ZOR – Methods and Models of Operations Research (1992) 36:3–21

# Approximations of Pseudo-Boolean Functions; Applications to Game Theory[1]

By P.L. Hammer[2] and R. Holzman[3]

*Abstract:* This paper studies the approximation of pseudo-Boolean functions by linear functions and more generally by functions of (at most) a specified degree. Here a pseudo-Boolean function means a real valued function defined on $\{0, 1\}^n$, and its degree is that of the unique multilinear polynomial that expresses it; linear functions are those of degree at most one. The approximation consists in choosing among all linear functions the one which is closest to a given function, where distance is measured by the Euclidean metric on $\boldsymbol{R}^{2^n}$. A characterization of the best linear approximation is obtained in terms of the average value of the function and its first derivatives. This leads to an explicit formula for computing the approximation from the polynomial expression of the given function. These results are later generalized to handle approximations of higher degrees, and further results are obtained regarding the interaction of approximations of different degrees. For the linear case, a certain constrained version of the approximation problem is also studied. Special attention is given to some important properties of pseudo-Boolean functions and the extent to which they are preserved in the approximation. A separate section points out the relevance of linear approximations to game theory and shows that the well known Banzhaf power index and Shapley value are obtained as best linear approximations of the game (each in a suitably defined sense).

*Zusammenfassung:* In dieser Arbeit wird die Approximation Pseudo-Boole'scher Funktionen durch lineare Funktionen bzw. allgemeiner durch Funktion kleiner/gleich eines festen Grades studiert. Dabei ist eine Pseudo-Boole'sche Funktion eine reellwertige Funktion definiert auf $\{0, 1\}^n$ und ihr Grad ist jener des eindeutig bestimmten multilinearen Polynoms, durch die sie dargestellt werden kann. Lineare Funktionen sind jene mit einem Grad kleiner oder gleich Eins. Die Approximation besteht darin, daß unter allen linearen Funktionen jene mit dem kleinsten Euklidischen Abstand in $R^{2^n}$ gewählt wird. Es wird eine Charakterisierung der besten linearen Approximation durch den Mittelwert der Funktion und ihrer ersten Ableitungen angegeben. Sie führt auf eine explizite Formel, um die Approximation aus der Polynomform gegebenen Funktion zu berechnen. Diese Ergebnisse werden später verallgemeinert, um Approximationen höheren Grades zu behandeln. Ferner werden Ergebnisse bezüglich des Zusammenhanges von Approximationen verschiedenen Grades gewonnen. Im linearen Fall wird auch eine im gewissen Sinne eingeschränkte Approximationsaufgabe behandelt. Spezielle Betrachtung wird jenen wichtigen Eigenschaften Pseudo-Boole'scher Funktionen gezollt, die bei der

---

[1] Supported by the Air Force Office of Scientific Research (under grant number AFOSR 89-0512 and AFOSR 90-0008 to Rutgers University), as well as the National Science Foundation (under grant number DMS 89-06870).
[2] Peter L. Hammer, RUTCOR, Rutgers Center for Operations Research, Rutgers University, New Brunswick, NJ 08903, USA.
[3] Ron Holzman, Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science, 76100 Rehovot, Israel.

Approximation erhalten bleiben. Ein weiterer Abschnitt zeigt die Relevanz linearer Approximationen in der Spieltheorie auf und zeigt Verbindungen zwischen den hier erzielten Ergebnissen und dem wohlbekannten Banzhaf Index auf.

*Key words:* pseudo-Boolean functions, linear approximations, least squares, best $k$th approximations, power indices.

## 1 Introduction

When a function describing some complicated relationship is given, one often seeks to replace it by a simpler functional form, usually linear, which approximates the given one. Examples of this classical theme are: (i) the local approximation of a differentiable function of several real variables by a linear function, and (ii) the global approximation of statistical data by a linear relationship (regression analysis).

Here we intend to study this sort of operation for *pseudo-Boolean functions*; these are functions $f: B^n \to \boldsymbol{R}$, where $B^n$ is the $n$-fold product of $B = \{0, 1\}$ and $\boldsymbol{R}$ is the set of real numbers. Such functions can express statistical data when the explanatory variables are Boolean while the explained variable is real. Many $0-1$ optimization problems admit natural formulations in terms of pseudo-Boolean functions. Furthermore, the main objects of cooperative game theory are pseudo-Boolean functions. Thus, these functions are a natural topic of study.

Any pseudo-Boolean function has a unique expression as a multilinear polynomial in $n$ variables:

$$f(x) = \sum_{T \subseteq N} \left[ a_T \prod_{i \in T} x_i \right] . \tag{1.1}$$

Here $N = \{1, \ldots, n\}$, the $a_T$ are real coefficients and $x = (x_1, \ldots, x_n) \in B^n$. (See Hammer and Rudeanu (1968) for the derivation of this expression and more on pseudo-Boolean functions.) We let deg $(f)$ denote the degree of the polynomial (1.1). A pseudo-Boolean function $f$ is *linear* if deg $(f) \leq 1$, i.e., if it can be written in the form

$$f(x) = a_0 + \sum_{i \in N} a_i x_i . \tag{1.2}$$

(We write $a_0$ for $a_\emptyset$ and $a_i$ for $a_{\{i\}}$.)

Our purpose is to approximate an arbitrary pseudo-Boolean function by a linear one. Following common practice in regression analysis, we shall use the least squares criterion to choose a best approximation. In other words, we shall approximate a function by the linear function that is closest to it in the Euclidean metric on $\boldsymbol{R}^{2n}$ (identifying a function with the vector listing its $2^n$ values).

The main feature that distinguishes our approach from standard regression analysis is that we regard the polynomial expression (1.1) as the canonical description of the function, rather than referring to a table description. Thus, while the definition of the best approximation refers to the $2^n$ values $f(x)$, we are ultimately interested in finding an expression of the best approximation in terms of the coefficients $a_T$ of the given function. This will permit a quick computation of the approximation when all but a few of these coefficients vanish (which is the case in many applications). Another benefit of focusing on the polynomial expression is that it suggests looking at approximations of degree higher than one. For instance, we may (and will) consider the best approximation of an arbitrary pseudo-Boolean function by a quadratic one.

