{9}$$

Assume (i). We shall prove that the canonical DNF of f satisfies the requirements of (ii). From Theorem 1 it follows without difficulty that this canonical DNF is written as

$$(a + bq^+q^-) \vee (a + cr^+r^-) \vee \ldots,$$

where $q^+, r^+, \ldots$ are positive literals or the constant 1 function, $q^-, r^-, \ldots$ are negative literals or the constant 1 function. Consider now two generic terms $a + bq^+q^-$ and $a + cr^+r^-$ of this DNF. Let us denote the function $a + \min(b, c)q^+r^-$ by I. We need to show that $I \leqslant f$. If this were not true, there would exist a $V = (v_1, \ldots, v_n)$ in $\mathbb{B}^n$ such that $I(V) > f(V)$. To derive a contradiction, assume that this is the case. Clearly, $I(V) = a + \min(b, c)$ and thus

$$q^+(V) = r^-(V) = 1$$

but both implicants $a + bq^+q^-$ and $a + cr^+r^-$ must have the value a on V, i.e.

$$q^+q^-(V) = q^+(V)q^-(V) = 0,$$

$$r^+r^-(V) = r^+(V)r^-(V) = 0$$

and it now follows that

$$q^-(V) = r^+(V) = 0.$$

This is possible if and only if q^- is a negative literal $\bar{x}_i$ (the negation of the ith variable) and r^+ is a positive literal x_j (the jth variable). For the corresponding components of V this implies $v_i = 1$, $v_j = 0$. Define now the vector U in $\mathbb{B}^n$ as having the same components as V, except that $u_i = 0$, and define W to have the same components as V except that $w_j = 1$. Obviously $U \leqslant V \leqslant W$. However, the implicant $a + bq^+q^-$ takes value $a + b$ on U, and $a + cr^+r^-$ takes value $a + c$ on W. Thus

$$f(U) \geqslant a + b,$$

$$f(W) \geqslant a + c,$$

$$f(U) \wedge f(W) \geqslant a + \min(b, c).$$

But from the assumption $I(V) > f(V)$ it would now follow (since $I(V) = a + \min(b, c)$) that $f(U) \wedge f(W) > f(V)$, contradicting (9). Condition (ii) is thus proved from (i).

Conversely, assume we have a DNF of f as described by (ii). By Theorem 1, f is submodular. If f^d were not supermodular, then (8) would fail for some U, W in $\mathbb{B}^n$, i.e. we would have

$$f(UW) \wedge f(U \vee W)) > f(U). \tag{10}$$

The value $f(UW)$ would be the value on UW of some implicant $a + bq^+q^-$ of the given DNF, and $f(U \vee W)$ would also be the value on $U \vee W$ of some implicant $a + cr^+r^-$, i.e.

$$f(UW) = a + b,$$

$$q^+(UW) = 1,$$

$$f(U \vee W) = a + c,$$

$$r^-(U \vee W) = 1$$

and consequently

$$q^+(U) = 1 \qquad r^-(U) = 1,$$

$$f(U) \geqslant [a + \min(b, c)q^+r^-](U) = a + \min(b, c).$$

But this contradicts (10) because

$$a + \min(b, c) = f(UW) \wedge f(U \vee W).$$

Thus f^d must be submodular. □

In a finite Boolean lattice L, often thought of as a "discrete cube", a *half cube* is an order interval either of the form $[a, \max L]$ where a is an atom of L, or of the form $[\min L, \beta]$ where β is a coatom of L. For any $e \in L$, the complement of e in L is denoted by $\bar{e}$.

Lemma 5. *Let L_1 and L_2 be disjoint non-empty sublattices of a finite Boolean lattice L such that $L_1 \cup L_2 = L$. If $\min L$ and $\max L$ belong to different L_i then L_1 and L_2 are complementary half cubes of L, and if both belong to L_1 then L_1 is the union of two half cubes of L.*

Proof. We prove the Lemma by induction on the dimension of L. The statement is easy to verify for dimension 1.

