HAL
open science

Hölder functions and Deception of Genetic Algorithms

Jacques Lévy Véhel, Evelyne Lutton

To cite this version:

Jacques Lévy Véhel, Evelyne Lutton. Hölder functions and Deception of Genetic Algorithms. IEEE Transactions on Evolutionary Computation, 1998, 2 (2), pp.56-71. 10.1109/4235.728208 . inria-00592413

HAL Id: inria-00592413

https://inria.hal.science/inria-00592413

Submitted on 12 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hölder Functions and Deception of Genetic Algorithms

Evelyne Lutton, Jacques Lévy Véhel

Abstract—**We present a deception analysis for Hölder functions. Our approach uses a decomposition on the Haar basis, which reflects in a natural way the Hölder structure of the function. It allows to relate the deception, the Hölder exponent, and some parameters of the genetic algorithms (GAs). These results prove that deception is connected to the irregularity of the fitness function, and shed a new light on the schema theory. In addition, this analysis may assist in understanding the influence of some of the parameters on the performance of a GA.**

Keywords— **Genetic Algorithms, Deception Analysis, Hölder functions, Hölder exponent, Fractals.**

I. Introduction

TWO main factors make the optimization of certain functions difficult: local irregularity (for instance, non-differentiability) resulting in wild oscillations, and the existence of several local extrema. Stochastic optimization methods were developed to tackle these difficulties: one of their characteristic features is that no a priori hypotheses are made on the function to be optimized - no differentiability is required - and the function is not assumed to have only one local maximum (or minimum). This makes stochastic methods useful in numerous "difficult" applications (of course often at the expense of high computation times), as for example in inverse problems appearing in material optimization, image analysis, or process control.

In addition to theoretical investigations about their convergence properties, the main challenge in the field of stochastic optimization is to set the parameters of the methods so that they are the most efficient. This problem is of obvious practical interest but it also yields some theoretical insight on the behaviour of these optimization techniques.

It is difficult to derive rules for tuning the parameters without making any assumption on the studied function. On the other hand, if we are to make restrictive assumptions, they should not rule out "interesting" functions, as for instance non-differentiable functions with many local extrema. In this work, we consider a class of functions which is both quite general, as it includes smooth functions as well as very irregular ones, and sufficiently constrained so as to obtain useful results. This class is that of Hölder functions, whose definition is recalled in section II.

Essentially, Hölder functions are continuous functions which may have, up to a certain amount, wild variations. In particular, many non-differentiable continuous functions, as long as their irregularity can be bounded in a certain sense, belong to this class. Hölder functions cannot in general be optimized through usual, e.g. gradient-based, methods. Some "fractal" functions, as for instance the Weierstrass one (see section II), are Hölder functions which possess infinitely many local extrema. Since such functions motivate the use of stochastic optimization methods, they are a good test to assess their efficiency.

We focus on genetic algorithms (GAs), which belong to the pool of artificial evolution methods, i.e. methods inspired from natural evolution principles, and show that the Hölder framework allows to obtain more specific results. Evolutionary methods in general have been used since about 40 years, and are known as particularly efficient in numerous applications (see [15], [28], [1], [30], [19], [10], [6]). They have been widely studied in various domains, from a theoretical as well as from a practical point of view. Theoretical analyses of GAs are mainly based on two different approaches:

- proofs of convergence based on Markov chain modeling: for example, Davis [7] has established a mutation probability decreasing scheme that ensures the theoretical convergence of the canonical algorithm,
- deceptive functions analysis, based on schema analysis and Holland's original theory [16], [11], [12], [13], which characterizes the efficiency of a GA, and allows shedding light on GA-difficult functions.
  Deception has been intuitively related to the biological notion of epistasis [6], which can be understood as a sort of "nonlinearity" degree. Deception depends on:
  - the parameter setting of the GA,
  - the shape of the function to be optimized,
  - the coding of the solutions, i.e. the "way" of scanning the search space.

In this paper, we concentrate on the deception approach that provides a simple model of the GA behaviour. This model allows for making some computations, as we will see below, that are much more complicated or even infeasible for other GA models. But as schema theory is often considered as disputed and has some known limitations, the practical implications of the analysis presented in this paper have to be considered with care and mainly as "tendency" analyses. However, in [27] a result similar to the schema theorem has been proven with the help of a Markov chain model, i.e. with finite size populations. This new result has characteristics similar (yet more complex) to Holland's formula, and provides a theoretical lower bound to the expected number of representatives of a schema at the next generation with respect to its current number of representative, the parameters of the GA, and the characteristics of the schema to be considered. This result may

INRIA - Rocquencourt - B.P. 105, 78153 LE CHESNAY Cedex, France - Tel : 33 1 39 63 55 23 - Fax : 33 1 39 63 59 95 - email : evelyne.lutton@inria.fr, jacques.levy_vehel@inria.fr
http://www-rocq.inria.fr/fractales/

shed a new light on the validity of some qualitative results derived from the schema theory.

Section III recalls some basic facts about deception analysis. In section IV, a deception analysis is made for Hölder functions, and in section V, we analyze the influence of the parameter on deception. We conclude in section VI with some considerations about the usefulness and the limitations of this analysis.

II. Hölder functions

Definition 1 (Hölder function of exponent h)
Let (X, d_X) and (Y, d_Y) be two metric spaces. A function $F : X \to Y$ is called a *Hölder function of exponent* $h \geq 0$, *if for each* $x, y \in X$ *such that* $d_X(x, y) < 1$, *we have:*

$$d_Y(F(x), F(y)) \leq k.d_X(x, y)^h \qquad (x, y \in X) \qquad (1)$$

for some constant $k > 0$.

The following results are classical:

Proposition 1: If F is Hölder with exponent h, it is Hölder with exponent h' for all $h' \in (0, h]$.

Proposition 2: Let F be a Hölder function. Then F is continuous.

Although a Hölder function is always continuous, it need not be differentiable (see the example of Weierstrass functions below).

Intuitively (see Figures 3 and 4), a Hölder function with a low value of h looks much more irregular than a Hölder function with a high value of h (in fact, this statement only makes sense if we consider the highest value of h for which (1) holds).

The frame of Hölder functions, while imposing a condition that will prove useful for tuning the parameters of the GA, allows us to consider very irregular functions, as the Weierstrass function displayed in Figure 1 and defined by:

$$W_{b,s}(x) = \sum_{i=1}^{\infty} b^{i(s-2)} \sin(b^i x) \qquad (2)$$
$$\text{with } b > 2 \text{ and } 1 < s < 2$$

This function is nowhere differentiable, possesses infinitely many local optima, and may be shown to satisfy a Hölder condition with $h = s$ [9]. For such "monofractal" functions (i.e. functions having the same irregularity at each point), it is often convenient to talk in terms of box dimension d (sometimes referred to as "fractal" dimension), which, in this simple case, is $2 - h$.

Hölder functions appear naturally in some practical situations where no smoothness can be assumed and/or where a fractal behaviour arises (as for example to solve the inverse problem for iterated functions systems (IFS) [26], in constrained material optimization [29], or in image analysis tasks [22], [3]). It is thus important to obtain even very preliminary clues that allow tuning the parameters of a stochastic optimization algorithm, like a GA, in order to perform an efficient optimization on such functions.

Finally note that the well-known "onemax" test-function (i.e. the number of "1s" in the bit string) is a very irregular function that can be considered as the sampling of a Hölder function with $h = 0$, see Figure 2.

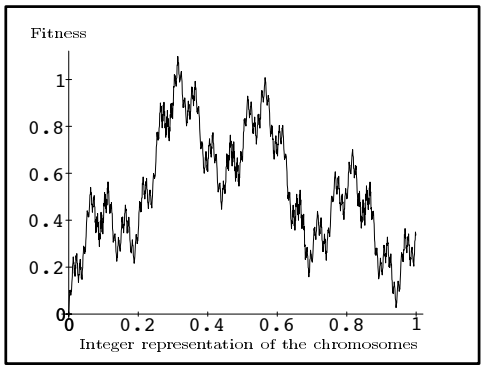

Fig. 1. Weierstrass function of dimension 1.5.