Linear approximations will be studied in Section 2, and those of higher degrees in Section 4. Meanwhile, in Section 3, we shall interpret linear approximations in the context of game theory. It will be shown how they can offer new insights into known solution concepts as well as suggest new solution concepts.

The literature on pseudo-Boolean functions has already dealt with the association of linear functions with given functions of higher degree. In particular, we have in mind the roof duality concept of Hammer et al. (1984). That theory differs from the one we develop here in at least two ways: it considers approximations by majorants only, and it uses the least sum of deviations criterion (rather than least squares). Both differences reflect a basic distinction of objectives: roof duality is geared towards the solution of maximization problems, whereas the present study aims at an unbiased approximation of a function.

The game theoretic literature has dealt extensively with the association of payoff vectors with cooperative games. Yet the approach discussed here, namely the identification of payoff vectors with linear games and the minimization of distance to the given game, has received little attention. We refer to Charnes et al. (1985) and some works mentioned there for previous attempts along such lines.

Finally we mention another body of literature that is relevant to our work. This is the study of switching functions in the context of the design of electrical circuits. Essentially, switching functions are just Boolean functions $f: B^n \to B$ (thus, a subclass of pseudo-Boolean functions), but sometimes the set $B = \{0, 1\}$ is replaced in the domain and/or the range by $\{-1, +1\}$. Using one of these formalizations an expression for the coefficients of the best linear approximation (by the least squares criterion) was obtained in Coleman (1961). It was pointed out in Kaplan and Winder (1965) that these coefficients are identical to a set of parameters of the switching function, known as the Chow parameters. This observation is essentially equivalent to our Corollary 3.1, concerning the Banzhaf power index in game theory.

## 2 Linear Approximations: Theory and Computation

*Definition 2.1:* Let $f: B^n \to \boldsymbol{R}$ be a pseudo-Boolean function. The best linear approximation of $f$ is the linear function $l: B^n \to \boldsymbol{R}$ which minimizes $\sum_{x \in B^n} [f(x) - l(x)]^2$ among all linear functions. We write $l = A(f)$.

Existence and uniqueness of the best linear approximation follow from the theory of orthogonal projections in Euclidean spaces. Indeed, if we identify pseudo-Boolean functions with vectors in $\boldsymbol{R}^{2n}$ (assuming a fixed ordering of the $2^n$ elements of $B^n$) and denote by $L$ the subspace of $\boldsymbol{R}^{2n}$ corresponding to linear pseudo-Boolean functions, then $A$ is the orthogonal projection onto $L$. It follows also that $A$ is a linear operator, i.e.,

$$A(\lambda f + \mu g) = \lambda A(f) + \mu A(g) \tag{2.1}$$

for all pseudo-Boolean functions $f$ and $g$ and all real numbers $\lambda$ and $\mu$. In particular, $A$ is covariant with respect to addition of linear functions: if $l$ is linear then

$$A(f+l) = A(f) + A(l) = A(f) + l \ . \tag{2.2}$$

We proceed now to characterize the best linear approximation of a function. We shall use repeatedly the following well known fact.

*Lemma 2.2:* Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be a sequence of real numbers. Then the expression $\sum_{t=1}^{m} (\lambda_t - x)^2$ is minimized uniquely by $x = \text{Ave}\{\lambda_t\}$, i.e., the arithmetical average of $\lambda_1, \ldots, \lambda_m$.

Let $l = A(f)$. Since $l + a$ is a linear function for all real $a$, the expression $\sum_{x \in B^n} [f(x) - l(x) - a]^2$ is minimized by $a = 0$. Hence by the lemma $0 = \text{Ave}\{f(x) - l(x) : x \in B^n\}$, or equivalently

$$\text{Ave}\{l(x) : x \in B^n\} = \text{Ave}\{f(x) : x \in B^n\} \ . \tag{2.3}$$

Next, let $i \in N$. Since $l + a x_i$ is a linear function for all real $a$, the expression $\sum_{x \in B^n} [f(x) - l(x) - a x_i]^2$ is minimized by $a = 0$. But $a$ affects only the terms where $x_i = 1$, so $\sum_{x_i = 1} [f(x) - l(x) - a]^2$ is minimized by $a = 0$. Hence by the lemma $0 = \text{Ave}\{f(x) - l(x) : x_i = 1\}$, or equivalently

$$\text{Ave}\{l(x) : x_i = 1\} = \text{Ave}\{f(x) : x_i = 1\} \ . \tag{2.4}$$

From (2.3) and (2.4) it follows that

$$\operatorname{Ave}\{l(x):x_i=0\}=\operatorname{Ave}\{f(x):x_i=0\}\ . \tag{2.5}$$

We recall that the (first) derivative of a pseudo-Boolean function $f$ with respect to $x_i$ at the point $x\in B^n$ is defined as

$$\Delta_i f(x)=f(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n)-f(x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_n)\ . \tag{2.6}$$

Note that $\Delta_i f$ depends only on the components $x_j$ for $j\neq i$, but we still regard it as a function on $B^n$. From (2.4) and (2.5) it follows that

$$\operatorname{Ave}\{\Delta_i l(x):x\in B^n\}=\operatorname{Ave}\{\Delta_i f(x):x\in B^n\}\ . \tag{2.7}$$

For $l(x)=a_0+\sum_{i\in N}a_i x_i$ we have $\Delta_i l\equiv a_i$, $i\in N$, so that the Equations (2.7) determine $a_i$, $i\in N$. Once these are determined, (2.3) determines $a_0$. Thus we have the following.

*Theorem 2.3:* Given a pseudo-Boolean function $f: B^n\to \boldsymbol{R}$, its best linear approximation is characterized as the unique linear function $l: B^n\to\boldsymbol{R}$ that agrees with $f$ in average value and in average first derivatives (as in (2.3) and (2.7)).

This result can be interpreted by analogy with the approximation of a function $f:\boldsymbol{R}^n\to\boldsymbol{R}$ by a hyperplane in the neighborhood of a point $x\in \boldsymbol{R}^n$ where $f$ is differentiable. The latter is obtained as the unique linear function that agrees with $f$ in value and in first derivatives at $x$. Our approximations have a similar characterization, except that agreement holds on average rather than at a point, which is due to the fact that we are doing global, not local, approximation.