Let L have dimension n greater than 1, and suppose that the statement is true for dimension $n - 1$. If $\min L \in L_1$ and $\max L \in L_2$, then it is easy to see that every atom of L except one (say a) belongs to L_1 and every coatom except one (say β) belongs to L_2. It follows that

$$L_1 = [\min L, \bar{a}],$$

$$L_2 = [\bar{\beta}, \max L].$$

If $\min L$ and $\max L$ belong to L_1, then it is still true that there is an atom a such that every atom except a belongs to L_1 and

$$[\min L, \bar{a}] \subseteq L_1,$$

$$L_2 \subseteq [a, \max L].$$

Now $L' = [a, \max L]$ is a Boolean lattice of dimension $n-1$, $L_1' = L_1 \cap L'$ and $L_2' = L_2 \cap L' = L_2$ are non-empty lattices partitioning L', $\min L' = a$ belongs to L_2' and $\max L' = \max L$ belongs to L_1'. By the induction by hypothesis L_1' and L_2' are complementary half cubes in L', and for some coatom β of L'

$$L_2' = [a, \beta].$$

But $L_2' = L_2$, β is also a coatom of L, and

$$L_1 = [\min L, \bar{a}] \cup [\bar{\beta}, \max L] \qquad \square$$

The next result is specific to Boolean functions.

Theorem 8. *For a Boolean function f the following conditions are equivalent*:

(i) *both f and f^d are disjunctive submodular*;
(ii) *f has a DNF with at most one complemented and at most one uncomplemented variable occurrence*;
(iii) *f is constant*, *or it is a literal*, *or it has a DNF of the form $x \vee \bar{y}$ or $x\bar{y}$.*

Proof. The equivalence of (ii) and (iii) is easy to verify, and so is the implication (iii) $\Rightarrow$ (i).

Note that (i) is equivalent to saying that both the true and the false points of f constitute sublattices of $\mathbb{B}^n$. Then (iii) follows by Lemma 5. $\square$

Theorem 9. *For a pseudo-Boolean function f the following two conditions are equivalent*:

(i) *both f and f^d are disjunctive submodular*;
(ii) *f is of the form $k + (ax \vee b\bar{y} \vee cx\bar{y})$ when k, a, b, c are constants and $a, b, c \geqslant 0$.*

Proof. Note that (i) holds if and only if both of the following inequalities hold for all $U, W \in \mathbb{B}^n$

$$f(UW) \vee f(U \vee W) \leqslant f(U) \vee f(W), \tag{11}$$

$$f(UW) \wedge f(U \vee W) \geqslant f(U) \wedge f(W). \tag{12}$$

If (ii) holds then we may actually suppose $a < c, b < c$. A simple case analysis shows that (11) and (12) hold for all U, W.

Suppose, conversely, that (11) and (12) hold for all U, W. If f is constant, then clearly (ii) holds. Otherwise let us enumerate the range of f in increasing order, say

$$r_1 < \cdots < r_m, \quad m \geqslant 2.$$

For each $i = 1, \ldots, m-1$ let f_i be the Boolean function defined by

$$f_i(V) = 0 \Leftrightarrow f(V) \leqslant r_i.$$

Then for each i, f_i is non-constant, submodular, and f_i^d is also submodular. (This can be verified using inequalities like (11) and (12) with f_i instead of f.) Obviously,

$$f_1 > \cdots > f_{m-1}.$$

It follows from Theorem 8 that $m - 1 \leqslant 3$.

If $m - 1 = 1$ then

$$f = r_1 + (r_2 - r_1) f_1$$

and (ii) holds obviously.

If $m-1=2$ then we have one of the following two cases:

(a) f_1 is of the form $x \vee \bar{y}$ and f_2 is either x or $\bar{y}$ or $x\bar{y}$;
(b) f_2 is of the form $x\bar{y}$ and f_1 is either x or $\bar{y}$.

In either case

$$f = r_1 + (r_2 - r_1)f_1 + (r_3 - r_2)f_2 = r_1 + [(r_2 - r_1)f_1 \vee (r_3 - r_1)f_2]$$

and (ii) holds obviously.

Finally, if $m-1=3$, then for some variables x, y

$$f_1 = x \vee \bar{y},\ \ f_2 = x \text{ or } f_2 = \bar{y}\ \ f_3 = x\bar{y},$$

$$f = r_1 + [(r_2 - r_1)f_1 \vee (r_3 - r_1)f_2 \vee (r_4 - r_1)f_3]$$

and (iii) holds obviously. □

3. Implicant theory of basic conjunctions

Basic conjunctions generalize the notion of elementary conjunction, since every elementary conjunction is a basic conjunction and every pseudo-Boolean function is the join of a finite number of basic conjunctions. Within the class of Boolean functions, elementary and basic conjunctions coincide. In the class of pseudo-Boolean functions, however, the set of elementary conjunctions is not closed under meet. In fact the closure under meet of the set of pseudo-Boolean literals is the entire class of basic conjunctions.

We shall now outline a disjunctive representation theory of pseudo-Boolean functions based on basic rather than elementary conjunctions. Our terminology is justified within the framework of abstract lattice-theoretical implicant theory, proposed by Davio, Deschamps and Thayse [2] (see also [7]).