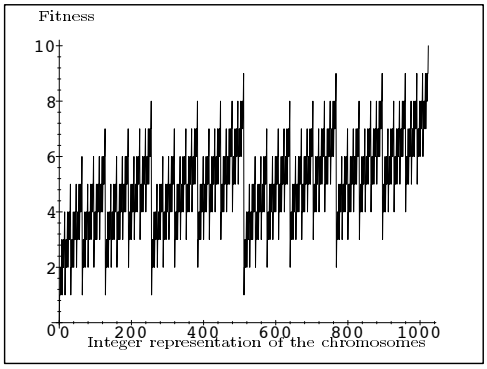

Fig. 2. Onemax function on 10 bits: the sampling of a Hölder function, with $h = 0$ (the abscissa is the usual integer representation of a binary string)

III. Deception Analysis

Our approach is based on Goldberg's deception analysis [11], [12], which uses a decomposition of the function to be optimized, f, on Walsh polynomials. This decomposition allows defining a new function f', which can be understood as a sort of statistic "preference" given by the GA to the points of the search space during the search. This function f' is in some sense an averaged version of f. The GA is said to be deceived when the global maxima of f and f' do not correspond to the same points of the search space.

A. Schema theory

More precisely, this approach is based on the schema theory [10], [16]. A schema represents a subspace of the search space, and quantifies the resemblance between its representing codes: for example the schema $\mathbf{01\star\star11\star0}$ is a subspace of the space of codes of eight bits in length ($\star$ represents a "wild card" that can be 0 or 1).

The GA modelled in schema theory is a canonical GA which acts on binary strings, and for which the creation of a new generation is based on three operators:

- proportional *selection*: the probability that a solution of the current population is selected is proportional to its relative fitness,
- the *genetic operators*: one-point crossover and bit-flip

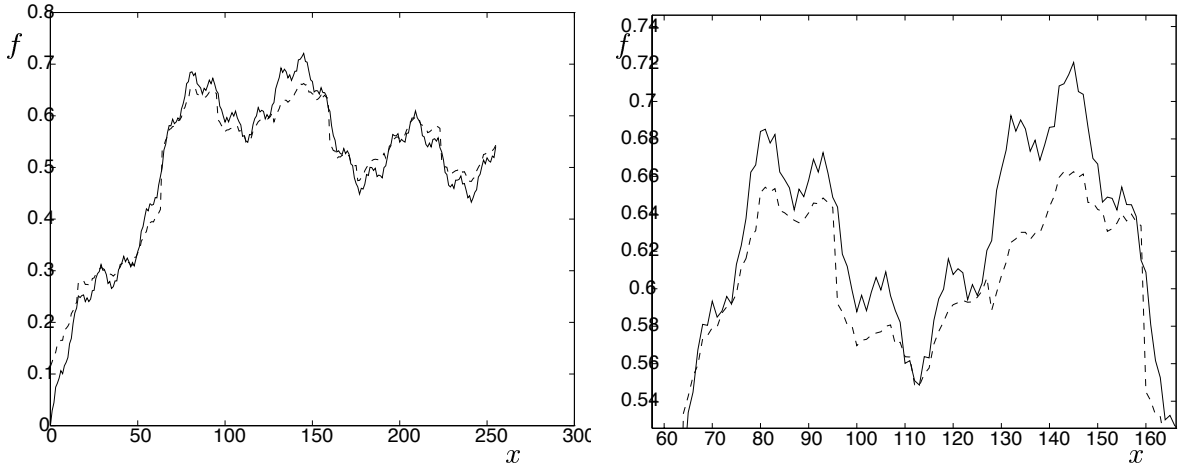

Fig. 3. *Left:* f (continuous) and f' (dotted) for a Weierstrass function of dimension 1.2 sampled on 8 bits. *Right:* zoom on the region of the first two maxima: the function is not 0-deceptive although it is 0.03-deceptive.

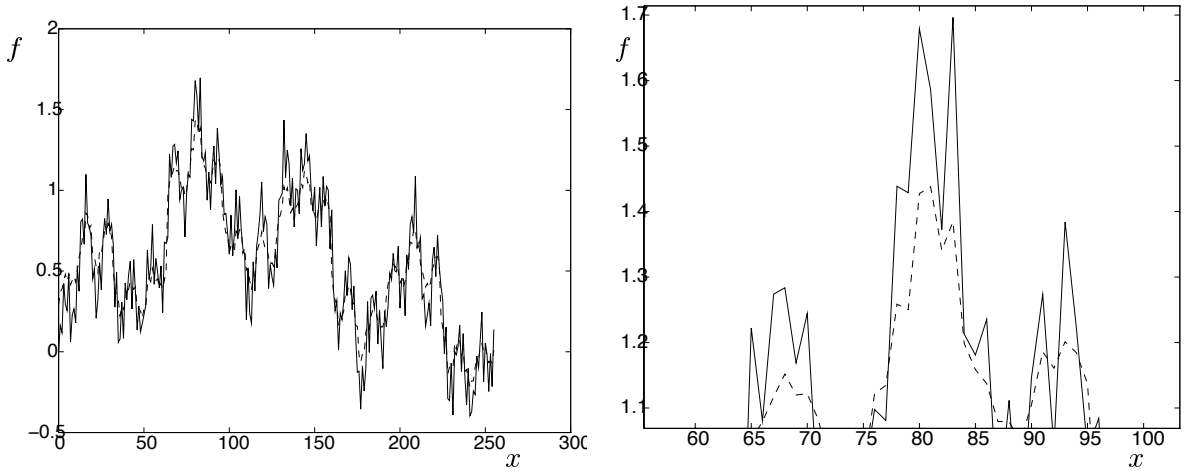

Fig. 4. *Left:* f (continuous) and f' (dotted) for a Weierstrass function of dimension 1.7 sampled on 8 bits. *Right:* zoom on the region of the first two maxima: the function is 0-deceptive although it is not 0.05-deceptive.

mutation, randomly applied, with probabilities p_c and p_m.

Schemata allow representing global information about the fitness function. It has to be understood that schemata are just tools which help to understand the codes' structure. A GA thus works on a population of N codes, and implicitly uses information on schemata that are represented in the current population.

We recall below the so-called "schema theorem" which is based on the observation that the evaluation of a single code allows us to deduce some knowledge about the schemata to which that code belongs.

Theorem 1 (schema theorem) (Holland)
For a given schema H, let:

- $m(H,t)$ be the relative frequency of the schema H in the population of the tth generation,
- $f(H)$ be the mean fitness of the elements of H,
- $\mathcal{O}(H)$ be the number of fixed bits in the schema H, called the *order* of the schema,
- $\delta(H)$ be the distance between the first and the last fixed bit of the schema, called the *definition length* of the schema.
- p_c be the crossover probability,
- p_m be the mutation probability of a gene of the code,
- $\bar{f}$ be the mean fitness of the current population.

Then:

$$m(H,t+1) \geq m(H,t)\frac{f(H)}{\bar{f}}[1 - p_c\frac{\delta(H)}{l-1} - \mathcal{O}(H)p_m]$$

The quantities $\delta(H)$ and $\mathcal{O}(H)$ help to model the influence of the genetic operators on the schema H : the longer the definition length of the schema is, the more frequently it is broken by a crossover (the schema theory has been developed for a one-point crossover). In the same way, the bigger the order of H is, the more frequently H is broken by a mutation.

From a qualitative view point, this formula means that the "good" schemata, having a short definition length and a low order, tend to grow very rapidly in the successive populations. These particular schemata are called *building blocks*.

The usefulness of the schema theory is twofold : first, it supplies some tools to check whether a given representation is well-suited for a GA (by answering the question : does this representation generate efficient bluiding blocks ?). Second, the analysis of the nature of the "good" schemata, using for instance Walsh functions [10], [17], can give some ideas regarding GA efficiency [6], via the notion of deception that we describe below.

B. Walsh polynomials and deception characterization

In order to test if a given function f is easy or difficult to optimize for a GA, one could verify the "building block" hypothesis :

1. identify the building blocks : i.e. compute all the mean fitnesses of the short schemata which are represented within a generation, and identify as building blocks the ones whose representation increases along the evolution,
2. verify whether or not the optimal solution belongs to these building blocks, to know if the building blocks may confuse the GA.

However, this procedure is obviously computationally intractable. Instead, Goldberg [11] has suggested using a method based on a decomposition of f on the orthogonal basis of Walsh functions on $[0..2^l-1]$, where $[0..2^l-1]$ denotes the set of integers of the interval $[0, 2^l-1]$.

On the search space $[0..2^l-1]$, we can define 2^l Walsh polynomials as :

$$\Psi_j(x) = \prod_{t=0}^{l-1}(-1)^{x_t j_t} = (-1)^{\sum_{t=0}^{l-1} x_t j_t} \qquad \forall x, j \in [0..2^l-1]$$

x_t and j_t are the values of the tth bit of the binary decomposition of x and j.