Let us now use the characterization to derive an explicit formula for the approximation. Suppose first that $f$ is of the form $f(x)=\prod_{i\in T}x_i$. Then for $i\in T$, $\Delta_i f(x)=\prod_{j\in T\setminus\{i\}}x_j$ and therefore $\operatorname{Ave}\{\Delta_i f(x):x\in B^n\}=\frac{1}{2^{|T|-1}}$; for all $i\in N\setminus T$, $\Delta_i f(x)\equiv 0$. Thus we conclude from (2.7) that $l(x)=a_0+\frac{1}{2^{|T|-1}}\times\sum_{i\in T}x_i$, with $a_0$ to be determined. Since $\operatorname{Ave}\{l(x):x\in B^n\}=a_0+\frac{|T|}{2^{|T|}}$ and $\operatorname{Ave}\{f(x):x\in B^n\}=\frac{1}{2^{|T|}}$, it follows from (2.3) that $a_0=-\frac{|T|-1}{2^{|T|}}$. Summing up, we have shown that

$$f(x)=\prod_{i\in T} x_i \Rightarrow (Af)(x) = -\frac{|T|-1}{2^{|T|}}+\frac{1}{2^{|T|-1}}\sum_{i\in T} x_i \ . \tag{2.8}$$

The formula for general $f$ follows from this by linearity of $A$, namely:

$$\begin{aligned} f(x) &= \sum_{T\subseteq N}\left[a_T\prod_{i\in T} x_i\right] \Rightarrow (Af)(x) \\ &= -\sum_{T\subseteq N}\frac{a_T(|T|-1)}{2^{|T|}}+\sum_{i\in N}\left[\sum_{T\, i\in T}\frac{a_T}{2^{|T|-1}}\right]x_i \ . \end{aligned} \tag{2.9}$$

*Example:* Let $f: B^5 \to \boldsymbol{R}$ be given by

$$f(x)=8-x_1+5x_2-x_1x_5+4x_3x_5-6x_2x_4x_5+2x_1x_2x_3x_4 \ . \tag{2.10}$$

Using (2.8), we approximate

| | | |
|---|---|---|
| $x_1x_5$ | by | $-\frac{1}{4}+\frac{1}{2}(x_1+x_5)$ , |
| $x_3x_5$ | by | $-\frac{1}{4}+\frac{1}{2}(x_3+x_5)$ , |
| $x_2x_4x_5$ | by | $-\frac{1}{4}+\frac{1}{4}(x_2+x_4+x_5)$ , |
| $x_1x_2x_3x_4$ | by | $-\frac{3}{16}+\frac{1}{8}(x_1+x_2+x_3+x_4)$ . |

The linear part of $f$ is its own approximation. Putting things together and using linearity we get:

$$\begin{aligned} (Af)(x) &= 8-x_1+5x_2-\left[-\frac{1}{4}+\frac{1}{2}(x_1+x_5)\right]+4\left[-\frac{1}{4}+\frac{1}{2}(x_3+x_5)\right] \\ &\quad -6\left[-\frac{1}{4}+\frac{1}{4}(x_2+x_4+x_5)\right]+2\left[-\frac{3}{16}+\frac{1}{8}(x_1+x_2+x_3+x_4)\right] \\ &= \frac{67}{8}-\frac{5}{4}x_1+\frac{15}{4}x_2+\frac{9}{4}x_3-\frac{5}{4}x_4 \ . \end{aligned} \tag{2.11}$$

It is often the case that a pseudo-Boolean function is given by an expression that involves also complemented variables. That is, the function $f: B^n \to R$ is given as a sum of terms of the form $a_{S,T} \prod_{i \in S} x_i \prod_{j \in T} \bar{x}_j$, where $S \subseteq N$ and $T \subseteq N$ are disjoint and $\bar{x}_j = 1 - x_j$. It is of course possible to substitute $1 - x_j$ for $\bar{x}_j$ and rearrange terms so as to obtain the canonical expression (1.1). In order to compute the approximation, however, we need not go through the canonical expression. Instead, we may compute the approximation of such a mixed term as though none of the variables were complemented, and then complement those variables in the linear expression which are complemented in the original term. This procedure is valid because the complementation of a subset of the variables preserves the distance between functions.

*Example:* Let $g: B^5 \to R$ be given by

$$g(x) = 8 - \bar{x}_1 + 5x_2 - x_1\bar{x}_5 + 4\bar{x}_3\bar{x}_5 - 6\bar{x}_2 x_4 x_5 + 2x_1\bar{x}_2 x_3 \bar{x}_4 \ .$$

Observing that, upon ignoring complementation, $g(x)$ is the same as $f(x)$ of (2.10), we can use the term-by-term approximations obtained there and carry over the complementation. Thus, we get:

$$\begin{aligned}(Ag)(x) = 8 - \bar{x}_1 + 5x_2 - &\left[ -\frac{1}{4} + \frac{1}{2}(x_1 + \bar{x}_5) \right] + 4\left[ -\frac{1}{4} + \frac{1}{2}(\bar{x}_3 + \bar{x}_5) \right] \\ &-6\left[ -\frac{1}{4} + \frac{1}{4}(\bar{x}_2 + x_4 + x_5) \right] + 2\left[ -\frac{3}{16} + \frac{1}{8}(x_1 + \bar{x}_2 + x_3 + \bar{x}_4) \right] \\ &= \frac{79}{8} + \frac{3}{4}x_1 + \frac{25}{4}x_2 - \frac{7}{4}x_3 - \frac{7}{4}x_4 - 3x_5 \ .\end{aligned}$$

We go on to observe that the best linear approximation preserves some important properties of the approximated function. In the following definitions and propositions $f$ is an arbitrary pseudo-Boolean function and $l(x) = a_0 + \sum_{i \in N} a_i x_i$ is its best linear approximation.

