Recursive Binary Sequences of Differences

Robert M. Richman
Mount Saint Mary's College,
Emmitsburg, MD 21727–7799

A subset of the Walsh functions is identified, its properties are explored, and algorithms are given that generate the functions. Each function is divided into 2^n intervals and is indexed by the integer n. It is shown that the nth function can be used to generate an nth derivative and the importance of this is illustrated for solving a class of minimization problems.

1. Introduction

A wide variety of problems exist requiring the identification of the best binary sequence from a choice of many. For example, consider the fairest way for "captain A" and "captain B" to choose sides for a pick-up game of basketball. It is traditional to alternate choices and if eight other players are available, captain A gets the first, third, fifth, and seventh choices, while captain B gets the second, fourth, sixth, and eighth choices. But is this sequence likely to result in the most equitable distribution of talent? The somewhat surprising answer is no: a closer game is likely if captain A has the first, fourth, sixth, and seventh choices, while captain B has the second, third, fifth, and eighth choices.

This inquiry began as an attempt to establish with mathematical rigor the optimal sequence for a class of problems that this exemplifies. A second example, called the "coffeepot problem," is considered in detail after the mathematics have been developed. But as is typically the case with fundamental contributions, scientifically significant applications of this work may not appear for some time.

2. The FCN functions

Consider the set of all functions over the domain (0,1) satisfying the following conditions.

1. The domain is divided into 2^n intervals of equal length, where n is a nonnegative integer. The intervals are numbered from $j = 0$ to $j = 2^n - 1$.

2. The value of the function over each interval is ± 1.

3. The value of the function over the first interval is $+1$.

4. Except for $n = 0$, which has only one interval, every function has equal numbers of intervals valued at $+1$ and -1.

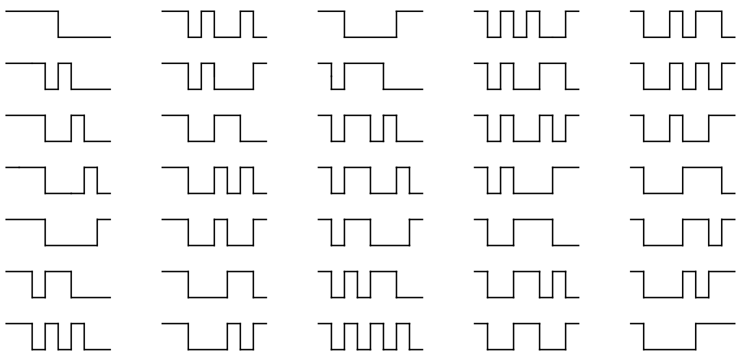


Figure 1. The functions $\mathrm{FCN}(3,i,x)$. No attempt has been made to establish an algorithm to assign the 35 values of i uniquely to the 35 graphs.

For $n = 0$, there is one function in the set. For all other n there are $2^n!/2(2^{n-1}!)^2$ functions in the set. The value of the ith function on the jth interval for a given value of n is designated $\mathrm{FCN}(n,i,j)$. The value at point x is $\mathrm{FCN}(n,i,x)$. Graphs of the 35 functions $\mathrm{FCN}(3,i,x)$ are shown in Figure 1.

3. The Walsh functions

A subset of the FCN functions was described by Walsh in 1923 [1]. Walsh functions enjoyed a resurgence of popularity in the 1960s as an alternative to Fourier analysis for decomposing electrical signals [2] and in the 1980s in the development of the theory of wavelets [3].

