Swarm and Evolutionary Computation 86 (2024) 101540

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

journal homepage: www.elsevier.com/locate/swevo

Characterization of rankings generated by pseudo-Boolean functions

Imanol Unanue [a,*], María Merino [b,c], Jose A. Lozano [a,c]

[a] Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, Manuel Lardizabal pasealekua 1, 20018, Donostia, Spain
[b] Department of Mathematics, University of the Basque Country UPV/EHU, Sarriena auzoa, 48940, Leioa, Spain
[c] BCAM - Basque Center for Applied Mathematics, Mazarredo Zumarkalea, 14, 48009, Bilbo, Spain

ARTICLE INFO

Keywords:
Binary-based combinatorial optimization problems
Pseudo-Boolean functions
Rankings
Dyck words
Theoretical analysis

ABSTRACT

In this paper we pursue the study of pseudo-Boolean functions as ranking generators. The objective of the work is to find new insights between the relation of the degree m of a pseudo-Boolean function and the rankings that can be generated by these insights. Based on a characterization theorem for pseudo-Boolean functions of degree m, several observations are made. First, we verify that pseudo-Boolean functions of degree $m < n$, where n is the search space dimension, cannot generate all the possible rankings of the solutions. Secondly, the sufficient condition for a ranking to be generated by a pseudo-Boolean function of dimension $(n-1)$ is presented, and also the necessary condition is conjectured. Finally, we observe that the same argument is not sufficient to prove which ranking can be generated by pseudo-Boolean functions of degree $m < n-1$.

1. Introduction

Combinatorial Optimization Problems (COPs) have received much attention from the research communities in the fields of computer science, mathematics, economics, industry and logistics. While most of the research has been carried out in the design of exact and metaheuristics algorithms to solve the problems in an efficient way [1–5], less attention has been devoted to the theoretical study of the problems. Furthermore, it is known that there is no algorithm that will perform on average better than a random search over all the problems [6]. Hence, given a particular COP or a specific instance of it, it is not clear how to choose the most efficient algorithm to solve it. A first step to find this association between "problem/instance - algorithm" is to group problem instances in such a way that those that share similar characteristics (with respect to finding their optimal solutions) are in the same group. This association would significantly reduce the costs of solving problems.

However, grouping problems or instances is a difficult task as they come defined in a variety of different forms. For example, the definition of the Unconstrained Binary Quadratic Problem is not apparently related with the definition of the well-known Knapsack Problem (even if the solutions of both problems are described by binary strings). One way to approach this diversity is to consider fitness functions as rankings of the solutions of the search space (an ordered list of the solutions according to their fitness function values). This makes sense as most algorithms (such as local search or evolutionary algorithms with tournament or ranking selection to name a few) only consider the ranking of the solutions in their machinery instead of the specific fitness function value of a solution. This avenue has been previously followed by [7], where the authors show that the studied permutation-based COPs cannot generate all the possible rankings of solutions and they present the intersection of COPs (instances that can be generated by several problems). Our desired goal is to present a characterization of the rankings according to their features which allows us to select the most "appropriate" algorithm (in terms of efficiency) to solve (an instance of) a problem.

This paper constitutes a further step in the previous direction. Inspired by the works of [7–10], we analyze for the first time the rankings generated by pseudo-Boolean functions of degree $m \leq n$, being n the size of the search space. Our main contributions are the following. First, we prove that there exist rankings that cannot be generated by a pseudo-Boolean function of degree $m < n$. Moreover, we exactly present the necessary conditions for a ranking to be generated by an m-degree pseudo-Boolean function. We provide a novel and easy-to-compute procedure to check when a ranking cannot be generated by an m-degree pseudo-Boolean function. Secondly, we study if the obtained necessary conditions are sufficient conditions to prove when a ranking can be generated by m-degree pseudo-Boolean functions. When $m = n-1$, we conjecture that the answer is affirmative and we calculate the exact number of rankings generated by $(n-1)$-degree pseudo-Boolean functions; whereas when $m < n-1$, the presented procedure is not sufficient to check if a ranking can be generated by an m-degree pseudo-Boolean function. Throughout this paper, we present several examples

* Corresponding author.
E-mail addresses: imanol.unanue@ehu.eus (I. Unanue), maria.merino@ehu.eus (M. Merino), jlozano@bcamath.org (J.A. Lozano).

https://doi.org/10.1016/j.swevo.2024.101540
Received 21 July 2023; Received in revised form 23 January 2024; Accepted 11 March 2024
Available online 16 March 2024

for the particular case of $m = 2$. This scenario is analogous to a well-known problem in the literature: the Unconstrained Binary Quadratic Problem, which is the NP-hard problem with the lowest possible degree polynomial function.

This paper is organized as follows. In Section 2, the required mathematical concepts are defined. In Section 3, the main results are shown: the analysis of the rankings of solutions generated by an m-degree pseudo-Boolean function. Finally, in Section 4, conclusions and future work are presented.

2. Preliminaries

In this paper, we focus on pseudo-Boolean functions, i.e., functions whose possible solutions are codified as 0–1 vectors.

Definition 1 (*Pseudo-Boolean Functions*)**.** Let $\Omega = \{0,1\}^n$ be the search space and $x = x_1 x_2 \dots x_n \in \Omega$ a solution (a binary string of length n). Then, a function $f : \Omega \longrightarrow \mathbb{R}$ is a pseudo-Boolean function. Any pseudo-Boolean function can be written uniquely as a multi-linear polynomial [11,12] (notice that for any bit x_i, if the rest of the bit values are fixed, then the function f is linear with respect to x_i):

$$f(x) = a_0 + \sum_{1\le i_1\le n} a_{i_1} x_{i_1} + \sum_{1\le i_1<i_2\le n} a_{i_1 i_2} x_{i_1} x_{i_2} + \sum_{1\le i_1<i_2<i_3\le n} a_{i_1 i_2 i_3} x_{i_1} x_{i_2} x_{i_3} + \cdots . \tag{1}$$

Definition 2 (*Degree of a Pseudo-Boolean Function*)**.** A pseudo-Boolean function is of degree $m \le n$ if the degree of its polynomial representation is m.

We highly recommend [11] for a deep introduction to pseudo-Boolean functions and their main properties. We consider the maximization problem, so the objective is to obtain

$$x^* = \arg\max_{x\in\Omega} f(x).$$

Definition 3 (*Ranking of Solutions*)**.** A ranking of solutions is an ordered list of all the solutions of Ω. We denote a ranking of solutions with the letter r.

We consider pseudo-Boolean functions as "rankings generators", an ordered list of the solutions according to their fitness function values. Notice that all evolutionary algorithms that use tournament or ranking selection and most local search based algorithms consider functions as rankings generators. Bear in mind that the same ranking can represent several functions, see Example 1. We denote a ranking generated by a pseudo-Boolean function f with the letter r_f.

Example 1. Let $\Omega = \{0,1\}^2$. The two different 1-degree pseudo-Boolean functions $f(x) = 3x_2 - 2x_1$ and $g(x) = -4 + 6x_2 - 2x_1$ generate the same ranking of solutions.

$$\begin{array}{c|cccc} x & 00 & 10 & 01 & 11 \\ \hline f(x) & 0 & -2 & 3 & 1 \\ g(x) & -4 & -6 & 2 & 0 \end{array} \Longrightarrow \begin{cases} f(01) > f(11) > f(00) > f(10) \\ g(01) > g(11) > g(00) > g(10) \end{cases}$$

$$\Longrightarrow r_f = r_g = \begin{bmatrix} 01 \\ 11 \\ 00 \\ 10 \end{bmatrix}$$

Moreover, for any fitness function f, real constant c and positive real constant c', the rankings generated by f, $f + c$ and $c' \cdot f$ are the same: r_f.

To simplify, let us assume that the studied pseudo-Boolean functions are injective. Even the presented analysis can be replicated for non-injective pseudo-Boolean functions, the notation needs to be much more tedious. With this simplification in mind, even though there are infinite n-dimensional pseudo-Boolean functions, the number of possible rankings that can be generated by them is $2^n!$, which is also the number of permutations of the group Σ_{2^n}. Consequently, we can group pseudo-Boolean functions that generate the same ranking of solutions and study COPs as the sets of all the rankings that can be generated by all the instances of the problems. Note that all the results we could obtain for a set of rankings can be extended to all the COPs that generate those rankings regardless of how they have been defined. For instance, the set of rankings that can be generated by both the Unconstrained Binary Quadratic Problem and the Number Partitioning Problem could be solved in the same way.

Notice that, given a pseudo-Boolean function f as in Eq. (1), the value of the coefficient a_0 does not change the ranking. Because of that, we assume that $a_0 = 0$ for the rest of the manuscript.

Next, let us define a partition of Ω based on the parity of zeros of the solutions. Definition 4 is analogous to the one presented in [13] or the *Hamming weight* [14,15].