We say that the variable $x_i$ is a dummy for $f$ if for all $x \in B^n$

$$f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) \ . \tag{2.12}$$

***Proposition 2.4:*** If $x_i$ is a dummy for $f$ then $x_i$ is a dummy for $l$ (or equivalently $a_i = 0$).

*Proof:* From Theorem 2.3, $a_i = \text{Ave}\{\Delta_i f(x) : x \in B^n\}$, but $\Delta_i f(x) \equiv 0$ so $a_i = 0$.

We say that a permutation $\pi$ of $N$ is a symmetry of $f$ for all $x \in B^n$

$$f(x_1, x_2, \ldots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}) \ . \tag{2.13}$$

*Proposition 2.5:* If $\pi$ is a symmetry of $f$ then $\pi$ is a symmetry of $l$ (or equivalently $a_i = a_{\pi(i)}$ for all $i \in N$).

*Proof:* If $\pi$ were not a symmetry of $l$ then $l(x_{\pi(1)}, \ldots, x_{\pi(n)})$ would define another linear approximation of $f$ that is as good as $l$, contradicting uniqueness.

We say that $f$ is nondecreasing in the variable $x_i$ if for all $x \in B^n$

$$f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n) \leq f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) \ . \tag{2.14}$$

*Proposition 2.6:* If $f$ is nondecreasing in $x_i$ then $l$ is nondecreasing in $x_i$ (or equivalently $a_i \geq 0$).

*Proof:* $\Delta_i f(x) \geq 0$ for all $x \in B^n$, hence $a_i = \text{Ave}\{\Delta_i f(x) : x \in B^n\} \geq 0$ .

For two distinct variables $x_i$ and $x_j$, we write $x_i \geq x_j$ for $f$ if for all $x \in B^n$ such that $x_i = x_j = 0$

$$f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) \geq f(x_1, \ldots, x_{j-1}, 1, x_{j+1}, \ldots, x_n) \ . \tag{2.15}$$

*Proposition 2.7:* If $x_i \geq x_j$ for $f$ then $x_i \geq x_j$ for $l$ (or equivalently $a_i \geq a_j$).

*Proof:* By averaging over the $2^{n-2}$ inequalities (2.15) we obtain $\text{Ave}\{f(x) : x_i = 1,\ x_j = 0\} \geq \text{Ave}\{f(x) : x_i = 0,\ x_j = 1\}$. Hence

$$\begin{aligned}
2a_i &= 2\,\text{Ave}\{\Delta_i f(x) : x \in B^n\} \\
&= 2\,\text{Ave}\{f(x) : x_i = 1\} - 2\,\text{Ave}\{f(x) : x_i = 0\} \\
&= \text{Ave}\{f(x) : x_i = 1,\ x_j = 1\} + \text{Ave}\{f(x) : x_i = 1,\ x_j = 0\} \\
&\quad - \text{Ave}\{f(x) : x_i = 0,\ x_j = 1\} - \text{Ave}\{f(x) : x_i = 0,\ x_j = 0\}
\end{aligned}$$

$$\geq \text{Ave}\{f(x): x_i = 1,\ x_j = 1\} + \text{Ave}\{f(x): x_i = 0,\ x_j = 1\}$$

$$- \text{Ave}\{f(x): x_i = 1,\ x_j = 0\} - \text{Ave}\{f(x): x_i = 0,\ x_j = 0\}$$

$$= 2\ \text{Ave}\{f(x): x_j = 1\} - 2\ \text{Ave}\{f(x): x_j = 0\}$$

$$= 2\ \text{Ave}\{\Delta_j f(x): x \in B^n\} = 2a_j\ .$$

In conclusion of this investigation of properties, we note that the converses of Propositions 2.4 – 2.7 are false; the example function (2.10) is a counterexample to all converses. This is not surprising. Indeed, $l$ has a property if $f$ has that property "on average", which may happen without $f$ having the property pointwise.

In certain applications of the theory developed here, one may be interested only in linear approximations that coincide with the given function at $x = \mathbf{0}$ and $x = \mathbf{1}$ (where $\mathbf{0}$ and $\mathbf{1}$ are shorthand for $(0, \ldots, 0)$ and $(1, \ldots, 1)$ respectively). One such field of application is game theory, on which we shall elaborate in the next section. Another field that we have in mind is the theory of subjective probability (see for example Shafer (1976)). There, one considers $n$ mutually exclusive and exhaustive events indexed by $1, \ldots, n$; $f(x)$ is the degree of belief that the true event lies in $\{i: x_i = 1\}$. $f$ is a pseudo-Boolean function and is required to satisfy $f(\mathbf{0}) = 0$, $f(\mathbf{1}) = 1$ and some other properties, but it need not be linear (as an objective probability has to be). It makes sense to approximate such a subjective probability function by an objective one, which of course has to agree with the given function at $\mathbf{0}$ and at $\mathbf{1}$.

With this motivation, we introduce the notation $L^f$ for the set of all linear functions $l: B^n \rightarrow \boldsymbol{R}$ that satisfy $l(\mathbf{0}) = f(\mathbf{0})$ and $l(\mathbf{1}) = f(\mathbf{1})$, where $f$ is a given pseudo-Boolean function.

*Definition 2.8:* Let $f: B^n \rightarrow \boldsymbol{R}$ be a pseudo-Boolean function. The best faithful linear approximation of $f$ is the function $\bar{l} \in L^f$ which minimizes $\sum_{x \in B^n} [f(x) - l(x)]^2$ among all functions $l \in L^f$. We write $\bar{l} = \bar{A}(f)$.

As in Definition 2.1, existence and uniqueness follow from the observation that $\bar{A}(f)$ is the orthogonal projection of $f$ on $L^f$, when the latter is viewed as an affine space in $\boldsymbol{R}^{2^n}$. Furthermore, it can be checked that $\bar{A}$ is a linear operator (but note that this checking takes more care than for $A$, since $L^f$ is not in general a subspace and it varies with $f$). Next we observe that since $L^f \subset L$, the projection of $f$ on $L^f$ may be realized by first projecting $f$ on $L$ and then projecting the obtained function on $L^f$. Thus we have the following.

*Lemma 2.9:* Let $f$ be a pseudo-Boolean function. Then $\bar{l} = \bar{A}(f)$ minimizes $\sum_{x \in B^n} [(Af)(x) - l(x)]^2$ among all functions $l \in L^f$.