It is well known that these Walsh polynomials form an orthogonal basis of the set of functions defined on $[0..2^l-1]$, and we let $f(x) = \sum_{j=0}^{2^l-1} w_j \Psi_j(x)$ be the decomposition of the function f.

The deception of f is characterized through the function f' [11], [12] defined as follows :

$$f'(x) = \sum_{j=0}^{2^l-1} w'_j \Psi_j(x) \qquad \text{with} \tag{3}$$

$$w'_j = w_j(1 - p_c\frac{\delta(j)}{l-1} - 2p_m\mathcal{O}(j)) \tag{4}$$

The quantities δ and $\mathcal{O}$ are defined for every j in a similar way as for the schemata : $\delta(j)$ is the distance between the first and the last non-zero bits of the binary decomposition of j, and $\mathcal{O}(j)$ is the number of non-zero bits of j.

For $\epsilon \geq 0$ let :

$$\begin{aligned} N_\epsilon &= \{x \in [0..2^l-1]/|f(x) - f^*| \leq \epsilon\} \quad \text{and} \\ N'_\epsilon &= \{x \in [0..2^l-1]/|f'(x) - f'^*| \leq \epsilon' = \frac{f'^* - w_0}{f^* - w_0}\epsilon\} \end{aligned}$$

where f^* (resp. f'^*) is the global optimum of f (resp. f'). Recall that w_0 is the mean value of both f and f'.

Definition 2 (ϵ-deception) f is said to be **ϵ-deceptive** if $N_\epsilon \not\subset N'_\epsilon$.

Remark 1 : ϵ-deception is not monotonic : for some 0-deceptive functions, an ϵ may be found such that the function is not ϵ-deceptive. Reversely, for some non-0-deceptive functions, we may also find an ϵ' such that the function is ϵ'-deceptive. This fact is particularly obvious in Figures 3 and 4.

Remark 2 : ϵ-deception is not strictly equivalent to the notion of deception based on the verification of the building block hypothesis, that is developed for example in [8], where sufficient conditions for deception have been derived.

IV. Haar polynomials for the deception analysis of Hölder functions

In order to perform a valuable deception analysis for Hölder functions, we have to replace the decomposition on the Walsh basis by one that is more suited. This new basis should allow us to relate deception to the irregularity of the Hölder function, i.e. to its Hölder exponent. Indeed, it is intuitively obvious that the more irregular the function is (i.e. the lower the Hölder exponent), the more deceptive it is likely to be. Figures 3 and 4 show f and f' for Weierstrass functions of dimension 1.2 and 1.7, both sampled on eight bits : the Weierstrass function of dimension 1.2 is here not deceptive while the Weierstrass function of dimension 1.7 is deceptive.

There exist simple bases which permit characterizing in a certain sense the irregularity of a function in terms of its decomposition coefficients. Wavelet bases possess such a property.

The wavelet transform (WT) of a function f consists in decomposing it into elementary space-scale contributions, associated to the so-called *wavelets* which are constructed from a single function, the *analyzing wavelet* ϕ, by means of translations and dilations. The WT of the function f is defined as :

$$T_\phi[f](b,a) = \frac{1}{a}\int_{-\infty}^{+\infty} \phi\left(\frac{x-b}{a}\right) f(x)dx$$

where $a \in \mathbf{R}^+$ is a scale parameter and $b \in \mathbf{R}$ is a space parameter. The analyzing wavelet ϕ is a square integrable

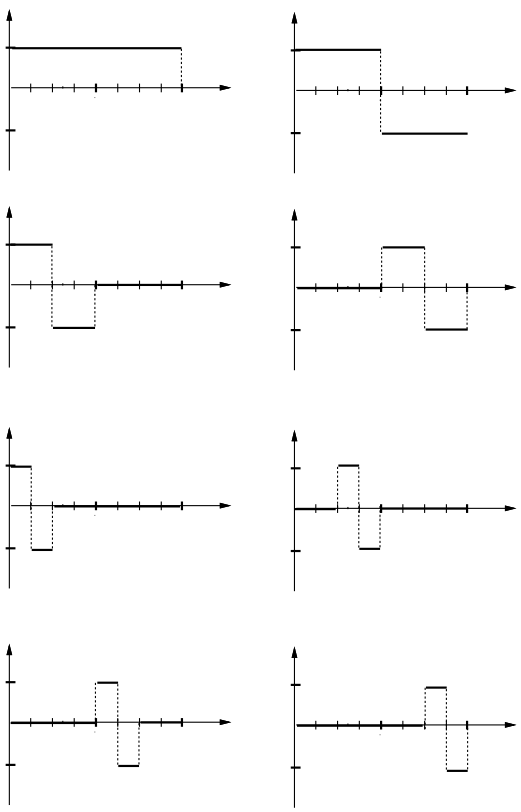

Fig. 5. Haar functions for $l = 3$.

function of zero mean, generally chosen to be well localized in both space and frequency.

Our approach is based on the use of the simplest wavelets, i.e. Haar wavelets, which are defined on the discrete space $[0..2^l - 1]$ as:

$$H_{2^q+m}(x) = \begin{cases} 1 & for \;\; (2m)2^{l-q-1} \leq x < (2m+1)2^{l-q-1} \\ -1 & for \;\; (2m+1)2^{l-q-1} \leq x < (2m+2)2^{l-q-1} \\ 0 & otherwise \;\; in \;\; [0..2^l - 1] \end{cases}$$

with $q = 0, 1, \ldots, l-1$ and $m = 0, 1, \ldots, 2^q - 1$: q is the degree of the Haar function, related to the scale of the wavelet and m corresponds to its localization (see Figure 5).

These functions form an orthogonal basis of the set of functions defined on $[0..2^l - 1]$. Any function f of $[0..2^l - 1]$ can be decomposed as:

$$f(x) = \sum_{j=0}^{2^l-1} h_j H_j(x) \qquad h_j = \frac{1}{2^{l-q}} \sum_{x=0}^{2^l-1} f(x) H_j(x)$$

A. Haar coefficients can be bounded

Suppose that the function f to be optimized is the sampling, with precision $\epsilon = \frac{1}{2^l}$, of a Hölder function F defined on $[0, 1]$:

$$\forall x \in [0..2^l - 1], \qquad f(x) = F(\frac{x}{2^l})$$

Using the definition of the Haar functions $H_j, \;\; j = 2^q + m$, we write:

$$h_j = \frac{1}{2^{l-q}}[\sum_{x=(2m)2^{l-q-1}}^{(2m+1)2^{l-q-1}} f(x) \;\; - \sum_{x=(2m+1)2^{l-q-1}}^{(2m+2)2^{l-q-1}} f(x)]$$

$$\Longleftrightarrow h_j = \frac{1}{2^{l-q}} \sum_{x=(2m)2^{l-q-1}}^{(2m+1)2^{l-q-1}} [f(x) - f(x + 2^{l-q-1})]$$

$$\Longleftrightarrow h_j = \frac{1}{2^{l-q}} \sum_{x=(2m)2^{l-q-1}}^{(2m+1)2^{l-q-1}} [F(\frac{x}{2^l}) - F(\frac{x}{2^l} + 2^{-q-1})]$$

Recall that:

$$\forall y \in [0, 1), \;\; y + \eta \in [0, 1), \quad |F(y) \; - \; F(y + \eta)| \; \leq \; k|\eta|^h$$

then

$$\forall x \in [0..2^l - 1] \;\; \forall q \in [0..l - 1],$$
$$|F(\frac{x}{2^l}) - F(\frac{x}{2^l} + 2^{-q-1})| \leq k2^{(-q-1)h}$$

We finally obtain the well-known bound for the Haar coefficients of a Hölder function [5]:

$$\forall j, \quad |h_j| \leq \frac{k}{2} 2^{-h(q+1)}$$

This inequality is illustrated in Figure 6. The following remark is relevant for practical implementation: the optimal value of k (i.e. the lowest one) depends on the sampling precision. The curves of Figure 6 are drawn with $k = 2.5$ for a Weierstrass function sampled on 12 bits, and with $k = 3$ for an FBM[1] sampled on 10 bits.

B. Deception for Hölder functions

The use of a Haar decomposition for deception analysis has already been proposed in [18], but it seems that the complete computation of the adjusted coefficients (i.e. the coefficients of the function f') was not explicit. We thus use here a transformation between Walsh and Haar bases to explicitly compute the adjusted Haar coefficients. Details of the computations are given in Appendices A to D, and only the main steps are presented below.