Definition 4 (*Even (odd) Solutions*)**.** Let $x \in \Omega$ be an even (odd) solution, labeled as E (O), if it contains an even (odd) number of 0 values. Let us denote by $\mathcal{E}$ ($\mathcal{O}$) the set of all even (odd) solutions.

By definition, $\{\mathcal{E}, \mathcal{O}\}$ is a partition of Ω such that $|\mathcal{E}| = |\mathcal{O}| = 2^{n-1}$. For the presented results in this work, there is no difference if we define even and odd solutions according to the number of ones in a solution. The definition of the set of even (odd) solutions can be extended and defines a partition according to a non-empty set of variables $S \subseteq \{1,\dots,n\}$.

Definition 5 (*Even (odd) Solutions Defined by S*)**.** Let $S \subseteq \{1,\dots,n\}$ be a non-empty set of variables and $x \in \Omega$. Then, x is an even (odd) solution defined by S, labeled as E_S (O_S), if it contains an even (odd) number of 0 values from the set of S. Moreover, let us denote by $\mathcal{E}_S$ ($\mathcal{O}_S$) the set of all even (odd) solutions defined by S. By definition, $\{\mathcal{E}_S, \mathcal{O}_S\}$ is a partition of Ω such that $|\mathcal{E}_S| = |\mathcal{O}_S|$.

When the subset S is clear from the context, we will simplify the notation and remove the subindex S from E and O.

3. Studying the rankings generated by pseudo-boolean functions

The main result of this section is to show and prove the existence of rankings of solutions that cannot be generated by any m-degree pseudo-Boolean function, where $m < n$. In addition, the necessary conditions for a ranking to be generated by an m-degree pseudo-Boolean function are presented.

3.1. Characterization of pseudo-boolean functions of degree $m < n$

Let us introduce a characterization of pseudo-Boolean functions according to the partitions of even and odd solutions. To present the characterization of pseudo-Boolean functions, we start with the following lemma.

Lemma 1. *Let $j, n \in \mathbb{N}$, $1 \le j < n$, a set of variables $\{i_1,\dots,i_j\} \subset \{1,\dots,n\}$ and a subset of variables $S \subseteq \{1,\dots,n\}$ such that $|S| > j$. Then, given a value to the variables with indices in $\{i_1,\dots,i_j\}$, the number of even and odd solutions defined by S is the same. In other words, for any two j-tuples $(c_1,\dots,c_j)$, $(d_1,\dots,d_j) \in \{0,1\}^j$, the following equality holds:*

$$|\{x \in \mathcal{E}_S : x_{i_1} = c_1 \wedge \cdots \wedge x_{i_j} = c_j\}| = |\{x \in \mathcal{O}_S : x_{i_1} = d_1 \wedge \cdots \wedge x_{i_j} = d_j\}|. \tag{2}$$

Proof. The lemma is deduced from the definition of the partition $\{\mathcal{E}_S, \mathcal{O}_S\}$. In terms of the relation between the sets $\{i_1,\dots,i_j\}$ and S, there are three types of possible scenarios: (a) $\{i_1,\dots,i_j\} \subset S$; (b) $\{i_1,\dots,i_j\} \not\subset S$ and $S \cap \{i_1,\dots,i_j\} \neq \emptyset$; and (c) $S \cap \{i_1,\dots,i_j\} = \emptyset$.

Without loss of generality, let us consider the case (a) and that S has $1 \leq s < n-j$ additional elements apart from $\{i_1,\dots,i_j\}$. Then, there are 2^{s-1} solutions of $\mathcal{E}_S$ of length $j+s$ and 2^{n-j-s} options for the rest of terms in $\{1,\dots,n\}$. Therefore, there are in total 2^{n-j-1} solutions of $\mathcal{E}_S$ and 2^{n-j-1} solutions of $\mathcal{O}_S$, where the bit values in the positions $\{i_1,\dots,i_j\}$ are determined as in Eq. (2). The cases (b) and (c) are proved analogously. $\square$

Now, let us present and prove the main result of Section 3.1, which is the characterization of pseudo-Boolean functions of degree $m < n$ (Theorem 2).

Theorem 2. *Let $\Omega = \{0,1\}^n$ and $f : \Omega \longrightarrow \mathbb{R}$ a pseudo-Boolean function. Then, f is a pseudo-Boolean function of degree $m < n$ if and only if*

$$\begin{cases} \forall S \subseteq \{1,\dots,n\} \text{ such that } |S| > m\,,\ \sum_{x\in\mathcal{E}_S} f(x) = \sum_{x\in\mathcal{O}_S} f(x) \\ \exists S \subseteq \{1,\dots,n\} \text{ such that } |S| = m \text{ and } \sum_{x\in\mathcal{E}_S} f(x) \neq \sum_{x\in\mathcal{O}_S} f(x) \end{cases} \tag{3}$$

Furthermore, when the equality holds, the sum $\sum_{x\in\mathcal{E}_S} f(x)$ is half of the sum of the function value of all the solutions of the search space:

$$2^{n-1}\cdot \bar{f} = 2^{n-2}\sum_{1\leq i_1\leq n} a_{i_1} + 2^{n-3}\sum_{1\leq i_1<i_2\leq n} a_{i_1 i_2} + \cdots + 2^{n-m-1}\sum_{1\leq i_1<i_2<\cdots<i_m\leq n} a_{i_1\dots i_m}, \tag{4}$$

where $\bar{f}$ is the average fitness function value of f.

Proof. $\boxed{\Longrightarrow}$ Let f be an m-degree polynomial defined over $\{0,1\}^n$. Considering Equality (2) of Lemma 1, for any set S such that $|S| > m$, there are the same number of solutions with $x_{i_1} = 1$, with $x_{i_1} = x_{i_2} = 1$, … and with $x_{i_1} = \cdots = x_{i_m} = 1$ in $\mathcal{E}_S$ and $\mathcal{O}_S$. Therefore, each coefficient a_{i_1}, $a_{i_1 i_2}$, …, $a_{i_1\dots i_m}$ appears the same number of times in $\sum_{x\in\mathcal{E}_S} f(x)$ and in $\sum_{x\in\mathcal{O}_S} f(x)$ and, consequently, $\sum_{x\in\mathcal{E}_S} f(x) = \sum_{x\in\mathcal{O}_S} f(x)$.

On the other hand, because f is an m-degree polynomial, there exists, at least, one non-null coefficient $a_{i_1\dots i_m}$. Consequently, when $S = \{i_1,\dots,i_m\}$, the solutions such that $x_{i_1} = \cdots = x_{i_m} = 1$ only appear in $\mathcal{E}_S$ whereas the rest of coefficients appear the same number of times in $\sum_{x\in\mathcal{E}_S} f(x)$ and $\sum_{x\in\mathcal{O}_S} f(x)$. So, it implies that $\sum_{x\in\mathcal{E}_S} f(x) \neq \sum_{x\in\mathcal{O}_S} f(x)$.

$\boxed{\Longleftarrow}$ Let $f : \Omega \longrightarrow \mathbb{R}$ be a function that fulfills Eq. (3). Let us consider a set of binary variables S such that $|S| = m$ and $\sum_{x\in\mathcal{E}_S} f(x) \neq \sum_{x\in\mathcal{O}_S} f(x)$. Because of Equality (2) of Lemma 1, for any nonempty subset of indexes $\{i_1,\dots,i_j\} \subset S$, the coefficient $a_{i_1\dots i_j}$ appears the same number of times in the sums $\sum_{x\in\mathcal{E}_S} f(x)$ and $\sum_{x\in\mathcal{O}_S} f(x)$. Therefore, because the solutions such that $x_{i_1} = \cdots = x_{i_m} = 1$ only appear in $\mathcal{E}_S$, the only coefficient which causes $\sum_{x\in\mathcal{E}_S} f(x) \neq \sum_{x\in\mathcal{O}_S} f(x)$ is the coefficient $a_{i_1\dots i_m}$, which implies that $a_{i_1\dots i_m} \neq 0$ and consequently the function f is at least an m-degree pseudo-Boolean function.

In addition, by hypothesis, for any set of binary variables S such that $|S| > m$, the equality $\sum_{x\in\mathcal{E}_S} f(x) = \sum_{x\in\mathcal{O}_S} f(x)$ holds. Then, the solutions such that $x_{i_1} = \cdots = x_{i_{|S|}} = 1$ have no relevance in the sums and therefore $a_{i_1\dots i_{|S|}}$ must be a null coefficient. Consequently, f is an m-degree polynomial.