We proceed now to characterize the best faithful linear approximation of a function. Let $\bar{l}=\bar{A}(f)$. We shall derive information on the difference $\bar{l}-Af$; for the moment we note that this is a linear function and introduce a notation for it:

$$(\bar{l}-Af)(x)=\hat{l}(x)=\hat{a}_0+\sum_{i\in N}\hat{a}_i x_i \ . \tag{2.16}$$

Let $i,j\in N$, $i\neq j$. Since $\bar{l}+a(x_i-x_j)\in L^f$ or all real $a$, it follows from Lemma 2.9 that the expression $\sum_{x\in B^n}[(Af)(x)-\bar{l}(x)-a(x_i-x_j)]^2$ is minimized by $a=0$. But $a$ affects only the terms where $x_i\neq x_j$, so $\sum_{x_i=0,x_j=1}[\hat{l}(x)-a]^2+\sum_{x_i=1,x_j=0}[-\hat{l}(x)-a]^2$ is minimized by $a=0$. Hence using Lemma 2.2,

$$0=\mathrm{Ave}\{\hat{l}(x):x_i=0,\ x_j=1\}+\mathrm{Ave}\{-\hat{l}(x):x_i=1,\ x_j=0\}$$

$$=\hat{a}_0+\hat{a}_j+\frac{1}{2}\sum_{k\in N\setminus\{i,j\}}\hat{a}_k-\hat{a}_0-\hat{a}_i-\frac{1}{2}\sum_{k\in N\setminus\{i,j\}}\hat{a}_k=\hat{a}_j-\hat{a}_i \ .$$

As $i,j$ are arbitrary, we conclude that $\hat{a}_1=\hat{a}_2=\ldots=\hat{a}_n$. Once $Af$ is determined (by Theorem 2.3), this fact completely determines $\bar{l}\in L^f$. Thus we have the following.

*Theorem 2.10:* Given a pseudo-Boolean function $f:B^n\to\boldsymbol{R}$, its best faithful linear approximation is characterized as the unique linear function $\bar{l}:B^n\to\boldsymbol{R}$ that agrees with $f$ at $\mathbf{0}$ and at $\mathbf{1}$ and whose coefficients $\bar{a}_1,\bar{a}_2,\ldots,\bar{a}_n$ differ from the corresponding average first derivatives of $f$ by a fixed amount.

*Example* (continued): Let us return to the function $f$ given by (2.10), and compute $(\bar{A}f)(x)=\bar{a}_0+\sum_{i\in N}\bar{a}_i x_i$. First $\bar{a}_0=f(\mathbf{0})=8$. Next, $\sum_{i\in N}\bar{a}_i=f(\mathbf{1})-f(\mathbf{0})=3$; since the sum of the corresponding coefficients of $Af$ is $\frac{7}{2}$ (from (2.11)), each of them has to be decreased by $\left(\frac{7}{2}-3\right)\Big/5=\frac{1}{10}$. Thus we get:

$$(\bar{A}f)(x)=8+\left(-\frac{5}{4}-\frac{1}{10}\right)x_1+\left(\frac{15}{4}-\frac{1}{10}\right)x_2+\left(\frac{9}{4}-\frac{1}{10}\right)x_3$$
$$+\left(-\frac{5}{4}-\frac{1}{10}\right)x_4+\left(0-\frac{1}{10}\right)x_5$$
$$=8-\frac{27}{20}x_1+\frac{73}{20}x_2+\frac{43}{20}x_3-\frac{27}{20}x_4-\frac{1}{10}x_5 \ .$$

As regards the preservation of properties, it is easy to see that the counterparts of Propositions 2.5 and 2.7 (concerning symmetries and the relation $x_i \geq x_j$) hold true for the best faithful linear approximation. The other two properties, however, are not preserved. Indeed, consider the expression (2.10) as defining a function $f: \boldsymbol{B}^6 \to \boldsymbol{R}$. Then $x_6$ is a dummy for $f$, and $f$ is nondecreasing in $x_6$. $Af$ is still given by (2.11), and to obtain $\bar{A}f$ each coefficient has to be decreased by $\left(\frac{7}{2} - 3\right) \Big/ 6 = \frac{1}{12}$. In particular, the coefficient of $x_6$ in $\bar{A}f$ becomes $-\frac{1}{12}$. Thus, we see that the best faithful linear approximation is not quite faithful when it comes to preservation of properties.

## 3 Linear Approximations and Game Theory

In game theory, the set $N = \{1, \ldots, n\}$ stands for the set of *players*, and its subsets are called *coalitions*; $2^N$ denotes the set of all coalitions. A *game* (more precisely, a side-payment game in characteristic function form) is a function $v: 2^N \to \boldsymbol{R}$ with $v(\emptyset) = 0$. One thinks of $v(S)$ as the worth of the coalition $S$.

Through the natural identification of coalitions with their characteristic vectors in $\boldsymbol{B}^n$, a game is seen to be a pseudo-Boolean function that assumes the value 0 at $\mathbf{0}$. We shall henceforth make this identification.

The polynomial expression (1.1) was introduced in game theory (Owen (1972)) as the *multilinear extension* of a game. When this reduces to a linear function as in (1.2) – with $a_0 = 0$ by assumption – the game is said to be *additive* (or also *inessential*).

A *simple game* is a game $v$ that satisfies $v(S) \in \{0, 1\}$ for all $S \subseteq N$ and moreover $S \subset T$ implies $v(S) \leq v(T)$. Coalitions with $v(S) = 1$ are called *winning*, the rest are *losing*. Simple games model the allocation of power in committees: a coalition is winning if it controls the decisions.

Given a simple game, one would like to assign to the players nonnegative numbers $p_1, \ldots, p_n$ which indicate their individual power in the game. Such an assignment is called a *power index*. A particular power index that has been studied theoretically as well as applied to settle constitutional issues in the courts is the Banzhaf (1965) index. It is defined through the notion of a *swing*, which occurs when a player, by joining a coalition, turns it from losing to winning. The (normalized) Banzhaf index assigns to player $i$ the probability $p_i$ that a swing occurs when $i$ joins a coalition picked at random from among the $2^{n-1}$ coalitions not including $i$.