The Walsh functions are a closed, mutually orthogonal set in which each function can take on only two values, ±1 [1, 2, 4]. The first 16 Walsh functions are shown in Figure 2. The top two functions have $n = 0$ and 1, respectively. Thereafter, an index k is also required to specify every Walsh function, according to the definitions [1]:

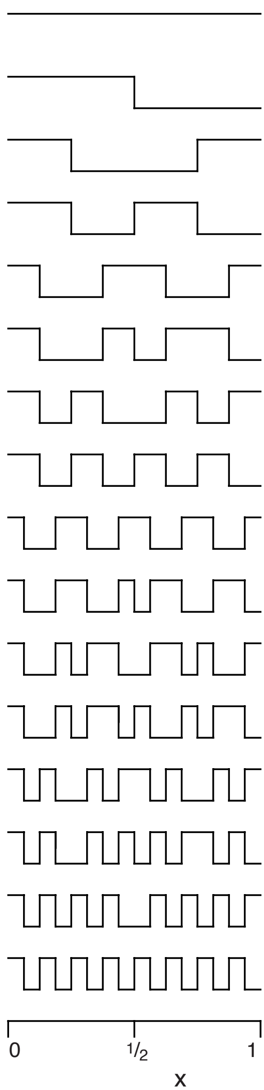
$$\mathrm{WAL}(n+1,2k-1,x) = \begin{cases} \mathrm{WAL}(n,k,2x) & 0 \le x < \frac{1}{2}, \\ (-1)^{k+1}\mathrm{WAL}(n,k,2x-1) & \frac{1}{2} < x \le 1 \end{cases} \tag{1}$$

$$\mathrm{WAL}(n+1,2k,x) = \begin{cases} \mathrm{WAL}(n,k,2x) & 0 \le x < \frac{1}{2}, \\ (-1)^{k}\mathrm{WAL}(n,k,2x-1) & \frac{1}{2} < x \le 1 \end{cases} \tag{2}$$

with $k = 1,2,3,\ldots,2^{n-1}$ and $n = 1,2,3,\ldots$.

The ordering in Figure 2 is that originally suggested by Walsh [1]. Called *sequency order*, the functions are arranged in ascending order of sign changes, often called *zero crossings* [2].

The Rademacher functions are obtained by using only equation (2) instead of incrementing through all allowable values of k. An orthogonal function set that is not closed, the Rademacher functions form an



S	D	N	R	O
0	0	0	0	0
1	1	8	1	1
2	3	12		2
3	2	4	2	
4	6	6		
5	7	14		3
6	5	10		
7	4	2	3	
8	12	3		
9	13	11		
10	15	15		4
11	14	7		
12	10	5		
13	11	13		
14	9	9		
15	8	1		

Figure 2. The first 16 Walsh functions (through $n = 4$) with positive phasing. The number of each function is given in S, sequency order; D, dyadic order; and N, natural order. The n indices are also shown for the first five Rademacher functions, R; and DIF functions, O.

important subset of the Walsh functions [5]. The Rademacher functions in Figure 2 are the Walsh functions with sequency numbers 0, 1, 3, 7, and 15. The complete set of Walsh functions may be generated as products of the Rademacher functions with each other, albeit in a different order called *dyadic order* [6].

The Walsh functions can also be generated from Hadamard matrices in what is called *natural order* [7].

4. The difference functions

We define the difference (DIF) functions by alternating equations (1) and (2) as n is incremented, using (2) when n is even and (1) when n is odd. The first five DIF functions ($n = 0, \ldots, 4$) represented as binary sequences are illustrated below. The values of the functions may be +1 or −1, abbreviated as + and −, on the 2^n intervals of equal length over the domain $(0, 1)$:

$$
\begin{array}{l}
+\\
+-\\
+--+\\
+--+-++-\\
+--+-++--++-+--+
\end{array}
$$

$\mathrm{DIF}(n, j)$ is the value of the nth function on the jth interval, where $0 \le j < 2^n$. The DIF functions may be generated recursively from $\mathrm{DIF}(0, 0) = +1$ in any of the following ways.

Translation. Draw the nth DIF function by first compressing the $(n-1)$th DIF function into the domain $(0, 1/2)$, then changing the sign of the $(n-1)$th DIF function and compressing that into the domain $(1/2, 1)$, that is,

$$
\mathrm{DIF}(n, j) = \begin{cases} \mathrm{DIF}(n-1, j) & \text{if } 0 \le j < 2^{n-1}, \\ -\mathrm{DIF}(n-1, j-2^{n-1}) & \text{if } 2^{n-1} \le j < 2^n. \end{cases}
$$

The translation algorithm follows directly from the definition above and illustrates the origin of the name "difference functions."