Finally, let us calculate the exact value of the sum $\sum_{x\in\mathcal{E}_S} f(x) = 2^{n-1}\bar{f}$. For a set of indexes $\{i_1,\dots,i_j\}$, $1 \leq j \leq n$, the number of solutions such that $x_{i_1} = \cdots = x_{i_j} = 1$ is 2^{n-j}. Because f is an m-degree pseudo-Boolean function, for any subset of indexes $\{i_1,\dots,i_j\}$ such that $m < j \leq n$, then $a_{i_i\dots i_j} = 0$, which implies that Eq. (4) is fulfilled. $\square$

Example 2. Let $n = 3$ and f be the following fitness function:

$$f(x_1x_2x_3) = -16 + 33x_1 + 34x_2 + 36x_3 - 64x_1x_2 - 64x_1x_3 - 64x_2x_3 + 128x_1x_2x_3.$$

The function is similar to the well-known function BINVAL. By definition of f, it is obvious that $\sum_{x\in\mathcal{E}} f(x) > \sum_{x\in\mathcal{O}} f(x)$. So, by Theorem 2, the function f cannot be rewritten as a pseudo-Boolean function of degree 2.

Notice that Eq. (3) depends on the cardinality of S, not on the indexes of S. Theorem 2 shows all the conditions that any m-degree pseudo-Boolean function must fulfill. In addition, from Theorem 2, the following corollary is obtained.

Corollary 3. *Let f be an m-degree pseudo-Boolean function defined over $\{0,1\}^n$ ($m < n$). For any subsets $S, S' \subseteq \{1,\dots,n\}$ such that $|S|, |S'| \geq m+1$, then the following holds,*

$$\sum_{x\in\mathcal{E}_S\cap\mathcal{O}_{S'}} f(x) = \sum_{x\in\mathcal{E}_{S'}\cap\mathcal{O}_S} f(x) \tag{5}$$

and

$$\sum_{x\in\mathcal{E}_S\cap\mathcal{E}_{S'}} f(x) = \sum_{x\in\mathcal{O}_S\cap\mathcal{O}_{S'}} f(x). \tag{6}$$

In addition, if $S \subset S'$, Equalities (5) and (6) are rewritten respectively as

$$\sum_{x\in\mathcal{E}_S\cap\mathcal{O}_{S'\setminus S}} f(x) = \sum_{x\in\mathcal{O}_S\cap\mathcal{O}_{S'\setminus S}} f(x) \tag{7}$$

and

$$\sum_{x\in\mathcal{E}_S\cap\mathcal{E}_{S'\setminus S}} f(x) = \sum_{x\in\mathcal{O}_S\cap\mathcal{E}_{S'\setminus S}} f(x). \tag{8}$$

Proof. By Theorem 2, for any subsets S, S' such that $|S|, |S'| \geq m+1$,

$$\sum_{x\in\mathcal{E}_S} f(x) = \sum_{x\in\mathcal{E}_{S'}} f(x) = \sum_{x\in\mathcal{O}_S} f(x) = \sum_{x\in\mathcal{O}_{S'}} f(x).$$

On the other hand, since for any subset S $\{\mathcal{E}_S, \mathcal{O}_S\}$ is a partition of Ω, we can decompose each summation:

$$\sum_{x\in\mathcal{E}_S} f(x) = \sum_{x\in\mathcal{E}_S\cap\mathcal{E}_{S'}} f(x) + \sum_{x\in\mathcal{E}_S\cap\mathcal{O}_{S'}} f(x).$$

Consequently,

$$\begin{aligned} \sum_{x\in\mathcal{E}_S} f(x) = \sum_{x\in\mathcal{E}_{S'}} f(x) \Longleftrightarrow \sum_{x\in\mathcal{E}_S\cap\mathcal{E}_{S'}} f(x) + \sum_{x\in\mathcal{E}_S\cap\mathcal{O}_{S'}} f(x) &= \sum_{x\in\mathcal{E}_{S'}\cap\mathcal{E}_S} f(x) \\ &+ \sum_{x\in\mathcal{E}_{S'}\cap\mathcal{O}_S} f(x) \\ \Longleftrightarrow \sum_{x\in\mathcal{E}_S\cap\mathcal{O}_{S'}} f(x) &= \sum_{x\in\mathcal{E}_{S'}\cap\mathcal{O}_S} f(x). \end{aligned}$$

Equality (6) is analogously obtained:

$$\sum_{x\in\mathcal{E}_S} f(x) = \sum_{x\in\mathcal{O}_{S'}} f(x) \Longleftrightarrow \sum_{x\in\mathcal{E}_S\cap\mathcal{E}_{S'}} f(x) = \sum_{x\in\mathcal{O}_S\cap\mathcal{O}_{S'}} f(x).$$

Finally, when $S \subset S'$:

- If $x \in \mathcal{E}_S \cap \mathcal{E}_{S'}$, then $x \in \mathcal{E}_S \cap \mathcal{E}_{S'\setminus S}$.
- If $x \in \mathcal{E}_S \cap \mathcal{O}_{S'}$, then $x \in \mathcal{E}_S \cap \mathcal{O}_{S'\setminus S}$.
- If $x \in \mathcal{O}_S \cap \mathcal{O}_{S'}$, then $x \in \mathcal{O}_S \cap \mathcal{E}_{S'\setminus S}$.
- If $x \in \mathcal{O}_S \cap \mathcal{E}_{S'}$, then $x \in \mathcal{O}_S \cap \mathcal{O}_{S'\setminus S}$.

Consequently,

$$\sum_{x\in\mathcal{E}_S\cap\mathcal{O}_{S'}} f(x) = \sum_{x\in\mathcal{E}_{S'}\cap\mathcal{O}_S} f(x) \Longleftrightarrow \sum_{x\in\mathcal{E}_S\cap\mathcal{O}_{S'\setminus S}} f(x) = \sum_{x\in\mathcal{O}_S\cap\mathcal{O}_{S'\setminus S}} f(x)$$

and

$$\sum_{x\in\mathcal{E}_S\cap\mathcal{E}_{S'}} f(x) = \sum_{x\in\mathcal{O}_{S'}\cap\mathcal{O}_S} f(x) \Longleftrightarrow \sum_{x\in\mathcal{E}_S\cap\mathcal{E}_{S'\setminus S}} f(x) = \sum_{x\in\mathcal{O}_S\cap\mathcal{E}_{S'\setminus S}} f(x). \quad \square$$

Once Theorem 2 and Corollary 3 are presented, our next goal is to show that there exist rankings of solutions that cannot be generated by pseudo-Boolean functions of degree $m < n$.

3.2. Study of pseudo-boolean functions of degree $m = n - 1$

Based on Theorem 2, several new results are obtained. The first result will prove that some rankings of solutions follow a pattern which implies that they do not fulfill the equalities of Eq. (3) of Theorem 2 (and consequently cannot be generated by a $(n-1)$-degree pseudo-Boolean function or, equivalently, the ranking can only be generated by an n-degree pseudo-Boolean function). This specific result is enough to prove that, when $m < n$, m-degree pseudo-Boolean functions cannot generate all the possible rankings from the space of solutions.[1]

To show that the pseudo-Boolean functions of degree $m < n$ cannot generate all the rankings of solutions, new definitions are required.

Definition 6 (*Word of a Ranking*)**.** Let f be a pseudo-Boolean function defined over $\{0,1\}^n$ and r_f the ranking generated by f. Let us denote by $r_f(i)$ the ith solution of the ranking r_f. Then, we define the word of the ranking r_f, denoted by W_f, as the ordered list of length 2^n with the alphabet $\{E, O\}$ in the following way:

$$W_f = \begin{bmatrix} w_1 \\ \vdots \\ w_{2^n} \end{bmatrix} \text{ s.t. } w_i = \begin{cases} E, & \text{if } r_f(i) \text{ is an even solution,} \\ O, & \text{if } r_f(i) \text{ is an odd solution.} \end{cases}$$

To simplify notation, when a word is considered without a function f, we will simplify the notation and remove the subindex f from W.

Example 3. Let us consider the fitness function $f(x) = x_1 - 3x_2 + 3x_3 - 2x_1x_2 + 7x_1x_3 - x_2x_3 + 11x_1x_2x_3$ and calculate the word of its ranking.

x	$f(x)$
000	0
100	1
010	−3
001	3
110	−4
101	11
011	−1
111	16

$$\Longrightarrow r_f = \begin{bmatrix} 111 \\ 101 \\ 001 \\ 100 \\ 000 \\ 011 \\ 010 \\ 110 \end{bmatrix} \Longrightarrow W_f = \begin{bmatrix} E \\ O \\ E \\ E \\ O \\ O \\ E \\ O \end{bmatrix}.$$

Moreover, we extend the definition of the words of a ranking and present two new definitions.