Clearly, $\Delta_i v(x) \in \{0, 1\}$ for a simple game $v$ (see (2.6)), and a swing for $i$ occurs if and only if $\Delta_i v$ at the corresponding $x$ is 1. Thus the Banzhaf index is rendered by

$$p_i = \text{Ave}\{\Delta_i v(x) : x \in B^n\} \ . \tag{3.1}$$

Now, from Theorem 2.3 we have the following.

*Corollary 3.1:* For a simple game $v$, the Banzhaf power index assigns to each player $i$ the coefficient of $x_i$ in the best linear approximation of $v$.

This result offers a new justification for the use of the Banzhaf index as opposed to other power indices that have been suggested.

A major problem in game theory is how to distribute the worth $v(N)$ of the total coalition among its members in a way that takes into account reasonably the worths of the various coalitions. Formally, such a distribution is a *payoff vector:* $(a_1, \ldots, a_n) \in \boldsymbol{R}^n$ satisfying $\sum_{i \in N} a_i = v(N)$. Various solution concepts that associate with every (not necessarily simple) game a payoff vector have been suggested and studied.

A payoff vector $(a_1, \ldots, a_n)$ may be identified with the additive game $l(x) = \sum_{i \in N} a_i x_i$ in which the worth of every coalition is the sum of payoffs to its members. Thus we have a one-to-one correspondence between the payoff vectors for a given game $v$ and the functions in $L^v$ (see the notation preceding Definition 2.8). This suggests to treat the game theoretic problem of associating a payoff vector to a game as a problem of approximating the game $v$ by another game that lies in $L^v$. This point of view is not new to game theory, but it has not played a central role in it.

One way to associate a payoff vector to a game is known as the Banzhaf value. It consists of generalizing the Banzhaf index via (3.1) to handle nonsimple games, and renormalizing it to obtain the payoff vector

$$\left( \frac{p_1}{\sum_{i \in N} p_i} v(N), \ldots, \frac{p_n}{\sum_{i \in N} p_i} v(N) \right) , \tag{3.2}$$

provided that $\sum_{i \in N} p_i \neq 0$. If however we take the approximation point of view and recall Theorem 2.10, we are led to the payoff vector

$$\left( p_1 + \frac{v(N) - \sum_{i \in N} p_i}{n}, \ldots, p_n + \frac{v(N) - \sum_{i \in N} p_i}{n} \right) . \tag{3.3}$$

We see that (3.3) differs from (3.2) in that the normalization is by an additive amount instead of by a multiplicative factor. A serious drawback of (3.3) as a solution concept in game theory is its failure to preserve dummies and monotonicity as discussed at the end of the previous section. The better known (3.2) also has its problems.

A much more popular solution concept in game theory is the Shapley value (1953). It associates with a game $v$ the payoff vector $\varphi v$ whose components are given by

$$(\varphi v)_i = \sum_{S \subseteq N \setminus \{i\}} \frac{s!\,(n-s-1)!}{n!} [v(S \cup \{i\}) - v(S)] \ , \tag{3.4}$$

where $s = |S|$. Comparing to (3.1), we see that the Shapley value replaces the arithmetical average by a certain weighted average.

The Shapley value is obtained with an axiomatic approach. It is natural to ask whether it can also be obtained with the approximation approach. Indeed it can, but the least squares criterion in choosing a best approximation has to be replaced by a suitable weighted least squares criterion. The precise result, which is not new, is quoted here without proof.

*Theorem 3.2:* Charnes et al. (1985). For any game $v$, the additive game $w$ that corresponds to the Shapley value of $v$ (i.e., $w(S) = \sum_{i \in S} (\varphi v)_i$) minimizes $\sum_{S \subseteq N} \mu_S [v(S) - w(S)]^2$ among all additive games $w$ satisfying $w(N) = v(N)$; provided $\mu_S = \binom{n-2}{|S|-1}^{-1}$ for $S \neq \emptyset, N$.

## 4 Approximations of Arbitrary Degree

If one is willing to accept polynomial expressions which are not necessarily linear, but are of low degree, as simple enough, one is led to the idea of approximating a pseudo-Boolean function by a function of degree at most $k$.

*Definition 4.1:* Let $f\colon B^n \to \mathbf{R}$ be a pseudo-Boolean function, and let $k$ be an integer, $0 \le k \le n$. The best $k$-th approximation of $f$ is the function $g\colon B^n \to \mathbf{R}$ of degree $\le k$ which minimizes $\sum_{x \in B^n} [f(x) - g(x)]^2$ among all functions of degree $\le k$. We write $g = A_k(f)$.

When $k = 0$, the definition gives the best constant approximation; when $k = 1$, it coincides with the definition of the best linear approximation given in Section 2; when $k = 2$, it gives the best quadratic approximation; finally, every $f\colon B^n \to \mathbf{R}$ is its own best $n$-th approximation.

As in the linear case, existence and uniqueness of the best $k$-th approximation follow from the fact that $A_k$ is the orthogonal projection onto the subspace $V_k$

of $\boldsymbol{R}^{2^n}$ consisting of the pseudo-Boolean functions of degree $\leq k$. Thus $A_k$ is a linear operator, and in particular is covariant with respect to addition of functions of degree $\leq k$ (in the sense of (2.1) and (2.2)). Since the subspaces $V_k$ are nested, the operators $A_k$ commute in the following sense:

$$k \leq k' \Rightarrow A_k(A_{k'}f) = A_k(f) \quad \text{for all} \quad f\ . \tag{4.1}$$