Reflection. Draw the nth DIF function by first compressing the $(n-1)$th DIF function into the domain $(0, 1/2)$ and then reversing the $(n-1)$th DIF function, changing the sign for odd n, and compressing that into the domain $(1/2, 1)$, that is,

$$
\mathrm{DIF}(n, j) = \begin{cases} \mathrm{DIF}(n-1, j) & \text{if } 0 \le j < 2^{n-1}, \\ (-1)^n \mathrm{DIF}(n-1, 2^n - j - 1) & \text{if } 2^{n-1} \le j < 2^n. \end{cases}
$$

Reflection provides the easiest way to see why the even-numbered sequences are palindromes.

Division. Replace every interval of +1 in the $(n-1)$th DIF function with intervals of +1 and −1 in the nth DIF function, and replace every −1 with −1 and +1, that is,

$$\mathrm{DIF}(n,j) = \begin{cases} \mathrm{DIF}\left(n-1, \frac{j}{2}\right) & \text{if } j \text{ is even,} \\ -\mathrm{DIF}\left(n-1, \frac{j-1}{2}\right) & \text{if } j \text{ is odd.} \end{cases}$$

The sense intended is that of cell division, not arithmetic division.

Thus, for example, $\mathrm{DIF}(3,4) = -1$ may be generated by translation as $-\mathrm{DIF}(2,0)$, by reflection as $-\mathrm{DIF}(2,3)$, and by division as $\mathrm{DIF}(2,2)$.

The DIF functions resemble the paper folding (PF) sequences characterized by Davis and Knuth in [8]. By analogy with the PF sequences and as a direct consequence of the division process described above, the initial sequence $\mathrm{DIF}(n,0), \mathrm{DIF}(n,1), \mathrm{DIF}(n,2), \ldots$ is identical to the subsequence $\mathrm{DIF}(n,0)$, $\mathrm{DIF}(n,2)$, $\mathrm{DIF}(n,4), \ldots$ and is the negative of the subsequence $\mathrm{DIF}(n,1)$, $\mathrm{DIF}(n,3)$, $\mathrm{DIF}(n,5), \ldots$. Following Okninski, this property is referred to as *self-similarity* [9]. Thus it may be possible to develop algorithms to generate fractal patterns.

5. Relationships

There are a number of interesting connections among the DIF, Rademacher, and Walsh functions.

For $n = 0$ there is one DIF function, and it is a Walsh function. Each subsequent n supplies 2^{n-1} additional Walsh functions. The Walsh functions through sequency number 2^n form a subset of $\mathrm{FCN}(n,i,j)$.

For each n there is one Rademacher function, forming a subset of the Walsh functions.

Represented as step functions, the first five DIF functions are the Walsh functions in Figure 2 with sequency numbers $0, 1, 2, 5$, and 10. For each n, there is one DIF function, forming a subset of the Walsh functions.

When numbered in dyadic or natural order, the Walsh functions that are Rademacher functions are $0, 1, 2, 4, 8, \ldots$. Expressed in binary notation, $m = \sum_{i=0}^{\infty} m_i 2^i$, these are $0000, 0001, 0010, 0100, 1000, \ldots$, that is, all m such that at most one $m_i = 1$.

When numbered in dyadic order, the Walsh functions that are DIF functions are $0, 1, 3, 7, 15, \ldots$; or, in binary notation, 0000, 0001, 0011, 0111, 1111, …; that is, all m such that if $m_k = 1$, then $m_i = 1$ for all $i < k$.

When numbered in natural order, the Walsh functions that are DIF functions are $0, 8, 12, 14$, and 15 for $n = 4$; or in binary notation, 0000, 1000, 1100, 1110, and 1111; or in general, all m such that if $m_k = 1$, then $m_i = 1$ for all $i > k$.