Definition 7 (*Word of a Ranking Defined by S*)**.** Let f be a pseudo-Boolean function defined over $\{0,1\}^n$, r_f the ranking generated by f and S a subset of binary variables. Let us denote by $r_f(i)$ the ith solution of the ranking r_f. Then, we define *the word of the ranking r_f defined by S*, denoted by W_f^S, as the ordered list of length 2^n with the alphabet $\{E, O\}$ in the following way:

$$W_f^S = \begin{bmatrix} w_1^S \\ \vdots \\ w_{2^n}^S \end{bmatrix} \text{ s.t. } w_i^S = \begin{cases} E, & \text{if } r_f(i) \text{ is an even solution defined by } S, \\ O, & \text{if } r_f(i) \text{ is an odd solution defined by } S. \end{cases}$$

To simplify notation, when a word defined by S is considered without a function f, we will simplify the notation and remove the subindex f from W^S.

Definition 8 (*Word of a Ranking with Constraints C*)**.** Let f be a pseudo-Boolean function defined over $\{0,1\}^n$ and C a set of constraints (specific bit values) defined over k bit values, $1 \le k \le n-1$. Let us denote by $f|_C$ the reduction of the function f to all the solutions that fulfill the constraints of C, and $r_{f|_C}$ the ranking generated by $f|_C$. Then, we define *the word of the ranking* $r_{f|_C}$, denoted by $W_{f|_C}$, as the ordered list of length 2^{n-k} with the alphabet $\{E, O\}$ in the following way:

$$W_{f|_C} = \begin{bmatrix} w_1 \\ \vdots \\ w_{2^{n-k}} \end{bmatrix} \text{ s.t. } w_i = \begin{cases} E & \text{if } r_{f|_C}(i) \text{ is an even solution,} \\ O & \text{if } r_{f|_C}(i) \text{ is an odd solution.} \end{cases}$$

Example 4. Let us consider the fitness function f of Example 3 and $S = \{2, 3\}$. Then, the word of the ranking defined by S is:

$$r_f = \begin{bmatrix} 111 \\ 101 \\ 001 \\ 100 \\ 000 \\ 011 \\ 010 \\ 110 \end{bmatrix} \Longrightarrow W_f^S = \begin{bmatrix} E \\ O \\ O \\ E \\ E \\ E \\ O \\ O \end{bmatrix}.$$

On the other hand, if we are considering the ranking of the solutions that satisfies the constraint C: $x_1 = 0$ (or, equivalently, the function $f|_{x_1=0}$), then the word of the ranking is:

$$r_{f|_C} = \begin{bmatrix} 001 \\ 000 \\ 011 \\ 010 \end{bmatrix} \Longrightarrow W_{f|_C} = \begin{bmatrix} E \\ O \\ O \\ E \end{bmatrix}.$$

Once we have defined the word of a ranking, we present a specific type of word: Dyck Words [16].

Definition 9 (*Dyck Word*)**.** Let W be a word of length 2^n and Δ_i the difference between the number of E and O letters for the first i letters in a word W, $1 \le i \le 2^n$. Then, a word is a Dyck Word, with E (O) as dominant letter, if for any i, $\Delta_i \ge 0$ ($\Delta_i \le 0$).

The Catalan number $Cat_{2^{n-1}}$ is the number of possible Dyck Words of length 2^n with a fixed dominant letter, where $Cat_n = \binom{2n}{n}/(n+1)$.

In the literature, there exist a large number of articles about Dyck Words and equivalent definitions (such as Dyck paths) are also analyzed [17,18].

Example 5. The word W_f from Example 3 is a Dyck Word with E as dominant letter.

$$r_f = \begin{bmatrix} 111 \\ 101 \\ 001 \\ 100 \\ 000 \\ 011 \\ 010 \\ 110 \end{bmatrix} \Longrightarrow W_f = \begin{bmatrix} E \\ O \\ E \\ E \\ O \\ O \\ E \\ O \end{bmatrix} \Longrightarrow \Delta = \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \\ \Delta_4 \\ \Delta_5 \\ \Delta_6 \\ \Delta_7 \\ \Delta_8 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

Once we have defined Dyck Words, we present Proposition 4 which shows rankings that cannot be generated by a $(n-1)$-degree pseudo-Boolean function.

Proposition 4. *Let $n \ge 2$ and $\Omega = \{0,1\}^n$. Let r be a ranking of solutions from Ω and W the word generated by r. If W is a Dyck Word, then r cannot be generated by a $(n-1)$-degree pseudo-Boolean function.*

Proof. By reduction ad absurdum. Let r be a ranking generated by a $(n-1)$-degree pseudo-Boolean function f ($r_f = r$) and W a Dyck Word generated by r with E as dominant letter. By definition of the ranking of solutions and Dyck Words, we can group the solutions by 2^{n-1} different pairs of even–odd solutions, (x_e, x_o), such that $f(x_e) > f(x_o)$ for all pairs. Therefore, $\sum_{x \in \mathcal{E}} f(x) > \sum_{x \in \mathcal{O}} f(x)$ is deduced, which goes against

[1] Note that any ranking of solutions generated by an m-degree pseudo-Boolean function can be generated by a pseudo-Boolean function of degree $m+1$ and, by induction, by a pseudo-Boolean function of degree $n \ge m$.

Theorem 2 with $S = \{1, \dots, n\}$. For Dyck Words with O as dominant letter, we obtain the opposite inequality. □

Observation 6 (*Rankings Generated Only by n-degree Pseudo-Boolean Functions*). *Due to Proposition 4 and Note 1, there exist rankings that cannot be generated by pseudo-Boolean functions of degree $m < n$.*

Proposition 4 shows a necessary condition that any ranking must fulfill to have the possibility of being generated by a $(n-1)$-degree pseudo-Boolean function. Our next step is to check the "opposite direction" of Proposition 4: if the word W of a ranking r is not a Dyck Word, is it possible for r to be generated by a $(n-1)$-degree pseudo-Boolean function?

In order to shed some light on this issue, we first prove the result for $n = 3$. This is done by exhaustively verifying that any ranking of solutions r whose word is not a Dyck Word can be generated by a 2-degree pseudo-Boolean function f. Then, for $n > 3$, we conjecture that the result is true (notice that for $n = 4$, the study of $2^4! \approx 2\cdot 10^{13}$ rankings is not computationally tractable) and, assuming that the conjecture is true, we give the exact number of rankings that cannot be generated by $(n-1)$-degree pseudo-Boolean functions.

Let us present the sufficient result of Proposition 4 when $n = 3$.

Proposition 5. *Let $n = 3$ and $\Omega = \{0,1\}^n$. Let r be a ranking of solutions from Ω and W the word generated by r. If W is not a Dyck Word, then there exists a 2-degree pseudo-Boolean function f whose generated ranking is r ($r_f = r$).*

Let us present a conjecture about the generalization of Proposition 5. For now on, we assume that the following conjecture is true.

Conjecture 6. *Let $n \geq 2$ and $\Omega = \{0,1\}^n$. Let r be a ranking of solutions from Ω and W the word generated by r. If W is not a Dyck Word, then there exists a $(n-1)$-degree pseudo-Boolean function f whose generated ranking is r ($r_f = r$).*

In Appendix, we present two observations which could be helpful to prove Proposition 5 in such a way that it can be extended for any $n \geq 4$ value and $(n-1)$-degree pseudo-Boolean functions, and prove Conjecture 6. The first observation analyzes the coefficients of the 2-degree pseudo-Boolean functions and their impact on the generated ranking of solutions. The second observation studies the fitness function value of the solution 111 and specifies in which positions the solution "can be inserted" to generate a feasible ranking.

Assuming that Conjecture 6 is true, Proposition 4 and Conjecture 6 allow us to count the number of rankings of solutions that cannot be generated by $(n-1)$-degree pseudo-Boolean functions.

Conjecture 7. *Let $n \geq 2$ and $\Omega = \{0,1\}^n$. Then, there are $\frac{2}{2^{n-1}+1} \cdot 2^n!$ rankings that cannot be generated by $(n-1)$-degree pseudo-Boolean functions.*

Proof. There are $Cat_{2^{n-1}}$ Dyck Words with E as dominant letter and the same number of Dyck Words with O as dominant letter. In addition, each E (O) letter of the Dyck Word corresponds to any even (odd) solution, which implies that there are $\left(2^{n-1}!\right)^2$ rankings that generate that particular Dyck Word. Consequently, the number of rankings that cannot be generated by $(n-1)$-degree pseudo-Boolean functions is

$$2 \cdot Cat_{2^{n-1}} \cdot \left(2^{n-1}!\right)^2 = 2 \cdot \frac{2^n!}{2^{n-1}!(2^{n-1}+1)!} \cdot \left(2^{n-1}!\right)^2 = \frac{2}{2^{n-1}+1} \cdot 2^n!. \quad \square$$

Furthermore, because of Note 1, when $\Omega = \{0,1\}^n$, a ranking that cannot be generated by $(n-1)$-degree pseudo-Boolean functions is impossible to be generated by m-degree pseudo-Boolean functions, where $m < n-1$. Consequently, the previous number is also an upper bound of the number of rankings that cannot be generated by pseudo-Boolean functions of degree $m < n-1$. Note that, the proportion of rankings that can only be generated by n-degree pseudo-Boolean functions (functions which fulfill $a_{1\ldots n} \neq 0$) tends to 0 when n tends to infinity.