Towards the characterization of the best $k$-th approximation of a function, we recall that the $m$-th order derivative of a pseudo-Boolean function $f$ with respect to $x_{i_1}, x_{i_2}, \ldots, x_{i_m}$ at the point $x \in B^n$ is defined inductively as

$$\Delta_{i_1 i_2 \ldots i_m} f(x) = \Delta_{i_1}(\Delta_{i_2 \ldots i_m} f)(x)\ , \tag{4.2}$$

where $\Delta_i f(x)$ is the (first) derivative defined by (2.6). For completeness, the 0-th order derivative of $f$ is $f$ itself. Note that $\Delta_{i_1 i_2 \ldots i_m} f$ depends only on the components $x_j$ such that $j \notin \{i_1, i_2, \ldots, i_m\}$, but we regard the derivatives of all orders as functions on $B^n$. The definitions of higher order derivatives may be made explicit in the manner of (2.6). For instance,

$$\begin{aligned}\Delta_{12} f(x) = {} & f(1,1,x_3,\ldots,x_n) - f(1,0,x_3,\ldots,x_n) \\ & - f(0,1,x_3,\ldots,x_n) + f(0,0,x_3,\ldots,x_n)\ .\end{aligned}$$

By considering such expressions it is obvious that

$$\Delta_i(\Delta_j f)(x) = \Delta_j(\Delta_i f)(x) \quad \text{for all} \quad x \in B^n\ , \tag{4.3}$$

and by repeated applications of (4.3) it is seen that $\Delta_{i_1 i_2 \ldots i_m} f$ is independent of the ordering of the subscripts $i_1, i_2, \ldots, i_m$. Finally, observe that

$$\deg(\Delta_i f) \leq \deg(f) - 1 \tag{4.4}$$

whenever $f \not\equiv 0$, and therefore all $k$-th order derivatives of a function of degree $\leq k$ are constant. Explicitly, if $\deg(f) \leq k$ and $i_1, i_2, \ldots, i_k$ are distinct then $\Delta_{i_1 i_2 \ldots i_k} f \equiv a_{\{i_1, i_2, \ldots, i_k\}}$, the latter being the corresponding coefficient in the polynomial expression of $f$.

By imitating the proof of the characterization of the best linear approximation, arguing successively about orders $0, 1, \ldots, k$, the following generalization of Theorem 2.3 is established.

*Theorem 4.2:* Given a pseudo-Boolean function $f: B^n \to \boldsymbol{R}$ and an integer $k$, $0 \le k \le n$, the best $k$-th approximation of $f$ is characterized as the unique function $g: B^n \to \boldsymbol{R}$ of degree $\le k$ that agrees with $f$ in all average $m$-th order derivatives for $m = 0, 1, \ldots, k$.

The analogy that we pointed out between Theorem 2.3 and linear approximations of differentiable functions carries over to an analogy between Theorem 4.2 and $k$-th order Taylor expansions.

The above characterization can be used to derive explicit formulae for $k$-th approximations. We shall consider here the case of quadratic approximation, i.e., $k = 2$. It will be clear that formulae for higher $k$ can be derived by the same method.

Suppose first that $f$ is of the form $f(x) = \prod_{i \in T} x_i$. Then for any pair of distinct $i,j$ such that $\{i,j\} \subseteq T$, $\Delta_{ij} f(x) = \prod_{h \in T \setminus \{i,j\}} x_h$ and therefore $\mathrm{Ave}\{\Delta_{ij} f(x) : x \in B^n\} = \frac{1}{2^{|T|-2}}$; for $\{i,j\} \nsubseteq T$, $\Delta_{ij} f(x) \equiv 0$. By the equality of average second order derivatives, the best quadratic approximation of $f$ must take the form

$$g(x) = a_0 + \sum_{i \in N} a_i x_i + \frac{1}{2^{|T|-2}} \sum_{\substack{i,j \in T \\ i<j}} x_i x_j \ .$$

From here

$$\Delta_i g(x) = \begin{cases} a_i + \frac{1}{2^{|T|-2}} \sum_{j \in T \setminus \{i\}} x_j & \text{if} \quad i \in T \ , \\ a_i & \text{if} \quad i \notin T \ . \end{cases}$$

Hence

$$\mathrm{Ave}\{\Delta_i g(x) : x \in B^n\} = \begin{cases} a_i + \frac{|T|-1}{2^{|T|-1}} & \text{if} \quad i \in T \ , \\ a_i & \text{if} \quad i \notin T \ . \end{cases}$$

However

$$\mathrm{Ave}\{\Delta_i f(x) : x \in B^n\} = \begin{cases} \frac{1}{2^{|T|-1}} & \text{if} \quad i \in T \ , \\ 0 & \text{if} \quad i \notin T \ . \end{cases}$$

By the equality of average first derivatives, $a_i$ are thus determined and we can rewrite

$$g(x) = a_0 - \frac{|T|-2}{2^{|T|-1}} \sum_{i \in T} x_i + \frac{1}{2^{|T|-2}} \sum_{\substack{i,j \in T \\ i<j}} x_i x_j \ .$$

From here

$$\mathrm{Ave}\{g(x) : x \in B^n\} = a_0 - \frac{|T|(|T|-2)}{2^{|T|}} + \frac{|T|(|T|-1)}{2^{|T|+1}} = a_0 - \frac{|T|^2 - 3|T|}{2^{|T|+1}} \ .$$

Since $\mathrm{Ave}\{f(x) : x \in B^n\} = \frac{1}{2^{|T|}}$, the equality of average values requires that

$$a_0 = \frac{|T|^2 - 3|T|}{2^{|T|+1}} + \frac{1}{2^{|T|}} = \frac{(|T|-1)(|T|-2)}{2^{|T|+1}} \ .$$

Summing up, we have shown that

$$f(x) = \prod_{i \in T} x_i \Rightarrow (A_2 f)(x)$$

$$= \frac{(|T|-1)(|T|-2)}{2^{|T|+1}} - \frac{|T|-2}{2^{|T|-1}} \sum_{i \in T} x_i + \frac{1}{2^{|T|-2}} \sum_{\substack{i,j \in T \\ i<j}} x_i x_j \ . \tag{4.5}$$

The computation of the best quadratic approximation for general $f$ is done by linearity of $A_2$, using (4.5) for each basic monomial.