Definition. A basic conjunction g is a *basic implicant* of a pseudo-Boolean function f if $g \leqslant f$; it is a *prime basic implicant* of f if it is maximal in the set of basic implicants of f.

Observe that if $(c_0+c_1P_1)\vee\cdots\vee(c_0+c_mP_m)$ is a basic implicant of f, then each $c_0+c_iP_i$ is an (elementary) implicant (and conversely, assuming $c_1 < \cdots < c_m$, $P_1 > \cdots > P_m$, as specified in condition (c) of Theorem 4).

Definition. The *nesting order* $\preceq$ is the partial order on the set of all elementary conjunctions given by

$$a + bP \preceq a' + b'P' \qquad a \leqslant a',\ b \leqslant b',\ P \geqslant P'.$$

From Theorem 4 it is clear that a pseudo-Boolean function is a basic conjunction if and only if it has a DNF $\bigvee_i(a_i+b_iP_i)$ where the various $a_i + b_iP_i$ form a chain in the nesting order. (They then form an antichain in the standard order on $\mathbb{R}^{\mathbb{B}^n}$.) A basic conjunction $g = (c_0 + c_1P_1) \vee \cdots \vee (c_0 + c_mP_m)$, $c_1 < \cdots < c_m$, $P_1 > \cdots > P_m$ is a prime basic implicant of a pseudo-Boolean function f if and only if $\{c_0 + c_iP_i : 1 \leqslant i \leqslant m\}$ is a maximal chain in the nesting-ordered set of (elementary) prime implicants of f.

It follows that, similarly to (elementary) prime implicants, every pseudo-Boolean function has only a finite number of prime basic implicants.

We now define a *consensus algorithm for basic implicants*. Let a pseudo-Boolean function f be expressed as a join of some basic implicants

$$f = g_1 \vee \cdots \vee g_m. \tag{13}$$

The algorithm performs repeatedly any of the following two procedures until none can be performed:

Absorption: If $g_i < g_j$, delete g_i from (13).

Adjunction of consensus: If for some prime basic implicant g of $g_i \vee g_j$ there is no g_k with $g \leqslant g_k$, then add g to (13).

Note. In adjunction of consensus, for a given pair g_i, g_j, there can be more than one g satisfying the stated condition that there is no $g_k \geqslant g$. E.g.

$$f = (x \vee 2xyz) \vee (y \vee 2xyt)$$

$$= [(2 - 2\bar{x}) \wedge (2 - \bar{y}) \wedge (2 - \bar{z})] \vee [(2 - 2\bar{y}) \wedge (2 - \bar{x}) \wedge (2 - \bar{t})].$$

Theorem 10. *The consensus algorithm for basic implicants terminates always with precisely the join of all the prime basic implicants of f.*

Proof. It suffices to prove that if none of the two procedures of the algorithm can be performed for (13), then for every basic implicant g of f there is a g_i such that $g \leqslant g_i$. In fact it is enough to show this for those g whose range is contained in that of f because for every basic implicant

$$g = c_0 + (c_1 P_1 \vee \cdots \vee c_m P_m),$$

$$c_1 < \cdots < c_m, \quad P_1 > \cdots > P_m,$$

we can define, for $1 \leqslant i \leqslant m$, c_i' as the smallest number greater than or equal to c_i such that $c_0 + c_i'$ is in the range of f, and then

$$g' = c_0 + (c_1' P_1 \vee \cdots \vee c_m' P_m)$$

is also a basic implicant of f, and $g \leqslant g'$, and the range of g' is contained in that of f.

So let $\mathscr{G}$ denote the set of those basic implicants of f whose range is contained in that of f: this set $\mathscr{G}$ is finite and it is endowed with the order inherited from $\mathbb{R}^{\mathbb{B}^n}$. If the Theorem fails, there is a minimal member g of $\mathscr{G}$ for which there is no g_i such that $g \leqslant g_i$. We shall derive a contradiction.

It is easy to see that g cannot be of the form $c_0 + cP$, $c > 0, P$ a Boolean minterm. Let $g = (c_0 + c_1 P_1) \vee \cdots \vee (c_0 + c_k P_k)$, where $0 < c_1 < \cdots < c_k$, $P_1 > \cdots > P_k$ Boolean elementary conjunctions.