The complete set of Walsh functions may be generated as products of the DIF functions with one another, in analogy with the Rademacher functions.

The Rademacher functions are obtained from the DIF functions from the expression $\mathrm{RAD}(n,x) = \mathrm{DIF}(n,x)\mathrm{DIF}(n-1,x)$, where the second index of each function is now the point x in the domain $(0,1)$ where each function is to be evaluated. The Walsh sequency numbers corresponding to these functions are additive, but the Walsh dyadic and natural numbers are subtracted. For example, the function numbers corresponding to $\mathrm{RAD}(4,x) = \mathrm{DIF}(4,x)\mathrm{DIF}(3,x)$ are sequency, $15 = 10 + 5$; dyadic, $8 = 15 - 7$; and natural, $1 = 15 - 14$.

The DIF functions may be generated from the Rademacher functions from the rather simple relationship, $\mathrm{DIF}(n,x) = \prod_{i=0}^{n} \mathrm{RAD}(i,x)$. For example, the $n = 4$ DIF function in Figure 2 is seen to be the product of the $n = 0,\ldots,4$ Rademacher functions.

Rademacher functions form a graphical basis for determining the binary digits of an integer [2], so this relationship inspires a fourth algorithm to generate the DIF functions by *bits*.

Express j as a binary number, $j = \sum_{i=0}^{\infty} j_i 2^i$. It follows from previous definitions that n is, at minimum, the number of binary digits required to express j. Add the binary digits $\sum_{i=0}^{n-1} j_i$, if this result is odd, then $\mathrm{DIF}(n,j) = -1$; if even, then $\mathrm{DIF}(n,j) = +1$; that is,

$$\mathrm{DIF}(n,j) = (-1)^{\sum_{i=0}^{n-1} j_i} \qquad \text{if } 0 \le j < 2^n.$$

This is illustrated in Figure 3.

Thus, for example, $\mathrm{DIF}(3,4) = -1$ may be generated by expressing 4 as 100_2, the sum of whose bits is 1.

This is the most elegant of the expressions for the DIF functions. It is the only one that is not iterative, and it shows most clearly that the values of j that exist are independent of n.

It is easy to see the equivalence of the j values generated by bits and by division. Halving an even binary number removes the zero from the

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	10	11	100	101	110	111	1000	1001	1010	1011	1100	1101	1110	1111
0	1	1	2	1	2	2	3	1	2	2	3	2	3	3	4
even	odd	odd	even	odd	even	even	odd	odd	even	even	odd	even	odd	odd	even

Figure 3. Illustration of the generation of the DIF sequences by bits. j values are expressed in base 10 on the first line and in base 2 on the second line. The number of ones in the base 2 representation is shown on the third line, and that number's divisibility by 2 is shown on the fourth line. The pattern in the words *even* and *odd* is that of the DIF sequences.

ones place and moves the remaining digits one place to the right, so the resulting binary number has the same number of ones. Subtracting one from an odd binary number reduces by one the number of ones, then halving the result gives no further change in the number of ones.

6. Difference becomes differential

The first derivative of $f(x)$ with respect to x is defined as [10]:

$$f^{(1)}(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

The second derivative would be

$$f^{(2)}(x) = \lim_{k\to 0} \lim_{h\to 0} \frac{\frac{f(x+h+k)-f(x+k)}{h} - \frac{f(x+h)-f(x)}{h}}{k}.$$

Typically, one sets $k = h$ [11], but if instead we let $k = 2h$, this simplifies to

$$f^{(2)}(x) = \lim_{h\to 0} \frac{f(x+3h) - f(x+2h) - f(x+h) + f(x)}{2h^2}.$$

Similarly, we can define the third derivative as

$$f^{(3)}(x) = \lim_{h\to 0} \frac{\begin{array}{l} f(x+7h) - f(x+6h) - f(x+5h) + f(x+4h) - f(x+3h) \\ \quad + f(x+2h) + f(x+h) - f(x) \end{array}}{8h^3}.$$