We have (see Appendix A):

$$H_j(x) \;\; = \;\; \frac{1}{2^q}(\sum_{k=0}^{2^q-1} (-1)^{\sum_{t=0}^{l-1} m_t k_t} \Psi_{2^{l-q-1}+k2^{l-q}}(x))$$

with $\quad j = 2^q + m,$

$q \in [0..l - 1] \quad$ and $\quad m \in [0..2^q - 1]$

[1]FBM stands for fractional Brownian motion. For definition and properties of the fractional Brownian motion (FBM) see for instance [25]. As Weierstrass functions, paths of FBM (almost surely) verify a Hölder property, the irregularity being the same at each point. Thus an FBM with Hölder exponent h has box dimension equal to $2 - h$.

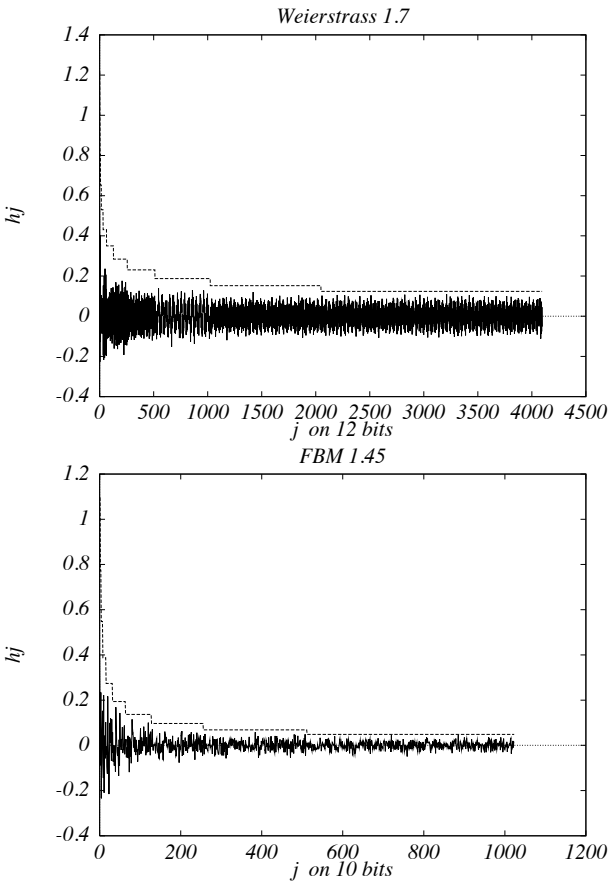

Fig. 6. Haar coefficients (continuous) and bound (dotted) for a Weierstrass function of dimension 1.7 sampled on 12 bits (upper) and an FBM of dimension 1.45 sampled on 10 bits (lower).

m_t and k_t represent the tth bit of the binary decomposition of m and k: $m = \sum_{t=0}^{l-1} m_t 2^t$ and $k = \sum_{t=0}^{l-1} k_t 2^t$.

Conversely (see Appendix B):

$$\begin{aligned} \Psi_j(x) &= \sum_{m=0}^{2^q-1} (-1)^{\sum_{t=0}^{q-1} k_t m_t} H_{2^q+m}(x) \\ \text{with} \quad & j = 2^{l-q-1}(1+2k), \\ & k \in [0..2^q-1] \quad \text{and} \quad q \in [0..l-1] \end{aligned}$$

We thus obtain the expression of Haar adjusted h'_j coefficients (see Appendices C and D):

$$f'(x) = \sum_{j=0}^{2^l-1} h'_j H_j(x)$$

$$\begin{aligned} h'_{2^q+m} &= h_{2^q+m}[1 - \frac{p_c}{l-1}(1 + \frac{1+(q-2)2^q}{2^q}) - 2p_m(1+\frac{q}{2})] \\ &- \frac{p_c}{2^q(l-1)} \sum_{u=0}^{q-1} (1-2^{u+1}) \sum_{r=0}^{2^{q-u-2}} h_{2^q+\sum_{t=0}^{u-1} m_t 2^t + (1-m_u)2^u + r2^{u+1}} \\ &- p_m \sum_{t=0}^{q-1} h_{2^q+m+(1-2m_t)2^t} \end{aligned}$$

We are now ready to compute an upper bound for the quantity $|f(x) - f'(x)|$:

$$\begin{aligned} |f(x) - f'(x)| &= |\sum_{j=1}^{2^l-1} (h_j - h'_j) H_j(x)| \\ &\leq \sum_{j=1}^{2^l-1} |h_j - h'_j||H_j(x)| \end{aligned}$$

Notice that for $x \in [0..2^l - 1]$:

$$H_j(x) \neq 0, \quad j = 2^q + m \quad \Longleftrightarrow$$
$$E(\frac{x}{2^{l-q-1}}) = 2m \quad or \quad E(\frac{x}{2^{l-q-1}}) = 2m+1$$

where $E()$ represents the integer part of x.

For a fixed x, and for each q, there exists a single value m_x of m such that $H_{2^q+m}(x) \neq 0$, and:

$$\forall\, x, \quad |f(x) - f'(x)| \leq \sum_{q=0}^{l-1} \sum_{m=0}^{2^l-1} |h_{2^q+m} - h'_{2^q+m}||H_{2^q+m}(x)|$$

$$|f(x) - f'(x)| \leq \sum_{q=0}^{l-1} |h_{2^q+m_x} - h'_{2^q+m_x}||H_{2^q+m_x}(x)|$$

with m_x such that $E(\frac{x}{2^{l-q-1}}) = 2m_x$ or $E(\frac{x}{2^{l-q-1}}) = 2m_x + 1$ and thus

$$|f(x) - f'(x)| \leq \sum_{q=0}^{l-1} |h_{2^q+m_x} - h'_{2^q+m_x}|$$

The bounds on the Haar coefficients of order q yield, after some computation:

$$\begin{aligned} &\forall m, \quad |h_{2^q+m} - h'_{2^q+m}| \leq \\ &\quad k2^{-h(q+1)}[\frac{p_c}{2^q(l-1)}[(1+(q-1)2^q] + p_m(1+q)] \end{aligned}$$

Thus:

$$\begin{aligned} &|f(x) - f'(x)| \\ &\leq k \sum_{q=0}^{l-1} (2^{-h(q+1)}[\frac{p_c}{2^q(l-1)}(1+(q-1)2^q) + p_m(1+q)]) \\ &\leq k \frac{p_c}{l-1} (\sum_{q=0}^{l-1} 2^{-h(q+1)-q}(1+(q-1)2^q)) \\ &\quad + k\, p_m (\sum_{q=0}^{l-1} 2^{-h(q+1)}(1+q)). \end{aligned}$$

We may now state Theorem 2 (see next page).

Since for all admissible values of l, p_m, p_c, B is a decreasing function of h, the relation 5 implies that the smaller h is (i.e. the more irregular the function is), the more different the functions f and f' may be, thus the more deceptive f is likely to be. This first fact bears some analogy with the results stated in [21] about sampling precision influence, and is confirmed by numerical simulations displayed in Figure 7.

Theorem 2: Let f be the sampling on l bits of a Hölder function of exponent h and constant k, defined on $[0,1]$, and let f' be defined as in (3). Then:

$$\forall x \in [0..2^l - 1], \qquad |f'(x) - f(x)| \leq k * B(p_m, p_c, l, h) \tag{5}$$

with

$$B(p_m,p_c,l,h) = \frac{p_c}{l-1}2^{-h}\left[\frac{2^{-l(h+1)}-1}{2^{-(h+1)}-1} + \frac{(1-2^{1-h})(2^{-hl}-1) - l2^{-hl}(1-2^{-h})}{(2^{-h}-1)^2}\right] + p_m\frac{2^{-h}}{(2^{-h}-1)^2}\left[1+2^{-hl}(l2^{-h}-l-1)\right]$$