Example 7. For $n = 3$, the number of rankings that cannot be generated by 2-degree pseudo-Boolean functions is

$$\frac{2}{2^{n-1}+1} \cdot 2^n! = \frac{2}{5} \cdot 8! = 16128.$$

Consequently, for $n = 3$, there are exactly 24 192 possible rankings that can be generated by 2-degree pseudo-Boolean functions out of 40320; that is, 60% of all the possible rankings.

3.3. Study of pseudo-boolean functions of degree $m < n-1$

The presented results up to this point are based on Theorem 2 when $m = n-1$. Our next step is to generalize and study the case of Dyck Words for m-degree pseudo-Boolean functions, where $m < n-1$. This section extends Proposition 4 for any $n \geq 3$ and $m < n-1$. However, this extension shows the necessary condition for a ranking to be generated by a pseudo-Boolean function of degree $m < n-1$, not the sufficient condition.

First, a variation of Definition 9 is presented.

Definition 10 (*Dyck Word Defined by S*). Let us consider Δ_i^S the difference between the number of E and O letters for the first i letters in a word W^S, $1 \leq i \leq 2^n$. Then, the word W^S is a Dyck Word, with E (O) as dominant letter, if for any i, $\Delta_i^S \geq 0$ ($\Delta_i^S \leq 0$).

With Definition 10, we present an extension of Proposition 4.

Lemma 8. *Let $n \geq 3$ and $\Omega = \{0,1\}^n$. Let r be a ranking of solutions from Ω, S a set of binary variables such that $n \geq |S| > m \geq 1$ and W^S the word of the ranking r defined by S. If W^S is a Dyck Word, then r cannot be generated by an m-degree pseudo-Boolean function.*

Proof. The proof of the lemma is analogous to the proof of Proposition 4. □

Bear in mind that Lemma 8 does not focus on a specific set S. Therefore, for a ranking r, there might be more than one possible way to apply Lemma 8 and to prove that r cannot be generated by a pseudo-Boolean function of degree $m \leq n-1$.

In addition, using a similar argument of the proof of Proposition 4, Eqs. (5) and (6) from Corollary 3 can be also used to show new rankings that cannot be generated by m-degree pseudo-Boolean functions. For example, it is possible to define new groups of solutions (such as $G_1 = \mathcal{E}_S \cap \mathcal{O}_{S'}$ and $G_2 = \mathcal{O}_S \cap \mathcal{E}_{S'}$, or $G_1 = \mathcal{E}_S \cap \mathcal{E}_{S'}$ and $G_2 = \mathcal{O}_S \cap \mathcal{O}_{S'}$) and show several new rankings that cannot be generated by a pseudo-Boolean function of degree m.

Example 8. For $n = 4$, the ranking

$$r = [1110 \;\; 1000 \;\; 0100 \;\; 1001 \;\; 0101 \;\; 1111 \;\; 0011 \;\; 0010 \;\; 0000 \;\; 1101 \;\; 0001 \;\; 1100 \;\; 1011 \;\; 1010 \;\; 0110 \;\; 0111]^T$$

cannot be generated by a 2-degree pseudo-Boolean function because for the set $S = \{1,2,3\}$

$$W^S = [E \;\; E \;\; E \;\; E \;\; E \;\; E \;\; E \;\; E \;\; O \;\; O \;\; O \;\; O \;\; O \;\; O \;\; O \;\; O]^T$$

is a Dyck Word.

Once Lemma 8 has been presented, the opposite research question is studied: for $n \geq 3$ and $m < n-1$, can we prove that any ranking which has no Dyck Words defined by all subset of variables S such that $|S| \geq m+1$ can be generated by an m-degree pseudo-Boolean function? In Example 9, a counterexample is presented.

Example 9. Let $n = 4$ and r be the following ranking:

$$r = [0010\ \ 1010\ \ 0001\ \ 1001\ \ 0000\ \ 1101\ \ 1100\ \ 0011\ \ 0101\ \ 0111\ \ 1110\ \ 0110\ \ 1011\ \ 1000\ \ 0100\ \ 1111]^T$$

We will observe that: (a) for any set of variables S such that $|S| \geq 3$, W^S is not a Dyck Word; and (b) the ranking cannot be generated by a 2-degree pseudo-Boolean function.

(a) Let us calculate the words W^S defined by the sets S such that $|S| \geq 3$.

S	W^S
$\{1,2,3,4\}$	$[O\ E\ O\ E\ E\ O\ E\ E\ E\ O\ O\ E\ O\ O\ O\ E]^T$
$\{1,2,3\}$	$[E\ O\ O\ E\ O\ O\ O\ E\ E\ O\ E\ O\ O\ E\ E\ E]^T$
$\{1,2,4\}$	$[O\ E\ E\ O\ O\ E\ O\ E\ O\ O\ O\ E\ O\ E\ E\ E]^T$
$\{1,3,4\}$	$[E\ O\ E\ O\ O\ O\ E\ O\ E\ O\ O\ E\ E\ E\ O\ E]^T$
$\{2,3,4\}$	$[E\ E\ E\ E\ O\ O\ E\ O\ O\ E\ O\ O\ O\ O\ O\ E]^T$

Therefore, for any set S such that $|S| \geq 3$, the presented ranking has no Dyck Words defined by S.

(b) Let $S = \{1,2,3\}$ and $S' = \{1,2,4\}$. By Corollary 3, if a 2-degree pseudo-Boolean function can generate r, then

$$\sum_{x \in \mathcal{E}_S \cap \mathcal{E}_{S'}} f(x) = \sum_{x \in \mathcal{O}_S \cap \mathcal{O}_{S'}} f(x)$$

must be fulfilled. However,

$\mathcal{E}_S \cap \mathcal{E}_{S'} = \{0011, 1000, 0100, 1111\}$ and

$\mathcal{O}_S \cap \mathcal{O}_{S'} = \{0000, 1100, 0111, 1011\}$

and, by definition of the ranking r,

$f(0000) > f(0011)$, $f(1100) > f(1000)$, $f(0111) > f(0100)$ and $f(1011) > f(1111)$.

Consequently,

$$\sum_{x \in \mathcal{E}_S \cap \mathcal{E}_{S'}} f(x) < \sum_{x \in \mathcal{O}_S \cap \mathcal{O}_{S'}} f(x),$$

which implies that r cannot be generated by a 2-degree pseudo-Boolean function.

In addition, based on Corollary 3, the following result is obtained.

Corollary 9. *Let $n \geq 3$ and $j \in \{1, \dots, n\}$. Let r be a ranking and $S = \{1, \dots, n\} \backslash \{j\}$ a set of variables. Let $C : x_j = c_j$ be a constraint and $W_{f|_C}$ the word of the ranking generated by $f|_C$. If $W_{f|_C}$ is a Dyck Word, then r cannot be generated by a $(n-2)$-degree pseudo-Boolean function.*

Proof. Let $S' = \{1, \dots, n\}$. Then, by Corollary 3,

$$\sum_{x \in \mathcal{E}_S \cap \mathcal{O}_{S' \backslash S}} f(x) = \sum_{x \in \mathcal{O}_S \cap \mathcal{O}_{S' \backslash S}} f(x) \quad \text{and} \quad \sum_{x \in \mathcal{E}_S \cap \mathcal{E}_{S' \backslash S}} f(x) = \sum_{x \in \mathcal{O}_S \cap \mathcal{E}_{S' \backslash S}} f(x)$$

$$\iff \sum_{x \in \mathcal{E}_S | x_j = 0} f(x) = \sum_{x \in \mathcal{O}_S | x_j = 0} f(x) \quad \text{and} \quad \sum_{x \in \mathcal{E}_S | x_j = 1} f(x) = \sum_{x \in \mathcal{O}_S | x_j = 1} f(x)$$

Consequently, if $W_{f|_C}$ is a Dyck Word (no matter if $x_j = 0$ or $x_j = 1$), one of the previous equalities is not fulfilled, which implies that r cannot be generated by a $(n-2)$-degree pseudo-Boolean function. □

Therefore, the study of Dyck Words allows us to recognize rankings that cannot be generated by any pseudo-Boolean function of degree $m < n-1$. The reason for not obtaining the opposite result of Lemma 8 (similar to the case of $n = 3$ with Proposition 5) is that, for all $|S| > m$, Lemma 8 studies each equality of Eq. (3) of Theorem 2 independently, whereas Theorem 2 ensures that all the equalities of Eq. (3) are fulfilled at the same time.

To remark the dissimilarities between Lemma 8 and Corollary 3, we present Example 10 for 3-dimensional linear pseudo-Boolean functions.