We recognize the pattern of $\mathrm{DIF}(1,j)$ in the numerator of $f^{(1)}(x)$, $\mathrm{DIF}(2,j)$ in $f^{(2)}(x)$, and $\mathrm{DIF}(3,j)$ in $f^{(3)}(x)$. Indeed, this method of generating higher order derivatives is analogous to generating higher order DIF functions by division. Hence, we generalize:

$$f^{(n)}(x) = (-1)^n \left[\prod_{i=0}^{n-1} 2^{-i}\right] \lim_{h\to 0} \left[\frac{\sum_{j=0}^{2^n-1} \mathrm{DIF}(n,j) f(x+jh)}{h^n}\right].$$

Due to the ease of generating the DIF functions from binary numbers, this could lead to more efficient algorithms for the numerical computation of nth derivatives.

When $f(x)$ is a polynomial of order n or less, $p_n(x)$, the limit is independent of h, so it is unnecessary to take the limit. Then

$$p_n^{(n)}(x) = (-1)^n \left[\prod_{i=0}^{n-1} 2^{-i}\right] \left[\frac{\sum_{j=0}^{2^n-1} \mathrm{DIF}(n,j) p_n(x+jh)}{h^n}\right].$$

When $f(x)$ is a polynomial of order $n-1$ or less, $p_{n-1}^{(n)}(x) = 0$, so

$$\sum_{j=0}^{2^n-1} \mathrm{DIF}(n,j) p_{n-1}(x+jh) = 0.$$

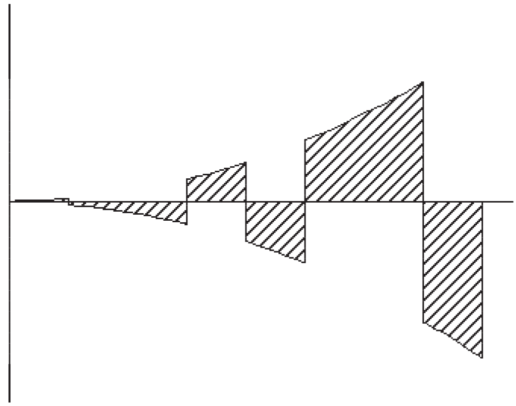


Figure 4. Graphical illustration that $\int_0^1 x^2 \mathrm{DIF}(3,x)\,dx = 0$.

Since this is true for any x along an interval, it is true for the sum of all x along that interval. Hence,

$$\int_0^1 p_{n-1}^{(n)}(x)\,dx = \int_0^1 \mathrm{DIF}(n,x)p_{n-1}(x)\,dx = 0.$$

In other words, the nth order DIF function is orthogonal to polynomials of order $n-1$. This is illustrated in Figure 4.

7. The coffeepot problem

The final result of the previous section has an interesting application. There is a flavor concentration gradient in a carafe of coffee: less and less flavor is available for extraction from the coffee grounds as the brewing process progresses, so the coffee is stronger at the bottom of the carafe than at the top. Swirling produces horizontal rather than vertical convection, so it does not eliminate the concentration gradient. In pouring two cups of coffee of equal volume from this carafe, how can one pour the coffee to most nearly equalize the solute concentrations in the two cups?

Let c be the concentration of interest and x be the height from the bottom of a cylindrical carafe, in units such that the range of x is from 0 to 1. It is likely that the concentration function c satisfies the conditions $c(x) > 0, c'(x) < 0$, and $c''(x) > 0$ over the entire interval $[0,1]$ as shown in Figure 5 (but the solution outlined below does not depend on this assumption). Examples of functions satisfying these conditions are $c \propto e^{-ax}$, $c \propto 1/(a+x)$, and $c \propto 1/(a+x)^2$.