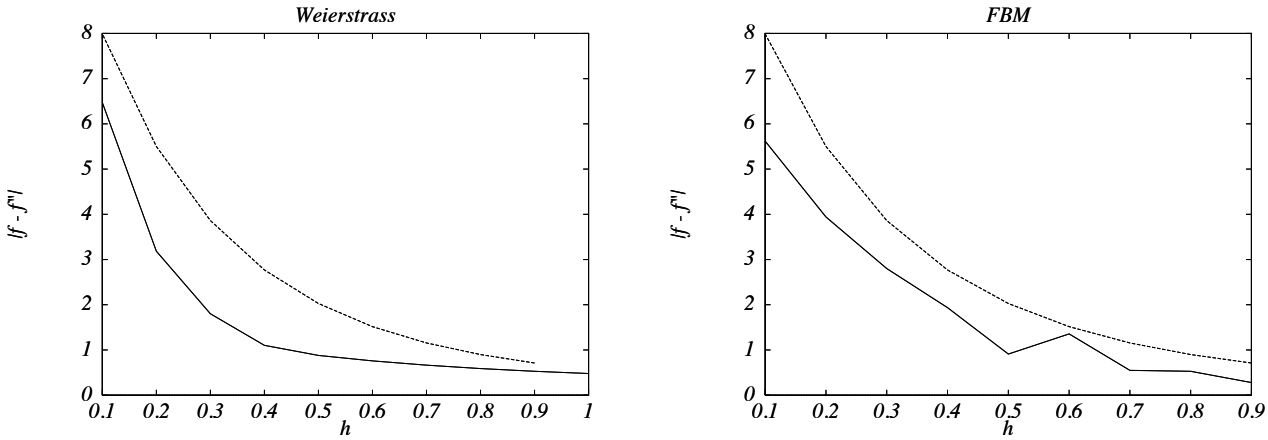

Fig. 7. $B(p_m,p_c,l,h)$ (dotted) and computed maximal differences between f and f' (continuous) as function of h, for Weierstrass functions (left), and FBM's (right), $l = 8$ bits, $p_c = 0.9$, $p_m = 0.25$.

V. Behaviour of B with respect to the GA parameters

A fine analysis of the function $B(p_m,p_c,l,h)$ is rather uneasy, because B defines a hypersurface of $\mathbf{R}^5$, but the following results may be stated (see Figure 8 to 12, which are 3D "cuts" of that surface).

- Dependence on l (Figures 8, 9, and 10):
  $B(p_m,p_c,l,h)$ has the following asymptotic behaviour when $l \to \infty$:

  $$\lim_{l\to\infty} B(p_m,p_c,l,h) = p_m\frac{2^{-h}}{(2^{-h}-1)^2}$$

  This limit does not depend on p_c (see Figure 12)[2]. We also have:

  $$B(p_m,p_c,2,h) = p_c 2^{-2h-1} + p_m(2^{-h} + 2^{1-2h})$$

  $B(p_m,p_c,l,h)$ increases with l for small values of l, and then decreases for larger values of l. It may be proved that the parameterized curves $B(p_m,p_c,\bullet,h)$ admit one and only one maximum at l_{max} in $[2,\infty)$. l_{max} increases when h decreases, i.e. when the function f becomes more and more irregular (see Figure 8 and 13).
  A sufficient condition for non-deception is $B(p_m,p_c,l,h) = 0$, which is in general not possible. A qualitative approach is then to keep B as small as possible. In that respect, a strategy to set the optimal value of l is the following one:
  - try to find a small value $l < l_{max}$ which is a tradeoff between a sufficiently fine sampling to correctly capture the optimum (according to [21]), while trying to limit the number of samples,
  - if no "small" values can be found, take a large value $l > l_{max}$, compatible with computational requirements.
- Dependence on p_c and p_m (Figures 14 and 15):
  Deception decreases as p_c and p_m decrease. This effect is more important for small values of h than for

[2]This fact is due to the definition of the mutation and crossover probabilities: *each gene* of the chromosome is mutated with probability p_m, while the crossover probability is defined on a whole chromosome. Thus when l tends to infinity, for fixed mutation and crossover probabilities, mutation becomes more and more important with respect to crossover. It may also be argued that the one-point crossover as it is defined here is meaningless when l is infinite.

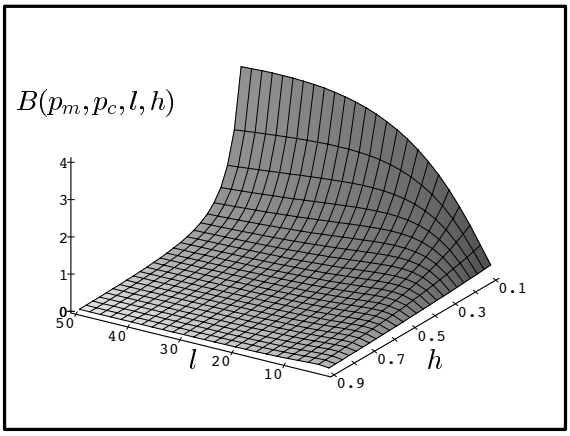

Fig. 8. $B(p_m, p_c, l, h)$ as a function of (l, h) for $p_m = 0.01$, $p_c = 0.7$.

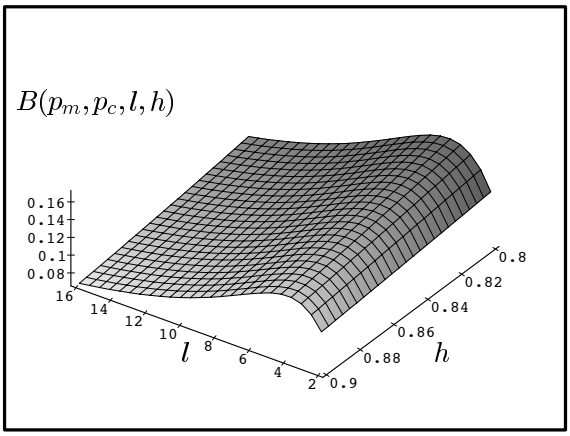

Fig. 9. Zoom on $B(p_m, p_c, l, h)$ for $p_m = 0.01$, $p_c = 0.7$ and large values of h.

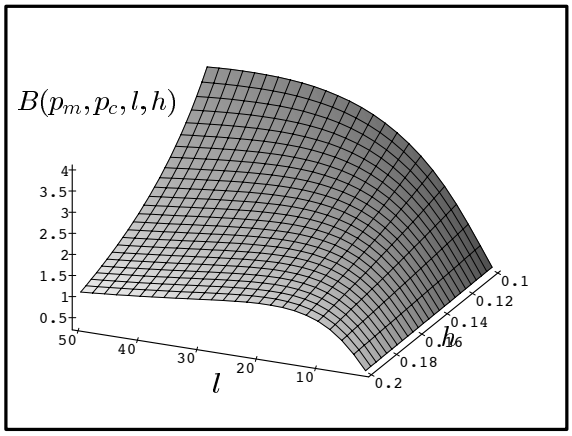

Fig. 10. Zoom on $B(p_m, p_c, l, h)$ for $p_m = 0.01$, $l = 8$ bits and small values of h.

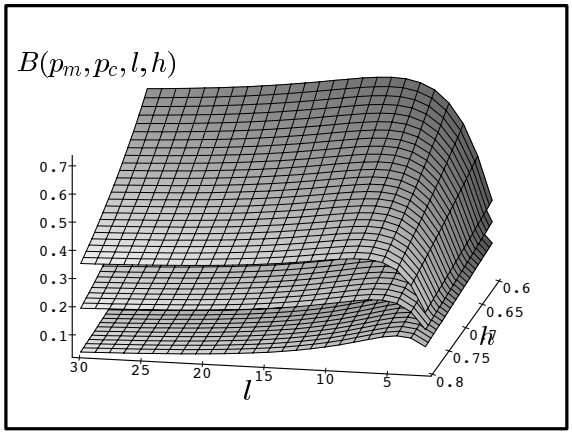

Fig. 11. Influence of p_m : $B(p_m, p_c, l, h)$ as a function of (l, h) for $p_c = 0.7$ fixed, and $p_m = 0.001, 0.05$ and 0.1.

large values of h. Note also that deception is less influenced by p_c than by p_m, and that the influence of p_m increases when h is small and when l is large. For example, if $h = 0.5$ and $l = 8$ bits, the influence of p_m on the deception is about 15 times more important than the influence of p_c. From a practical point of view, it means that decreasing p_m is much more efficient than decreasing p_c, in order to reduce deception. This fact also confirms the interest of the mutation probability decrease technique, especially for irregular functions. Mutation probability decrease has been theoretically justified for a simple model of GA, without crossover, with Markovian approaches (see [7]), and its practical efficiency has been experienced. Formula (5) shows that decreasing the mutation probability tightens the bound on $|f - f'|$, thus probably decreasing the deception of the function, i.e. making the convergence of the GA easier. Of course, this tendency is counterbalanced by the necessity of maintaining a reasonable value p_m in order to avoid premature convergences of the GA (effect that is not captured by the present model, that mainly takes into account the disruptive behaviour of the genetic operators).