Example 10. Let $n = 3$ and $m = 1$. In this example, we present: (a) the number of rankings of solutions that cannot be discarded by Lemma 8 (the upper bound of the number of possible rankings that can be generated by linear pseudo-Boolean functions); and (b) the number of rankings of solutions that can be generated by linear pseudo-Boolean functions (counted by Corollary 3).

In the following table, we show if a solution is even or odd according to a set S.

S	111	110	101	100	011	010	001	000
$\{1,2,3\}$	E	O	O	E	O	E	E	O
$\{1,2\}$	E	E	O	O	O	O	E	E
$\{1,3\}$	E	O	E	O	O	E	O	E
$\{2,3\}$	E	O	O	E	E	O	O	E

(a) Lemma 8 (or equivalently the equalities of Eq. (3) of Theorem 2) implies that any linear pseudo-Boolean function cannot have any Dyck Word over the sets S such that $|S| \geq 2$ (deduced from the following 4 equalities):

$(a.1)$ $\sum_{x \in \mathcal{E}_{\{1,2,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2,3\}}} f(x)$;
$(a.2)$ $\sum_{x \in \mathcal{E}_{\{1,2\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2\}}} f(x)$;
$(a.3)$ $\sum_{x \in \mathcal{E}_{\{1,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,3\}}} f(x)$;
$(a.4)$ $\sum_{x \in \mathcal{E}_{\{2,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{2,3\}}} f(x)$.

To bound the number of rankings, we have counted the rankings that generate a Dyck Word over one, two, three and four sets of variables ($\{1,2,3\}$, $\{1,2\}$, $\{1,3\}$ and $\{2,3\}$), and then we apply the inclusion-exclusion principle.

$$|\{\text{Rankings generated by linear pseudo-Boolean functions}\}| \leq \sum_{I \subseteq \{1,\dots,4\}} (-1)^{|I|} \left|\cap_{i \in I} R'_i\right| = 40320 - 64512 + 30720 - 4032 = 2496,$$

where R'_i is the set of rankings that fulfills the equality $(a.i)$.

(b) Corollary 3 implies that any linear pseudo-Boolean function must fulfill all the following 12 equalities:

$(b.1)$ $\sum_{x \in \mathcal{E}_{\{1,2,3\}} \cap \mathcal{O}_{\{1,2\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2,3\}} \cap \mathcal{E}_{\{1,2\}}} f(x)$;
$(b.2)$ $\sum_{x \in \mathcal{E}_{\{1,2,3\}} \cap \mathcal{E}_{\{1,2\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2,3\}} \cap \mathcal{O}_{\{1,2\}}} f(x)$;
$(b.3)$ $\sum_{x \in \mathcal{E}_{\{1,2,3\}} \cap \mathcal{O}_{\{1,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2,3\}} \cap \mathcal{E}_{\{1,3\}}} f(x)$;
$(b.4)$ $\sum_{x \in \mathcal{E}_{\{1,2,3\}} \cap \mathcal{E}_{\{1,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2,3\}} \cap \mathcal{O}_{\{1,3\}}} f(x)$;
$(b.5)$ $\sum_{x \in \mathcal{E}_{\{1,2,3\}} \cap \mathcal{O}_{\{2,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2,3\}} \cap \mathcal{E}_{\{2,3\}}} f(x)$;
$(b.6)$ $\sum_{x \in \mathcal{E}_{\{1,2,3\}} \cap \mathcal{E}_{\{2,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2,3\}} \cap \mathcal{O}_{\{2,3\}}} f(x)$;
$(b.7)$ $\sum_{x \in \mathcal{E}_{\{1,2\}} \cap \mathcal{O}_{\{1,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2\}} \cap \mathcal{E}_{\{1,3\}}} f(x)$;
$(b.8)$ $\sum_{x \in \mathcal{E}_{\{1,2\}} \cap \mathcal{E}_{\{1,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2\}} \cap \mathcal{O}_{\{1,3\}}} f(x)$;
$(b.9)$ $\sum_{x \in \mathcal{E}_{\{1,2\}} \cap \mathcal{O}_{\{2,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2\}} \cap \mathcal{E}_{\{2,3\}}} f(x)$;
$(b.10)$ $\sum_{x \in \mathcal{E}_{\{1,2\}} \cap \mathcal{E}_{\{2,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,2\}} \cap \mathcal{O}_{\{2,3\}}} f(x)$;
$(b.11)$ $\sum_{x \in \mathcal{E}_{\{1,3\}} \cap \mathcal{O}_{\{2,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,3\}} \cap \mathcal{E}_{\{2,3\}}} f(x)$;
$(b.12)$ $\sum_{x \in \mathcal{E}_{\{1,3\}} \cap \mathcal{E}_{\{2,3\}}} f(x) = \sum_{x \in \mathcal{O}_{\{1,3\}} \cap \mathcal{O}_{\{2,3\}}} f(x)$.

Similar to (a), we have counted the number of rankings that generate a Dyck Word over i sets of variables, $i = 1, \dots, 12$, and then we apply the inclusion-exclusion principle to count the exact number of rankings that can be generated by linear pseudo-Boolean functions.

$$
\begin{aligned}
&|\{\text{Rankings generated by linear pseudo-Boolean functions}\}| \\
&\quad = \sum_{I\subseteq\{1,\ldots,12\}} (-1)^{|I|} \left|\cap_{i\in I} R_i\right| \\
&\quad = 40320 - 322560 + 1196160 - 2693664 + 4082640 \\
&\qquad - 4368624 + 3368400 - 1875312 + 743376 - 203280 + 36288 \\
&\qquad - 3840 + 192 \\
&\quad = 96.
\end{aligned}
$$

where R_i is the set of rankings that fulfills the equality $(b.i)$.

Example 10 shows why Corollary 3 obtains the exact number rankings that can be generated by pseudo-Boolean functions of degree $m < n$ and shows why Lemma 8 does not. Therefore, Lemma 8 is not a sufficient condition to prove which rankings can be generated. A future work is to take advantage of the analysis of the words of the ranking and study the features that remain to achieve the sufficient condition.

4. Conclusions and future work

Throughout this work, we have presented a characterization of pseudo-Boolean functions of degree $m < n$, where n is the size of the search space, and the necessary conditions of a ranking of solutions to be generated by an m-degree pseudo-Boolean function. For the characterization, according to the parity of zeros of each solution (with respect to a set of variables), we have shown the equalities that any m-degree pseudo-Boolean function must fulfill. Based on those equalities, we present the word of a ranking (defined by a set of variables), we introduce Dyck Words and we present a necessary condition: if the word of a ranking defined by a set S such that $|S| > m$ is a Dyck Word, then the ranking is impossible to be generated by an m-degree pseudo-Boolean function. On the other hand, we present a conjecture about the sufficient condition of a ranking to be generated by a $(n-1)$-degree pseudo-Boolean function. In the Appendix section, we show two observations about the conjecture.

This work presents a promising avenue for the study of instances of any binary-based problem. However, there is still much work to do. Firstly, the conjecture presented in this paper about pseudo-Boolean functions of degree $(n-1)$ must formally be proved. Another possible future work is to get the necessary conditions of a ranking to be generated by an m-degree pseudo-Boolean function ($m < n$) without analyzing exhaustively the system of inequalities defined by a ranking of solutions. From a practical point of view, a procedure/algorithm which computes the words of a ranking and checks if there is a Dyck Word in an efficient way must be done.

In addition, it would be interesting to match the presented characterization of pseudo-Boolean functions with the studies and algorithms presented in the literature to solve COPs (or equivalent problems in other fields such as physics or engineering). For example, in the literature, several techniques have been proposed to redefine some problems and to solve them. Some examples are: the use of slack variables and/or penalty coefficients to reformulate constrained problems such as unconstrained problems, transformation of pseudo-Boolean functions into continuous functions and application of the algorithms designed for continuous problems, and the learning of surrogate models for black-box optimization. All these examples are currently being studied to improve the existing results in the literature. In this scenario, our future work is to consider the characterization of pseudo-Boolean functions and the analysis of Dyck Words in the mentioned areas to encourage future research and improvements. Furthermore, we want to consider the presented characterization to classify problem instances according to their difficulty to be solved and associate instances and algorithms for an efficient resolution.

CRediT authorship contribution statement

Imanol Unanue: Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **María Merino:** Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Jose A. Lozano:** Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This research has been partially supported by the Spanish Ministry of Science and Innovation [PID2022-137442NB-I00, PID2019-104933GB-I00 funded by MCIN/AEI/10.13039/501100011033] and BCAM Severo Ochoa accreditation, Spain [CEX2021-001142-S]; the Basque Government, Spain [BERC 2022–2025. IT1504-22, IT1494-22]; and the University of the Basque Country UPV/EHU [GIU20/054]. Open Access funding provided by University of Basque Country.