Let us apportion the coffee from the carafe equally into two cups, A and B, and let the notation AB designate a binary sequence of two pours, the top half of the carafe into cup A, then the bottom half into

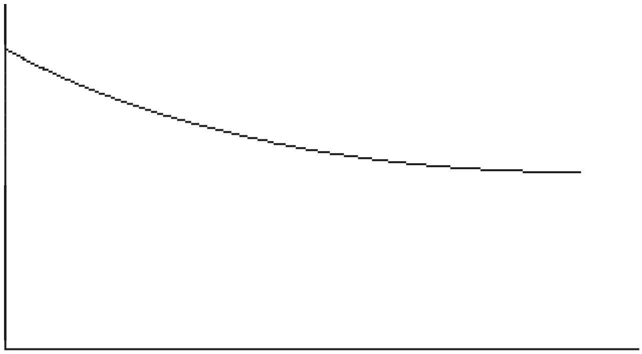


Figure 5. Graphical representation of the likely form of the concentration gradient as a function of height.

cup B. If one has the patience to make four pours of equal volume, the possible pouring sequences are $AABB$, $ABBA$, and $ABAB$. The objective is to make the difference in the two concentrations as close as possible to zero. Subtracting the concentration in cup B from the concentration in cup A is equivalent to multiplying the concentration of cup B by -1 and adding the two. Representing the sequences as step functions, they correspond to the FCN functions with $n = 2$. The best pouring sequence is obtained by minimizing the absolute value of the integral of the product of the concentration function times the sequence step function (Figure 6). That is, the concentration difference between cups A and B is minimized by using the step function $\text{FCN}(2, i, x)$ that minimizes $\left|\int_0^1 \text{FCN}(2, i, x) c(x)\, dx\right|$ over all i. The sequence that best minimizes the difference is $ABBA$, which is $\text{DIF}(2, j)$, whenever $c(x)$ is monotonic, a condition one would normally expect.

If one wishes to further reduce the difference and has more patience, one can make eight pours of equal volume, four in each cup. The number of possible sequences is now 35, as represented in Figure 1. One can write a simple computer program to identify, for a given concentration function $c(x)$, the step function that minimizes $\left|\int_0^1 \text{FCN}(3, i, x) c(x)\, dx\right|$. A Microsoft Excel macro written for this purpose systematically evaluates the integral for each sequence. The process is then repeated as a varies incrementally. The optimal sequence depends on the choice of $c(x)$, but it is $ABBABAAB$, or $\text{DIF}(3, j)$, for $c = e^{-ax}$ whenever $a \leq 1.5$, for $c = 1/(a + x)$ when $a \geq 1.3$, and for $c = 1/(a + x)^2$ when $a \geq 3$.

With even more patience, one may make 16 pours, eight into each cup. There are now 6435 possible pouring sequences. Different sequences minimize $\left|\int_0^1 \text{FCN}(4, i, x) c(x)\, dx\right|$ for different functions $c(x)$, but $ABBABAABBAABABBA$, or $\text{DIF}(4, j)$, has the lowest value for $c = e^{-ax}$ when $a \leq 0.6$, and for $c = 1/(a + x)$ when $a \geq 5$, and for

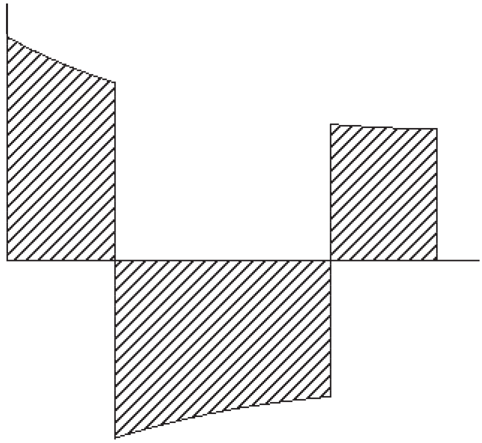


Figure 6. $\int_0^1 \text{DIF}(2,x)c(x)\,dx$, that is, the step function $\text{FCN}(2,i,x)$ that minimizes $\left|\int_0^1 \text{FCN}(2,i,x)c(x)\,dx\right|$.

$c = 1/(a+x)^2$ when $a \geq 8$. We see that the conditions for a become more stringent as n increases, having the effect of flattening the graphs of c as a function of x. Nonetheless, it seems significant that the difference sequence works best over a range of conditions when so many other options are available.