VI. Conclusion and limitations of the analysis

The use of Haar decomposition instead of Walsh decomposition yields some interesting results for the particular case of optimization of Hölder functions with GA. These theoretical results quantify the intuitive guess that the more irregular the function looks, the more deceptive it is likely to be.

The explicit formula obtained in section IV-B provides a relation between :

- an intrinsic parameter of the function to be optimized : its Hölder exponent h,
- the parameters of the GA : l, p_m and p_c.

A simple analysis of this formula sheds new light on previous results obtained by other theoretical approaches of GA. Formula (5) provides a relation involving mutation and crossover probabilities, which may help to set these probabilities in order to make the convergence of the GA easier.

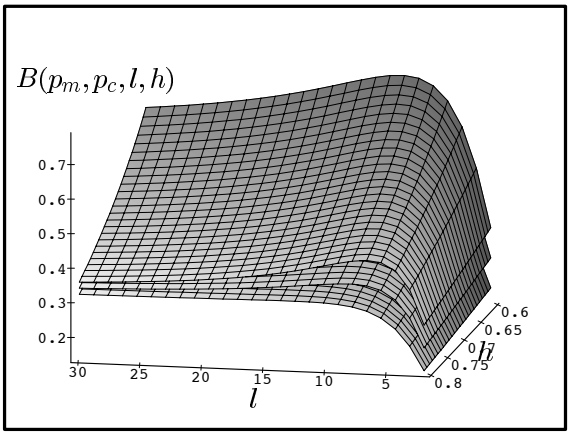

Fig. 12. Influence of p_c: $B(p_m, p_c, l, h)$ as a function of (l, h) for $p_m = 0.01$ fixed, and $p_c = 0.1, 0.5$ and 0.9.

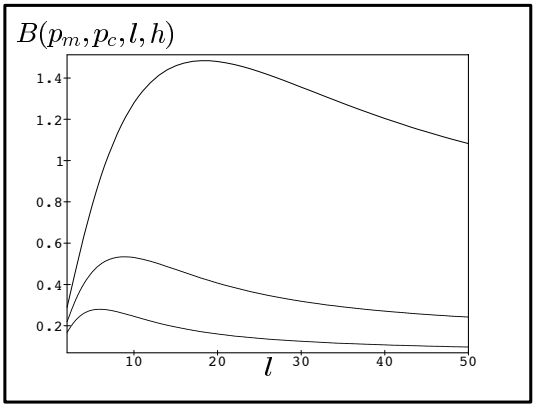

Fig. 13. $B(p_m, p_c, l, h)$ as a function of l, $p_m = 0.01$, $p_c = 0.7$, and for different values of h (up: 0.2, middle: 0.4, down: 0.6).

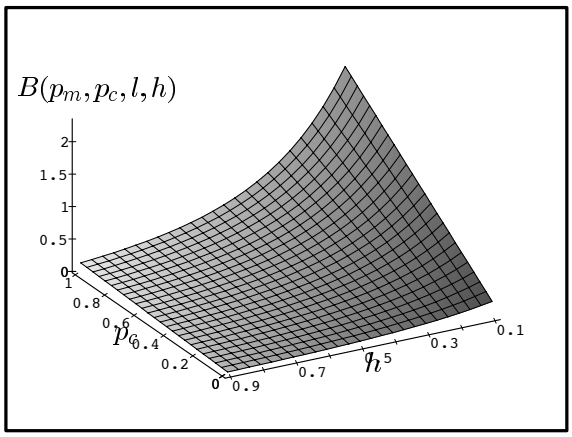

Fig. 14. $B(p_m, p_c, l, h)$ as a function of (p_c, h) for $l = 8$ bits, $p_m = 0.01$.

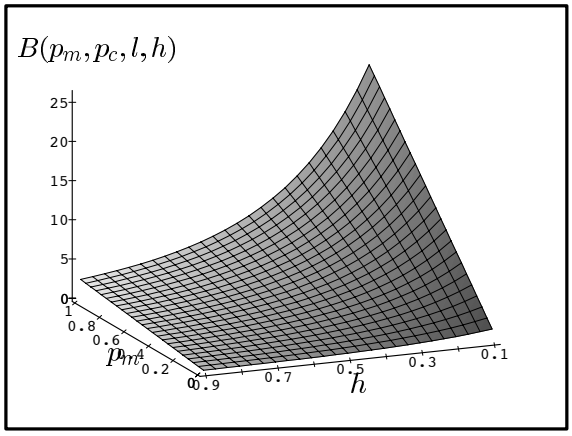

Fig. 15. $B(p_m, p_c, l, h)$ as a function of (p_m, h) for $l = 8$ bits, $p_c = 0.7$.

Notice however that this relation only gives a bound, which needs not to be optimal (non-deception can occur even if $B > 0$). Formula (5) points out that the function f' is located inside a strip of extent $2kB$ around f. Decreasing B decreases the maximal difference between f and f', thus the ability of f' to drive the GA onto a wrong optimum, which was defined as "deception". The extent of the strip around f may be tuned by changing the values of the parameters p_c, p_m, and l. This may suggest a sort of *a posteriori* validation test. Such a test is developed in [21].

Of course, the validity of this analysis depends on the validity of the deception work in general. Our purpose was not fundamentally to discuss the validity of deception analysis (see [27] for a rigorous analysis of the schema theorem), which has known weaknesses : it models only a simple GA, it takes into account the genetic operator only in a disruptive way, and does not consider populations of finite size. The results presented here thus only hold for the most simple of GAs. Nonetheless, our analysis relates the intuitive notion of "irregularity" (technically represented as the Hölder exponent of the function) to deception.

Furthermore, "Difficulty" for an optimization algorithm can come usually at least from two different sources[3] :

- size of the search space,
- irregularity (that can be related to deception) of the function.

It is possible to exhibit deceptive functions on a few number of bits, which are not "difficult" problems in the sense of a large search space (see for example [11], [12]). These problems are not strictly "difficult" to solve, but when the size of the search space grows, they become rapidly intractable. Experiments, as for example in [21] (where population sizes have been experimentally tuned), show the importance of the influence of the population size parameter on the performances and precision of results for large sized search spaces. Intuitively, we would like to relate this parameter to the size of the search space, but we were not able to include the population size parameter in the model, due to the theoretical limitations of schema theory.

In our analysis, we have related the deception to the irregularity of the function, i.e. we model the influence of p_m,

[3]There are a number of other origins for difficulty, see [14] for a more complete analysis.

p_c, and l on the GA. Separating in such a way the difficulty of the function from the size of the search space may appear as artificial and unrealistic (and the actual behaviour is for sure much more complicated), but corresponds to some intuitive guess. "Difficulty" in the sense of deceptive functions means isolated peaks surrounded by uninteresting areas, whereas other smoother regions are attractive.

Finally, we have supposed here that h and k are known for the function to be optimized : reliably estimating global Hölder exponents h on sampled function is not a simple problem, which we did not want to discuss here. However, several general methods have been proposed in other contexts for instance based on wavelets [23], [24].

Further work should be done in the following directions :

1. *Generalization to local Hölder characterization* : such an analysis would provide a variable-size strip around the function, yielding more precise results, at the expense, of course, of more complex computations.
2. *Use of other irregularity characterisations than a Hölder exponent* : a further work has been developed in [20]. It is based on "bitwise regularity coefficients" that are derived from grained Hölder exponents on a metric related to the Hamming distance. This irregularity analysis is no more based on the estimation of the Hölder exponent of an underlying one-dimensional function, and can provide more precise results especially for multidimensional problems. This work suggests also that the relative variations of the bound of Equation (5) may be used in order to evaluate a chromosome encoding. Numerical simulations with Gray code are reported in [20].
3. *Use of other analyses than the deception to quantify the efficiency of a GA* : As we have seen before, deception analysis is based on the schema theorem which models the action of genetic operators in a "negative" way, i.e. only the destruction of schemata by genetic operators is considered in the computation of the schema theorem bound. More recent approaches as [2], or based on Markov-based modelling as [27] or [4], seem to be of interest in the framework of Hölder functions, and will allow to consider the population size parameter in this framework.

Acknowledgements

The authors address special thanks to the anonymous reviewers and to the Editor-in-Chief whose numerous and fruitful comments led to important improvements of this paper.