Appendix. Observations about Conjecture 6

In this appendix, we present two observations about Conjecture 6. In Appendix A.1, we explain the influence of the coefficients of the 2-degree pseudo-Boolean functions to generate rankings of solutions. In Appendix A.2, we analyze the words of the rankings to study which the feasible rankings are. We believe that these observations could be helpful to prove Conjecture 6 without an exhaustive verification.

A.1. Analysis of the coefficients of 2-degree pseudo-boolean functions

Let us consider that the proof of Conjecture 6 could be done by induction. Let us consider the case $n = 3$ and f a pseudo-Boolean function of degree $m = 2$. Let us analyze the coefficients of f ($\{a_1, a_2, a_3, a_{12}, a_{13}, a_{23}\}$) and their influence to generate rankings of solutions. For a better comprehension of the argument below, Fig. A.1 shows the geometric relations between coefficients, solutions, their parity and fitness function values. Note that the 3-dimensional representation could be extended for larger dimensions. The fitness function values of the 8 solutions are the following:

x	$f(x)$
000	0
100	a_1
010	a_2
001	a_3
110	$a_1 + a_2 + a_{12}$
101	$a_1 + a_3 + a_{13}$
011	$a_2 + a_3 + a_{23}$
111	$a_1 + a_2 + a_3 + a_{12} + a_{13} + a_{23}$

In Fig. A.1(a), we present a cube whose vertexes are all the fitness function values of the solutions and in which two vertexes are connected if the Hamming distance between the solutions is 1; that is to say, if two solutions differ in 1 bit, their fitness function values are connected. In Fig. A.1(b), we show the parity of each solution (considered in Fig. A.1(a)). We observe that, according to the parity of zeros, the graph is a bigraph (perfectly balanced according to the partition $\{\mathcal{E}, \mathcal{O}\}$). Our analysis will focus on the reorderings of the fitness function values (Fig. A.1(a)) and their implications on the generated words (Fig. A.1(b)).

Let us divide the set of coefficients in two groups: the set of coefficients that depend on a single bit, $\{a_1, a_2, a_3\}$; and the set of coefficients that depend on two bits, $\{a_{12}, a_{13}, a_{23}\}$. The analysis starts from the rankings that can be generated with the former group of bits, and then we add the coefficients of the latter group to generate and to study all the rankings generated by pseudo-Boolean functions of degree 2. For each added coefficient, we observe which new rankings are generated and we check if their words are not Dyck Words.

In Fig. A.1(c), we show the coefficients a_1, a_2, a_3 that influence in the fitness function values of Fig. A.1(a). The study of Fig. A.1(c) is analogous to the study of pseudo-Boolean functions of degree 1 ($a_{12} = a_{13} = a_{23} = 0$). From this figure, several observations can be made:

(a) According to each edge, if we define a relative order between two values (if two adjacent vertex values are compared and an inequality is fixed), then the same relative order must be kept for all its parallel edges. The formal definition is the following one: for any a_i and $S \subseteq \{a_j, a_k\}$, if $a_i > 0$ ($a_i < 0$), then $a_i + \sum_{a\in S} a > \sum_{a\in S} a$ ($a_i + \sum_{a\in S} a < \sum_{a\in S} a$). This implies that several relative orderings between even and odd solutions are completely connected to other even and odd solutions (the number of connected relative orderings depends on the cardinality of S). Furthermore, if we want to swap two adjacent solutions in a ranking (reverse a relative order), all the solutions that are dependent on the same relative order must be swapped in the ranking at the same time.
In Fig. A.1(c), we have colored the edges in such a way that the edges of the same color are parallel and therefore they must keep the same fixed relative order.

(b) For any values of the coefficients a_1, a_2, a_3, if a vertex of the cube (Fig. A.1(c)) is the maximum value, then the opposite vertex (the vertex at Hamming distance 3) is the minimum value.

Therefore, if we consider the parity of Fig. A.1(b), there are only four possible words generated for 1-degree pseudo-Boolean functions: $EOOOEEEO$, $EOOEOEEO$, $OEEEOOOE$ and $OEEOEOOE$. In any case, the generated word is not a Dyck Word. In addition, this scenario proves that the number of different rankings that can be generated by 1-degree pseudo-Boolean functions is 96 (12 possible rankings starting from each vertex).

From this scenario, to prove Conjecture 6 for $n = 3$ without an exhaustive verification, it remains to be proved that for any coefficient values a_{12}, a_{13}, a_{23}, the addition of these coefficients to any ranking generated by a pseudo-Boolean function of degree 1 does not generate a Dyck Word. To do so, the new words generated by adding the remaining coefficients one by one are analyzed.

The influence of the coefficients a_{ij} is analogous to the formal definition of the relative orders defined by the coefficients a_i, a_j. For example, for $k \neq i, j$, if $a_i > a_i + a_j + a_{ij}$ ($a_i < a_i + a_j + a_{ij}$), then $a_i + a_k > a_i + a_j + a_k + a_{ij}$ ($a_i + a_k < a_i + a_j + a_k + a_{ij}$).

The first case is to consider any 1-degree pseudo-Boolean function f and to add a coefficient a_{ij}: $g(x) = f(x) + a_{ij}x_i x_j$. In this scenario, the fitness function values of g that differ from f are $g(111)$ (specifically, $g(111) = f(111) + a_{ij}$) and $g(x)$ such that $x_i = x_j = x_k + 1 = 1$ (specifically, $g(x) = f(x) + a_{ij}$). Particularly, the "critical" values in which the addition of the coefficient a_{ij} changes the ranking of solutions are $-a_i$, $-a_j$, $-a_i - a_j$, $a_k - a_i$, $a_k - a_j$ and $a_k - a_i - a_j$. For the first values, a_{ij} causes two swaps in the ranking, whereas the last value implies one swap. It can be observed that all the rankings generated by g have no Dyck Words.

The next step (and the most difficult one) is to consider any 1-degree pseudo-Boolean function f and to add two coefficients a_{ij}, a_{ik}: $g(x) = f(x) + a_{ij}x_i x_j + a_{ik}x_i x_k$. The difficulty of this case, in comparison to the previous case, is that there exist some adjacent swaps in the rankings that are influenced by the sum $a_{ij} + a_{ik}$ and consequently by the value $g(111)$. So, some relative orders might be lost. The final step is to add the three coefficients $\{a_{12}, a_{13}, a_{23}\}$ and combine the previous analyses. This combination would prove that the rankings whose word is not a Dyck Word can be generated by a 2-degree pseudo-Boolean function.

A.2. Construction of rankings without Dyck words

Let $n = 3$. Let f be a 2-degree pseudo-Boolean function defined by $\sum_{i=1}^{2}\binom{3}{i} = 6$ real coefficients and $f(111) = \sum_{x\in\mathcal{O}} f(x) - \sum_{x\in\mathcal{E}\setminus\{111\}} f(x)$. By definition, a ranking r without the solution 111 can always be generated by an appropriate selection of the coefficients (each fitness function value apart from 0 is defined with an independent coefficient that allows the solution to be fixed in the desired position):

x	$f(x)$
000	0
100	a_1
010	a_2
001	a_3
110	$f(100) + f(010) + a_{12}$
101	$f(100) + f(001) + a_{13}$
011	$f(010) + f(001) + a_{23}$
111	$f(110) + f(101) + f(011) - f(100) - f(010) - f(001)$

Moreover, because multiplying all the coefficients by any positive real value keeps the ranking invariant, each ranking can be generated by infinite possible selections of the coefficients. Furthermore, it can be ensured that the difference between the fitness function values of two adjacent solutions (with respect to the ranking) to be higher or lower than a real value. Let us denote r^* as the ranking r without the solution 111 and W^* as the word generated by r^*. Proposition 5 is proved if and only if, for any r^*, by an appropriate selection of the coefficients, the solution 111 can be inserted at any position that generates a ranking r whose word is not a Dyck Word.

Starting from the word W^* of a ranking r^*, first we observe in which positions of r^* the insertion of the solution 111 generates a ranking r such that W is not a Dyck Word. Specifically, the study of the sequence $\Delta_1, \ldots, \Delta_7$ in W^* allows us to know the exact positions where the solution 111 can be inserted to generate r. For any W^*, there are four possible scenarios.

(a) If $\Delta_i < 0$, for all $i \in \{1, \ldots, 7\}$, then inserting the solution 111 at the top of the ranking r^* and defining r, the word W is not a Dyck Word.
(b) If $W^* = [O \;\; E \;\; O \;\; E \;\; O \;\; E \;\; O]^T$, then inserting the solution 111 at any position except for the top and the bottom of the ranking r^* generates a ranking r such that W is not a Dyck Word.
(c) If $W^* \neq [O \;\; E \;\; O \;\; E \;\; O \;\; E \;\; O]^T$ and there exists an integer i such that $\Delta_i = -1$ and $\Delta_{i+1} = 0$, then inserting the solution 111 at any position $j \leq i + 2$ of the ranking r^* generates a ranking r such that W is not a Dyck Word.
(d) If there exists at least one value $i \in \{1, \ldots, 5\}$ such that $\Delta_i = 1$, $\Delta_{i+1} = 0$ and $\Delta_{i+2} = -1$, then inserting the solution 111 at any position $j \geq i + 2$ of the ranking r^* generates a ranking r whose word is not a Dyck Word.