Why do these solutions work for all three forms of the concentration function? And why do they work only under certain conditions?

Consider the Taylor series for each of these three functions:

$$e^{-ax} = \sum_{n=0}^{\infty} \frac{(-ax)^n}{n!}$$
$$\frac{1}{a+x} = \frac{1}{a} \sum_{n=0}^{\infty} \left(\frac{-x}{a}\right)^n$$
$$\frac{1}{(a+x)^2} = \frac{1}{a} \sum_{n=1}^{\infty} n \left(\frac{-x}{a}\right)^{n-1}.$$

When $c(x) = e^{-ax}$, as positive values of a approach zero, the higher order terms in the polynomial expansion decrease in proportion to a^n. When $c(x) = 1/(a+x)$ or $c(x) = 1/(a+x)^2$, increasing values of a make the higher order terms decrease as a^{-n-1} and a^{-n}, respectively. Since $\text{DIF}(3,x)$ is already orthogonal to all polynomial terms through x^2, the integral of the product becomes smaller as terms in x^3 and above decrease more rapidly. Similarly, $\text{DIF}(4,x)$ is already orthogonal to all polynomial terms through x^3, so the integral of the product is minimized when terms in x^4 and above decrease rapidly.

This result may be applicable to more important problems. For example, it is often difficult to distribute pigment uniformly through paint. These functions could be applied to such situations, where homogenization of inhomogeneous mixtures is difficult [12]. It is likely that uniformity from one sample to the next can best be achieved by distribution following the pattern of the DIF functions.

8. Conclusion

The difference (DIF) functions form a mutually orthogonal set that can be used to generate the Walsh and Rademacher functions. The difference sequences are easily generated by several algorithms, emerging as a fundamental property of binary numbers. A simple expression for the nth derivative of a function can be derived in terms of the nth sequence. As a consequence, the nth function is orthogonal to polynomials of order $n-1$. Because of this property, these sequences form solutions to a class of minimization problems.

References

[1] J. L. Walsh, "A Closed Set of Normal Orthogonal Functions," *Ann. J. Math*, **55** (1923) 5–24.

[2] K. G. Beauchamp, *Applications of Walsh and Related Functions* (Academic Press, London, 1984).

[3] G. G. Walter, *Wavelets and Other Orthogonal Systems With Applications* (CRC Press, Boca Raton, 1994).

[4] D. S. Stoffer, "Walsh-Fourier Analysis and Its Statistical Applications," *Journal of the American Statistical Association*, **86** (1991) 461–479.

[5] H. Rademacher, "Einige Sätze von Allgemeinen Orthogonalfunktionen," *Mathematische Annalen*, **87** (1922) 112–138.

[6] G. Alexits, *Convergence Problems of Orthogonal Series* (Pergamon, New York, 1961).

[7] C. K. Rushforth, "Fast Fourier-Hadamard Decoding of Orthogonal Code," *Information Control*, **15** (1969) 33–37.

[8] C. Davis and D. E. Knuth, "Number Representations and Dragon Curves I," *Journal of Recreational Mathematics*, **3** (1970) 66–81.

[9] A. Okninski, "Groups and Dynamical Systems: Discrete Time Dynamics on the E(2) Group," *Physica D*, **55** (1992) 358–367.

[10] T. M. Apostol, *Calculus*, volume 1, second edition (Wiley, New York, 1967).

[11] *User's Guide, Mathcad Plus 6 for Macintosh*, Professional Edition (Mathsoft, Cambridge, MA, 1996).

[12] C. D. Rielly and S. S. H. Burmester, "Homogenization of Liquids with Different Densities and Viscosities in Industrial Mixing Technology: Chemical and Biological Applications," in G. B. Tatterson, volume editor, R. V. Calabrese and W. R. Perrey, volume co-editors (American Institute of Chemical Engineers, Symposium Series; number 299, volume 90, 1994, pp. 175–180).