References

[1] J. Albert, F. Ferri, J. Domingo, and M. Vincens. An Approach to Natural Scene Segmentation by Means of Genetic Algoritms with Fuzzy Data. In *Pattern Recognition and Image Analysis*, pages 97–113, 1992. Selected papers of the 4th Spanish Symposium (Sept 90), Perez de la Blanca Ed.

[2] L. Altenberg. The Schema Theorem and Price's Theorem. In *Foundation of Genetic Algorithms 3*, ed. Darrell Whitley and Michael Vose. pp. 23-49. Morgan Kaufmann, San Francisco, 1995.

[3] P. Andrey and P Tarroux. Unsupervised image segmentation using a distributed genetic algorithm. *Pattern Recognition, 27, 659-673*, 1993.

[4] R. Cerf. Asymptotic Convergence of Genetic Algorithms. *PhD thesis, Université Montpellier II*, 1994.

[5] I. Daubechies. Ten Lecture on Wavelets CBMS-NF Regional Conference series in Applied Mathematics, 1992.

[6] Y. Davidor. Genetic Algorithms and Robotics. A heuristic Strategy for Optimization. World Scientific, 1990. *World Scientific Series in Robotics and Automated Systems - vol 1.*

[7] T. E. Davis and J. C. Principe. A Simulated Annealing Like Convergence Theory for the Simple Genetic Algorithm. In *Proceedings of the Fourth International Conference on Genetic Algorithm*, pages 174–182, 1991. 13-16 July.

[8] K. Deb and D. E. Goldberg. Sufficient conditions for deceptive and easy binary functions. *Annals of Mathematics and Artificial Intelligence*, 10, pages 385-408, 1994.

[9] K. J. Falconer. Fractal Geometry : Mathematical Foundation and Applications. *John Wiley & Sons, 1990.*

[10] D. E. Goldberg. Genetic Algorithms in Search, Optimization, and Machine Learning. *Addison-Wesley, January 1989.*

[11] D. E. Goldberg. Genetic Algorithms and Walsh functions : Part I, a Gentle Introduction. *TCGA Report No. 88006*, University of Alabama, Tuscaloosa, US, 1988.

[12] D. E. Goldberg. Genetic Algorithms and Walsh functions : Part II, Deception and its Analysis. *TCGA Report No. 89001*, University of Alabama, Tuscaloosa, US, 1989.

[13] D E. Goldberg. Construction of High–order Deceptive Functions Using Low–order Walsh Coefficients. *IlliGAL Report 90002*, University of Illinois at Urbana–Champaign, Urbana, IL 61801, December 1990.

[14] D E. Goldberg, K. Deb and J. Horn. Massive multimodality, deception, and genetic algorithms *PPSN 2, Parallel Problem Solving from Nature*, pages 37–46, 1992

[15] A. Hill and C. J. Taylor. Model-Based Image Interpretation using Genetic Algorithms. *Image and Vision Computing*, 10(5):295–301, June 1992.

[16] J. H. Holland. Adaptation in Natural and Artificial System. *Ann Arbor, University of Michigan Press*, 1975.

[17] A. Homaifar and X. Qi. Analysis of Genetic Algorithms Deception by Hadamard Transform. In *International symposium machine learning and neuronal networks*, pages 75–78, October 1990. org. by IASTED.

[18] S. Khuri. Walsh and Haar functions in Genetic Algorithms. *Proceedings of the 1994 ACM Symposium on Applied Computing.* 1994, ACM Press.

[19] J. R. Koza. Genetic Programming. *MIT Press*, 1992.

[20] B. Leblanc and E. Lutton. Bitwise regularity and GA-hardness. ICEC 98, May 5-9, Anchorage, Alaska, 1998.

[21] E. Lutton and J. Lévy Véhel. Some Remarks on the Optimization of Hölder Functions with Genetic Algorithms, *Inria Research Report No 2627*, 1996.

[22] E. Lutton and P. Martinez. A Genetic Algorithm for the Detection of 2d Geometric Primitives in Images, In *12-ICPR*, 1994. Jerusalem, Israel, 9-13 October.

[23] S. Mallat and S. Zhong, Characterization of Signals from MultiScale Edges, *EEE Transactions on Pattern Analysis and Machine Intelligence, 14(7), pp. 710-732, 1992*

[24] S. Mallat and W.L. Hang, Singularity Detection and Processing with Wavelets, *IEEE Trans. Inform. Theory, Vol. 38, pp. 617-643, March 1992.*

[25] B.B. Mandelbrot and J.W. Van Ness. Fractional Brownian Motion, Fractional Gaussian Noises and Applications. *SIAM Review 10*, 4:422–437, 1968.

[26] D. J. Nettleton and R. Garigliano. Evolutionary Algorithms and a Fractal Inverse Problem. *Biosystems*, (33):221–231, 1994. Technical note.

[27] S. Rochet. Convergence des Algorithmes Génétiques : Modèles Stochastiques et Épistasie Phd Thesis Université de Provence, Mathematics and Computer Science center, January 1998.

[28] G. Roth and M. D. Levine. Geometric Primitive Extraction Using a Genetic Algorithm. In *IEEE Computer Society Conference on CV and PR*, pages 640–644, 1992.

[29] P. Trompette, J. L. Marcelin, and C. Schmelding. OptImal Damping of Viscoelastic Constrained Beams or Plates by Use of a Genetic Algotithm. In *IUTAM*, 1993. Zakopane, Poland.

[30] S. Truvé. Using a Genetic Algorithm to Solve Constraint Satisfation Problems Generated by an Image Interpreter. In *Theory and Applications of Image Analysis : 7th Scandinavian Conference on Image Analysis*, pages 378–386, August 1991. Aalborg, Denmark.

Appendix

I. Expression of the Haar functions in the Walsh basis

For every real function f, defined on $[0..2^l-1]$, let:

$$\forall x \in [0..2^l-1] \qquad f(x) = \sum_{j=0}^{2^l-1} \omega_j \Psi_j(x)$$

with $\Psi_j(x) = \prod_{t=0}^{l-1}(-1)^{x_t j_t}$, x_t and j_t being the values of the tth bit of the binary decomposition of x and j. We will sometimes write $\Psi_j^l(x)$ to emphasize the dependance on l.

The Walsh coefficients are given by:

$$\omega_j = \frac{1}{2^l}\sum_{x=0}^{2^l-1} f(x)\Psi_j(x)$$

Let m_t and k_t be the tth bit of the binary decomposition of m and k: $m = \sum_{t=0}^{l-1} m_t 2^t$. and $k = \sum_{t=0}^{l-1} k_t 2^t$

Proposition 3: Every function H_j can be decomposed in the Walsh basis as follows:

$$\boxed{H_j(x) = \frac{1}{2^q}\left(\sum_{k=0}^{2^q-1}(-1)^{\sum_{t=0}^{l-1} m_t k_t}\Psi_{2^{l-q-1}+k2^{l-q}}(x)\right)} \quad (6)$$

with $j = 2^q + m, \quad q \in [0..l-1]$ and $m \in [0..2^q-1]$.

Proof:

Let T_j be the righthand side term of (6):

$$T_j(x) = \frac{1}{2^q}\left(\sum_{k=0}^{2^q-1}(-1)^{\sum_{t=0}^{l-1} m_t k_t}\Psi_{2^{l-q-1}+k2^{l-q}}(x)\right)$$

with $j = 2^q + m$ and $m \in [0..2^q-1]$.

We have to prove:

$$\begin{aligned} T_j(x) &= H_j(x) \\ &= \begin{cases} 1 & \text{if } (2m)2^{l-q-1} \le x < (2m+1)2^{l-q-1} \\ -1 & \text{if } (2m+1)2^{l-q-1} \le x < (2m+2)2^{l-q-1} \\ 0 & \text{else} \end{cases} \end{aligned}$$

Define: $j' = 2^{l-q-1} + k2^{l-q}$. We have:

$$\Psi_{2^{l-q-1}+k2^{l-q}}(x) = \Psi_{j'}(x) = (-1)^{\sum_{t=0}^{l-1} x_t j'_t}$$

with $j' = \sum_{t=0}^{l-1} j'_t 2^t = 2^{l-q-1} + k2^{l-q} = 2^{l-q-1} + \sum_{t=0}^{q-1} k_t 2^{t+(l-q)}$

and:

$$\begin{cases} j'_t = 0 & \text{if } t \in [0..l-q-2] \\ j'_{l-q-1} = 1 & \\ j'_t = k_{t-(l-q)} & \text{if } t \in [l-q..l-1] \end{cases}$$

Thus:

$$\Psi_{j'}(x) = (-1)^{x_{l-q-1} + \sum_{t=l-q}^{l-1} k_{t-(l-q)} x_t}$$

Replacing in the formula giving T_j:

$$\begin{aligned} & T_j(x) \\ = & \frac{1}{2^q}\sum_{k=0}^{2^q-1}(-1)^{\sum_{t=l-q}^{l-1} m_t k_t + x_{l-q-1} + \sum_{t=0}^{l-1} k_{t-(l-q)} x_t} \\ = & \frac{(-1)^{x_{l-q-1}}}{2^q}\sum_{k=0}^{2^q-1}(-1)^{\sum_{t=0}^{q-1} k_t (m_t + x_{t+(l-q)})} \end{aligned}$$

We consider two cases:

1 — $\underline{x \in [m2^{l-q}..(m+1)2^{l-q})}$

We may write: $x = m2^{l-q} + \alpha$, $\alpha \in [0..2^{l-q})$ with $\alpha = \sum_{t=0}^{l-q-1} \alpha_t 2^t$.