In Fig. A.2, one example of each of the mentioned scenarios is displayed. In Fig. A.2(a), the Δ_i values of r^* are negative values, and inserting the solution 111 at the top of the ranking, we generate r whose word is not a Dyck Word. In Fig. A.2(b), inserting the solution 111 in the fifth position we generate r such that W is not a Dyck Word. In Fig. A.2(c), $\Delta_3 = -1$ and $\Delta_4 = 0$, so inserting the solution 111 in the fifth position we generate r whose word is not a Dyck Word. Finally, in Fig. A.2(d), $\Delta_1 = 1$, $\Delta_2 = 0$ and $\Delta_3 = -1$, which implies that inserting the

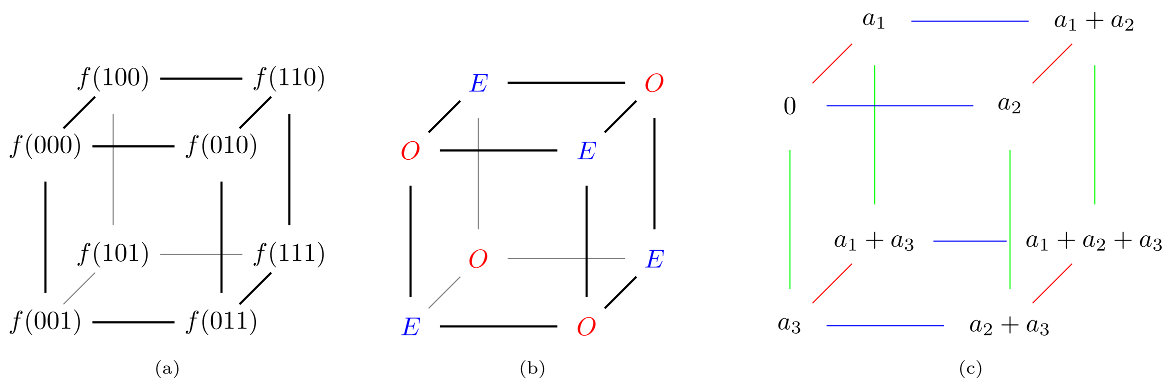

Fig. A.1. Graphical representation for the case $n = 3$. (*a*) Fitness function values and edges defined by the Hamming distance; (*b*) Parity of the solutions; (*c*) Exact fitness function values when $m = 1$.

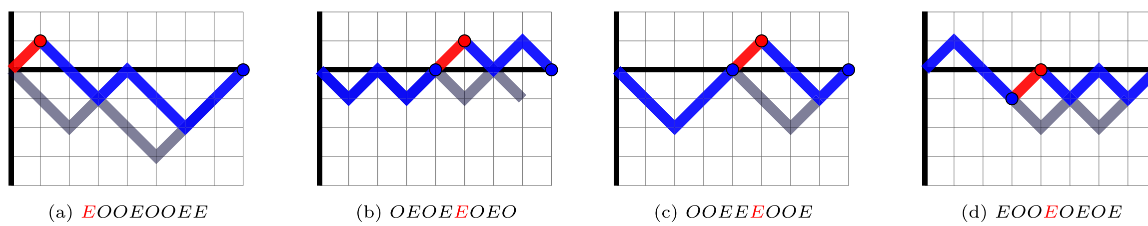

Fig. A.2. Possible r^* scenarios and how to generate a ranking r without a Dyck Word.

solution 111 in the fourth position, the word of the generated ranking r is not a Dyck Word.

Consequently, for any r^*, there always exists a position to insert 111 and to generate a ranking r whose word is not a Dyck Word. In addition, if inserting the solution 111 at the top or at the bottom of the ranking r^* generates a ranking r such that W is not a Dyck Word, then the solution 111 can be inserted at any position of the ranking r^* and will generate a ranking without a Dyck Word.

Finally, it remains to be explained why the solution 111 can be inserted in the desired position to generate the ranking r (such that W is not a Dyck Word): that is to say, why the ranking r is possible to be generated. As previously mentioned, the facts that multiplying the coefficients keeps the same ranking and that we can ensure a minimum distance between two adjacent values allow us to increase or decrease some specific coefficients without changing r^* and to increase or decrease the value $f(111)$.

References

[1] R.K. Ahuja, K. Mehlhorn, J. Orlin, R.E. Tarjan, Faster algorithms for the shortest path problem, J. ACM 37 (2) (1990) 213–223, http://dx.doi.org/10.1145/77600.77615.

[2] G. Laporte, The traveling salesman problem: An overview of exact and approximate algorithms, European J. Oper. Res. 59 (2) (1992) 231–247, http://dx.doi.org/10.1016/0377-2217(92)90138-Y.

[3] G.A. Kochenberger, F. Glover, B. Alidaee, C. Rego, A unified modeling and solution framework for combinatorial optimization problems, Or Spectr. 26 (2) (2004) 237–250, http://dx.doi.org/10.1007/s00291-003-0153-3.

[4] G. Kochenberger, J.-K. Hao, F. Glover, M. Lewis, Z. Lü, H. Wang, Y. Wang, The unconstrained binary quadratic programming problem: A survey, J. Comb. Optim. 28 (1) (2014) 58–81, http://dx.doi.org/10.1007/s10878-014-9734-0.

[5] C.H. Papadimitriou, K. Steiglitz, Combinatorial Optimization: Algorithms and Complexity, Prentice Hall, 1982.

[6] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE Trans. Evol. Comput. 1 (1) (1997) 67–82, http://dx.doi.org/10.1109/4235.585893.

[7] L. Hernando, A. Mendiburu, J.A. Lozano, Characterising the rankings produced by combinatorial optimisation problems and finding their intersections, in: Proceedings of the Genetic and Evolutionary Computation Conference, 2019, pp. 266–273, http://dx.doi.org/10.1145/3321707.3321843.

[8] J. Ceberio, A. Mendiburu, J.A. Lozano, Are we generating instances uniformly at random? in: 2017 IEEE Congress on Evolutionary Computation, CEC, IEEE, 2017, pp. 1645–1651, http://dx.doi.org/10.1109/CEC.2017.7969499.

[9] L. Hernando, A. Mendiburu, J.A. Lozano, Journey to the center of the linear ordering problem, in: Proceedings of the 2020 Genetic and Evolutionary Computation Conference, 2020, pp. 201–209, http://dx.doi.org/10.1145/3377930.3390241.

[10] A. Elorza, L. Hernando, J.A. Lozano, Characterizing permutation-based combinatorial optimization problems in Fourier space, Evol. Comput. (2023) 1–37, http://dx.doi.org/10.1162/evco_a_00315.

[11] E. Boros, P.L. Hammer, Pseudo-boolean optimization, Discrete Appl. Math. 123 (1–3) (2002) 155–225, http://dx.doi.org/10.1016/S0166-218X(01)00341-9.

[12] P.L. Hammer, S. Rudeanu, Boolean Methods in Operations Research and Related Areas, Vol. 7, Springer Science & Business Media, 2012, http://dx.doi.org/10.1007/978-3-642-85823-9.

[13] T.M. Thompson, From Error-Correcting Codes Through Sphere Packings to Simple Groups, Mathematical Association of America, 1983, http://dx.doi.org/10.5948/UPO9781614440215.

[14] E. Boros, Y. Crama, E. Rodríguez-Heck, Compact quadratizations for pseudo-boolean functions, J. Comb. Optim. 39 (3) (2020) 687–707, http://dx.doi.org/10.1007/s10878-019-00511-0.

[15] I.S. Reed, A Class of Multiple-Error-Correcting Codes and the Decoding Scheme, Tech. Rep. 4, Massachusetts Inst of Tech Lexington Lincoln Lab, 1953, pp. 38–49, http://dx.doi.org/10.1109/TIT.1954.1057465.

[16] E. Deutsch, Dyck path enumeration, Discrete Math. 204 (1) (1999) 167–202, http://dx.doi.org/10.1016/S0012-365X(98)00371-9.

[17] P. Duchon, On the enumeration and generation of generalized dyck words, Discrete Math. 225 (1–3) (2000) 121–135, http://dx.doi.org/10.1016/S0012-365X(00)00150-3.

[18] K. Manes, I. Tasoulas, A. Sapounakis, P. Tsikouras, Counting pairs of noncrossing binary paths: A bijective approach, Discrete Math. 342 (2) (2019) 352–359, http://dx.doi.org/10.1016/j.disc.2018.10.016.