*Example* (continued): Let us compute the best quadratic approximation of the function $f$ given by (2.10). Using (4.5), we approximate

$$x_2 x_4 x_5 \quad \text{by } \frac{1}{8} - \frac{1}{4}(x_2 + x_4 + x_5) + \frac{1}{2}(x_2 x_4 + x_2 x_5 + x_4 x_5) \ ,$$

$$x_1 x_2 x_3 x_4 \text{ by } \frac{3}{16} - \frac{1}{4}(x_1 + x_2 + x_3 + x_4) + \frac{1}{4}(x_1 x_2 + x_1 x_3 + x_1 x_4 + x_2 x_3 + x_2 x_4 + x_3 x_4) \ .$$

The quadratic part of $f$ is its own approximation. Putting things together and using linearity we get:

$$
\begin{aligned}
(A_2f)(x) &= 8 - x_1 + 5x_2 - x_1x_5 + 4x_3x_5 \\
&\quad - 6\left[\frac{1}{8} - \frac{1}{4}(x_2 + x_4 + x_5) + \frac{1}{2}(x_2x_4 + x_2x_5 + x_4x_5)\right] \\
&\quad + 2\left[\frac{3}{16} - \frac{1}{4}(x_1 + x_2 + x_3 + x_4)\right. \\
&\qquad \left. + \frac{1}{4}(x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4)\right] \\
&= \frac{61}{8} - \frac{3}{2}x_1 + 6x_2 - \frac{1}{2}x_3 + x_4 + \frac{3}{2}x_5 \\
&\quad + \frac{1}{2}x_1x_2 + \frac{1}{2}x_1x_3 + \frac{1}{2}x_1x_4 - x_1x_5 + \frac{1}{2}x_2x_3 \\
&\quad - \frac{5}{2}x_2x_4 - 3x_2x_5 + \frac{1}{2}x_3x_4 + 4x_3x_5 - 3x_4x_5 \ .
\end{aligned}
$$

The reader is invited to compare this to the best linear approximation of $f$, given by (2.11). Note that (2.11) is not the linear part of $A_2f$, but it is its best linear approximation (as it should be by (4.1)).

We have considered here two families of linear operators on pseudo-Boolean functions: approximations and derivatives. For each family we have a commutativity law: (4.1) for the former, (4.3) for the latter. Do members of the two families commute with each other? In other words, does $\Delta_i(A_kf) = A_k(\Delta_if)$ hold? Since the degree of the right hand side is in general $k$, whereas that of the left hand side, by (4.4), is at most $k-1$, this cannot be true. But once this problem is taken care of, we have the following

*Proposition 4.3:* $\Delta_i(A_kf) = A_{k-1}(\Delta_if)$, $k = 1, \ldots, n$.

*Proof:* According to (4.4), $\deg[\Delta_i(A_kf)] \leq k-1$. Hence by Theorem 4.2 $\Delta_i(A_kf)$ must be the best $(k-1)$-th approximation of $\Delta_if$ if it agrees with it in all average $m$-th order derivatives, $m = 0, \ldots, k-1$. But $m$-th order derivatives of $\Delta_i(A_kf)$ and $\Delta_if$ are $(m+1)$-th order derivatives of $A_kf$ and $f$ respectively, and as $m+1 \leq k$ the required agreement follows from Theorem 4.2 applied to $f$.

Clearly, higher order derivatives can be handled by repeated use of Proposition 4.3. For instance

$$\Delta_{ij} A_k = \Delta_i(\Delta_j A_k) = \Delta_i(A_{k-1}\Delta_j) = A_{k-2}(\Delta_i \Delta_j) = A_{k-2}\Delta_{ij} \ .$$

To complete our study of $k$-th approximations, we consider the question of preservation of properties.

*Proposition 4.4:* If $x_i$ is a dummy for $f$ then $x_i$ is a dummy for $A_k f$.

*Proof:* If $k=0$ there is nothing to prove, so assume $k \geq 1$. Since $\Delta_i f \equiv 0$, also $A_{k-1}(\Delta_i f) \equiv 0$. Hence by Proposition 4.3 $\Delta_i(A_k f) \equiv 0$, which means that $x_i$ is a dummy for $A_k f$.

*Proposition 4.5:* If $\pi$ is a symmetry of $f$ then $\pi$ is a symmetry of $A_k f$.

*Proof:* By uniqueness, just as in the proof of Proposition 2.5.

The other two properties discussed in Section 2 are not preserved in general. Indeed, let $f: \boldsymbol{B}^4 \to \boldsymbol{R}$ be defined by $f(x) = x_1 x_2 x_3$. By (4.5)

$$(A_2 f)(x) = \frac{1}{8} - \frac{1}{4}(x_1 + x_2 + x_3) + \frac{1}{2}(x_1 x_2 + x_1 x_3 + x_2 x_3) \ .$$

Clearly $A_2 f$ is nondecreasing in $x_1$, even though $f$ is, and moreover $x_1 \geq x_4$ is false for $A_2 f$, even though it is true for $f$. Thus, while $A_2 f$ gets closer to $f$ than $Af$, it is less faithful than $Af$ in terms of preserving the properties of $f$. This can be explained as follows. All $A_k f$ preserve these properties "on average", as seen by Theorem 4.2, but only for linear functions is satisfaction "on average" of these properties the same as pointwise satisfaction.

## References

Banzhaf JF III (1965) Weighted voting doesn't work: A mathematical analysis. Rutgers Law Review 19:317–343

Charnes A, Golany B, Keane M, Rousseau J (1985) Extremal principle solutions of games in characteristic function form: Core, Chebychev and Shapley value generalizations. Report CCS 526. The Center for Cybernetic Studies, University of Texas, Austin

Coleman RP (1961) Orthogonal functions for the logical design of switching circuits. IRE Trans EC 10:379–383
Hammer PL, Hansen P, Simeone B (1984) Roof duality, complementation and persistency in quadratic 0–1 optimization. Math Progr 28:121–155
Hammer PL, Rudeanu S (1968) Boolean methods in operations research and related areas. Springer, Berlin Heidelberg New York
Kaplan KR, Winder RO (1965) Chebyshev approximation and threshold functions. IEEE Trans EC 14:250–252
Owen G (1972) Multilinear extensions of games. Management Science 18:P64–P79
Shafer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton
Shapley LS (1953) A value for $n$-person games. In: Kuhn HW, Tucker AW (eds) Contributions to the theory of games II. Princeton University Press, Princeton

Received July 1989
Revised version received January 1990