If $k = 1$, then P_1 is not a minterm; and if $k \geqslant 2$, then $P_1 > P_2$ implies that P_1 is not a minterm. Let x be a variable not occurring in P_1. Let

$$g^+ = (c_0 + c_1 P_1 x) \vee \cdots \vee (c_0 + c_k P_k x),$$

$$g^- = (c_0 + c_1 P_1 \bar{x}) \vee \cdots \vee (c_0 + c_k P_k \bar{x}).$$

For some g_i, g_j, $g^+ \leqslant g_i, g^- \leqslant g_j$ because $g^+ < g$, $g^- < g$. But $g = g^+ \vee g^-$ and thus g could be added to (13) by adjunction of consensus: contradiction. □

Theorem 11. *Let a basic conjunction p be represented as a meet of literals*

$$\bigwedge_i (a_i + b_i \tilde{x}_i) \tag{14}$$

with all $b_i > 0$, so that no variable occurs in more than one factor literal in (14) and no factor $a_i + b_i \tilde{x}_i$ is redundant.

(a) *p is an elementary conjunction if and only if all the factors have the same minimum,*
(b) *expression (14) is the canonical CNF of p if and only if all the factors have the same maximum.*

Proof. (a) If all the factors have the same minimum, then (14) is by definition an elementary conjunction. Conversely, if all the factors do not have the same minimum, then let a_1 be the lowest of these minima, and let a_2 be the second lowest. We have $a_i + b_i > a_2$ for every i, and thus the range of the function p includes at least the three distinct numbers

$$a_1, a_2, \bigwedge_i (a_i + b_i).$$

Thus (14) cannot be an elementary conjunction if all the literals in (14) do not have the same minimum.

(b) If all the factors in (14) have the same maximum, then (14) is easily seen to be the canonical CNF of p. Conversely, (14) is a CNF by definition only if $a_i + b_i$ are the same for all i. □

The following theorem then follows by duality:

Theorem 12. *Let a basic disjunction d be represented as a join of literals*

$$\bigvee_i (a_i + b_i \tilde{x}_i) \tag{15}$$

so that no variable occurs in more than one join term of (15) and no join term is redundant.

(a) *d is an elementary disjunction if and only if all the join terms have the same maximum,*
(b) *expression (15) is the canonical DNF of d if and only if all the join terms have the same minimum.*

Acknowledgements

The results of this paper were obtained while the first named author was at RUTCOR. The partial support of the National Science Foundation (Grant DMS-9806389) and of the Office of Naval Research (Grant N00014-92-J-1375) are gratefully acknowledged.

References

[1] E. Boros, P.L. Hammer, Pseudo-Boolean optimization—a survey, Discrete Appl. Math. 123 (2002) 155–225.
[2] M. Davio, J.-P. Deschamps, A. Thayse, Discrete and Switching Functions, McGraw-Hill, New York, 1978.
[3] O. Ekin, S. Foldes, P.O. Hammer, L. Hellerstein, Equational characterizations of Boolean function classes, Discrete Math. 211 (2000) 27–51.
[4] O. Ekin, P.L. Hammer, U. Peled, Horn functions and submodular Boolean functions, Theoret. Comput. Sci. 175 (1997) 257–270.
[5] S. Foldes, P.L. Hammer, Disjunctive and conjunctive normal forms of pseudo-Boolean functions, Discrete Appl. Math. 107 (2000) 1–26.
[6] S. Foldes, P.L. Hammer, Monotone, Horn and quadratic pseudo-Boolean functions, J. Univ. Comput. Sci. 6-1 (2000) 97–104, http://www.iicm.edu/jucs_6_1.
[7] S. Foldes, P.L. Hammer, Disjunctive and conjunctive forms in finite lattices and convexity spaces, Discrete Math. 258 (2002) 13–25.
[8] S. Fujishige, Submodular Functions and Optimization, Annals of Discrete Mathematics, Vol. 47, North-Holland, Amsterdam, New York, Oxford, Tokyo, 1991.
[9] P.L. Hammer, I. Rosenberg, S. Rudeanu, On the minimization of pseudo-Boolean functions, Stud. Cerc. Matn. 14 (1963) 359–364.
[10] P.L. Hammer, S. Rudeanu, Boolean Methods in Operations Research and Related Areas, Springer, Berlin, 1968.
[11] L. Lovász, Submodular functions and convexity, in: A. Bachem, M. Groetschel, B. Korte (Eds.), Mathematical Programming—The State of the Art, Springer, Berlin, 1983.
[12] H. Narayanan, Submodular Functions and Electrical Networks, Annals of Discrete Mathematics, Vol. 54, North-Holland, Amsterdam, New York, Oxford, Tokyo, 1997.
[13] D.M. Topkis, Supermodularity and Complementarity, Princeton University Press, Princeton, NJ, 1998.