Quadratic Boolean functions on the cube

Yuval Filmus

March 27, 2022

Which functions $f\colon \{-1,1\}^n \to \{-1,1\}$ have degree at most 2 and depend on all coordinates? In this short note, we will provide a complete list, up to *NPN equivalence*. This means that two functions are considered equivalent if they differ by (i) negation of some of the inputs, (ii) permutation of the inputs, (iii) negation of the output.

We start with the following observation. If $f\colon \{-1,1\}^n \to \{-1,1\}$ has degree 2 then the function

$$g(x_1, \ldots, x_n, u) = f(x_1 u, \ldots, x_n u)$$

also has degree 2 (since $(x_i u)(x_j u) = x_i x_j$), and is moreover even. Therefore it suffices, as a first step, to classify all even functions. We will do so up to *NN equivalence*, that is, not allowing permutation of the inputs.

Classifying all even functions Suppose that $f\colon \{-1,1\}^n \to \{-1,1\}$ is an even degree 2 function depending on all coordinates. Then

$$f = c + \sum_{ij} c_{ij} x_i x_j,$$

where $c^2 + \sum_{ij} c_{ij}^2 = 1$ by Parseval's identity. We can describe the quadratic part of f as a graph G_f on the vertex set $\{1, \ldots, n\}$, where (i,j) exists if $c_{ij} \neq 0$, and has weight c_{ij}.

Since f depends on all coordinates, G_f has no isolated vertices. We claim that it is also connected. If not, we can write $f = f_1 + f_2$, where each of f_1, f_2 is a non-constant function, and f_1, f_2 depend on disjoint sets of coordinates. Since f_1 is not constant, we can find inputs under which it evaluates to at least two different values, $a_1 < b_1$. Similarly, f_2 evaluates to $a_2 < b_2$. Since f_1, f_2 depend on different coordinates, f evaluates to the three different values $a_1 + a_2 < a_1 + b_2 < a_2 + b_2$, and so cannot be Boolean.

The key to the classification is determining what the neighborhood of a vertex can look like. Consider the function $D_i f = \sum_j c_{ij} x_j$. A simple calculation shows that $D_i f(x) = [f(x|_{i \leftarrow 1}) - f(x|_{i \leftarrow -1})]/2$, where $x|_{i \leftarrow b}$ is obtained from x by setting $x_i = b$. In particular, $D_i f$ is $\{-1,0,1\}$-valued. A simple case analysis shows that $D_i f$ is one of the following functions:

$$\pm x_j, \frac{\pm x_j \pm x_k}{2},$$

where the two signs in the second case are independent. Accordingly, in the graph G_f, for every vertex i, one of the following two cases holds:

(a) v has degree 1, and its adjacent edge has label ± 1.

(b) v has degree 2, and its adjacent edges have labels $\pm \frac{1}{2}$.

The two types do not mix, and so since G_f is connected, it is either empty, or consists of a single edge, or of a cycle of length $\ell \geq 3$. In the latter case, there are exactly ℓ many non-zero c_{ij}'s, each satisfying $c_{ij}^2 = \frac{1}{4}$, and so $\ell \leq 4$.

G_f is empty In this case $f = 1$ up to NN equivalence.

G_f is an edge In this case Parseval's identity shows that $c = 0$. Therefore $f(x, y) = xy$ up to NN equivalence. This is the *XOR* function, if we switch from encoding Boolean values using $\{-1, 1\}$ to encoding them using $\{0, 1\}$.

G_f is a triangle The effect of negating an input is negating the weights of the two adjacent edges in G_f, and the effect of negating the output is negating all weights. Applying these two operations, we can make all weights positive, and so $f(x, y, z) = c + \frac{xy+xz+yz}{2}$ up to NN equivalence. Parseval's identity implies that $c = \pm\frac{1}{2}$. Considering the input $x = y = z = 1$, we see that $c = -\frac{1}{2}$. This is the *all-equal* function: $f(x, y, z) = [x = y = z]$, where $[B] = 1$ if B holds and $[B] = -1$ otherwise.

G_f is a square In this case, Parseval's identity shows that $c = 0$. There are three different possible graphs: using x, y, z, w for the vertices, these are the cycles $xyzw, xzyw, zxyw$. Suppose for definiteness that the cycle is $xyzw$. If all edge signs are the same, say all are positive, then $f(1, 1, 1, 1) = \frac{4}{2} = 2$, which is impossible. By possibly negating f, we can assume that exactly three signs are positive. Applying input negations, we can ensure that the negative edge is xw. Therefore for the graph $xyzw$, we have $f(x, y, z, w) = \frac{xy+yz+zw-xw}{2}$ up to NN equivalence. This is the *4-sortedness* function: $f(x, y, z, w) = [x \le y \le z \le w \text{ or } x \ge y \ge z \ge w]$. The other two options differ on the position of z.

Classifying all functions Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be an arbitrary degree 2 function depending on all coordinates. If f is not even then $g(x_1, \ldots, x_n, u) = f(x_1 u, \ldots, x_n u)$ depends on all coordinates. Above we have classified g up to NN equivalence. Since g is even, we can assume that u is not negated. Therefore we can derive f, up to NN equivalence, by going over all Boolean even degree 2 functions g, and substituting 1 in their last coordinate:

- If $g = xy$ then $f = x$ is the *identity* function.
- If $g = [x = y = z]$ then $f = [x = y = 1]$ is the *OR* function.
- If $g = [x \le y \le z \le w \text{ or } x \ge y \ge z \ge w]$ then $f = [x \le y \le z]$ is the 3-*sortedness* function. There are two other similar cases: $f = [x \le z \le y]$ and $f = [z \le x \le y]$.

In the last case, we can also describe f as a *depth* 2 *decision tree*: if $y = 1$ then $f = z$, and if $y = -1$ then $f = -x$.

Summarizing, here is a complete list of all Boolean degree 2 functions depending on all coordinates, up to NPN equivalence:

- $f = 1$.
- $f = x$, the identity function.
- $f = xy$, the XOR function.
- $f = [x = y = 1]$, the OR function.
- $f = [x = y = z]$, the all-equal function.
- $f = [x \leq y \leq z]$, the 3-sortedness function.
- $f = [x \leq y \leq z \leq w \text{ or } x \geq y \geq z \geq w]$, the 4-sortedness function.