Thus:

$$\begin{aligned} x &= \sum_{t=0}^{q-1} m_t 2^{l-q+t} + \sum_{t=0}^{l-q-1} \alpha_t 2^t \\ &= \sum_{t=l-q}^{l-1} m_{t-(l-q)} 2^t + \sum_{t=0}^{l-q-1} \alpha_t 2^t \\ & \forall\, t \in [l-q..l-1] \ , \ x_t = m_{t-(l-q)} \end{aligned}$$

$$T_j(x) = \frac{(-1)^{x_{l-q-1}}}{2^q}\sum_{k=0}^{2^q-1}(-1)^{\sum_{t=0}^{q-1} k_t (2m_t)}$$

Since $\sum_{t=0}^{q-1} k_t(2m_t)$ is even, the expression $(-1)^{\sum_{t=0}^{q-1} k_t(2m_t)}$ equals 1, and:

$$T_j(x) = (-1)^{x_{l-q-1}}$$

Thus:

if $\quad x \in [(2m)2^{l-q-1}..(2m+1)2^{l-q-1})$

$\quad T_j(x) = 1$ because $x_{l-q-1} = 0$

if $\quad x \in [(2m+1)2^{l-q-1}..(2m+2)2^{l-q-1})$

$\quad T_j(x) = -1$ because $x_{l-q-1} = 1$

2 — $\underline{x \notin [m2^{l-q}..(m+1)2^{l-q})}$

This is equivalent to:

$$\exists t \in [l-q..l-1] \quad \text{such that} \quad x_t \neq m_{t-(l-q)}$$

$$\Longleftrightarrow \exists t \in [0,..-1] \quad \text{such that} \quad x_{t+(l-q)} \neq m_t$$

using

$$\forall t, \quad m_t \in \{0,1\} \ \text{ and } \ x_t \in \{0,1\}$$

we get:

$$m_t \neq x_{t+(l-q)} \Longrightarrow m_t + x_{t+(l-q)} = 1$$

Let us define T_1 as:

$$T_1 = \{t \in [0..q-1] \ \ / m_t + x_{t+(l-q)} = 1\}$$

We know that $T_1 \neq \emptyset$, thus:

$$T_j(x) = \frac{(-1)^{x_{l-q-1}}}{2^q}[\sum_{k=0}^{2^q-1}(-1)^{\sum_{t\in T_1} k_t}]$$

Let $b \in [0..2^l-1]$ be such that:

$$\begin{cases} b_t = 1 & t \in T_1 \\ b_t = 0 & t \notin T_1 \end{cases}$$

Then:

$$T_j(x) = \frac{(-1)^{x_{l-q-1}}}{2^q}[\sum_{k=0}^{2^q-1}(-1)^{\sum_{t=0}^{q-1} b_t k_t}]$$

The term $(-1)^{\sum_{t=0}^{q-1} b_t k_t}$ is equal to $\Psi_b^q(k)$. We can thus write:

$$T_j(x) = \frac{(-1)^{x_{l-q-1}}}{2^q}[\sum_{k=0}^{2^q-1}\Psi_b^q(k)]$$

The term $\sum_{k=0}^{2^q-1}\Psi_b^q(k)$ is zero if b is not 0, which is the case here, because there is at least one bit of b that equals 1.
We finally obtain:

$$x \notin [m2^{l-q}..(m+2)2^{l-q}] \ \ \Rightarrow \ \ T_j(x) = 0.$$

□

II. Expression of the Walsh functions in the Haar basis

For $i = 2^q + m$, $q \in [0..l-1]$ and $m \in [0..2^q-1]$, we have:

$$H_i(x) = \frac{1}{2^q}(\sum_{k=0}^{2^q-1}(-1)^{\sum_{t=0}^{l-1} m_t k_t}\Psi_{2^{l-q-1}+k2^{l-q}}(x))$$

.

The coefficients of transformation matrix between Walsh and Haar bases are thus:

$$\begin{cases} m_{ij} = 0 & \text{if} \quad j \neq 2^{l-q-1} + k2^{l-q} \\ m_{ij} = \frac{1}{2^q}(-1)^{\sum_{t=0}^{q-1} m_t k_t} & \text{else} \end{cases}$$

$$H_i(x) = \sum_{j=0}^{2^{l-1}} m_{ij}\Psi_j(x)$$

As the two bases are othogonal bases, which are non-orthonormal, the inverse formula is:

$$\Psi_i(x) = \sum_{j=0}^{2^{l-1}} m_{ji}\frac{||\Psi_i||^2}{||H_j||^2}H_j(x)$$

with:

$$||\Psi_i||^2 = \sum_{x=0}^{x=2^l-1}[\Psi_i(x)]^2 = 2^l$$

$$||H_i||^2 = \sum_{x=0}^{x=2^l-1}[H_i(x)]^2 = \sum_{x=m2^{l-q}}^{x=(m+1)2^{l-q}} 1 = 2^{l-q},$$

$$\text{because} \ \ i = 2^q + m$$

Thus

$$\Psi_i(x) = \sum_{j=0}^{2^{l-1}} 2^q m_{ji} H_j(x)$$

For every integer $i \in [1..2^l-1]$, whose binary decomposition is $i = \sum_{t=0}^{l-1} i_t 2^t$, there exists a unique couple $(q,k), k \in [0..2^q-1]$ such that:

$$i = 2^{l-q-1} + k2^{l-q}$$

In the expression of Ψ_i, the only m_{ji} coefficients which are non zero correspond to the $j = 2^q + m$ such that:

$$\exists k \in [0..2^q-1] \ \text{ such that } \ i = 2^{l-q-1} + k2^{l-q}$$

For $q > 0$:

$$\begin{aligned} i = \sum_{t=0}^{l-1} i_t 2^t &= 2^{l-q-1} + \sum_{t=0}^{q-1} k_t 2^{t+(l-q)} \\ &= 2^{l-q-1} + \sum_{t=l-q}^{l-1} k_{t-(l-q)} 2^t \end{aligned}$$

- if $t \in [0..l-q-1)$ $\ i_t = 0$
- if $t = l-q-1$ $\ i_t = 1$
- if $t \in [l-q..l-1]$ $\ i_t = k_{t-(l-q)}$

$l-q-1$ is the first non-zero bit of i.
Thus:

$$\Psi_i(x) = \sum_{m=0}^{2^q-1}(-1)^{\sum_{t=0}^{q-1} i_{t+(l-q)} m_t} H_{2^q+m}(x)$$

Remark: this relation also holds for $q = 0$ (in this case $m = 0$), with the convention $\sum_{t=0}^{t=-1} \star = 0$.
Finally:

$$\Psi_i(x) = \sum_{m=0}^{2^q-1}(-1)^{\sum_{t=0}^{q-1} k_t m_t} H_{2^q+m}(x) \qquad (7)$$

$$\text{with} \ \ i = 2^{l-q-1}(1+2k), \quad k \in [0..2^q-1], \quad \text{and} \ \ q \in [0..l-1]$$

III. Expression of the Haar coefficients as a function of the Walsh coefficients and conversely

For any function f defined on $[0..2^l-1]$, write:

$$f(x)=\sum_{i=0}^{2^l-1}\omega_i\Psi_i(x)=\sum_{j=0}^{2^l-1}h_jH_j(x)$$

with $\quad h_j=\frac{1}{2^{l-q}}\sum_{x=0}^{2^l-1}f(x)H_j(x)$
and $\quad \omega_i=\frac{1}{2^l}\sum_{x=0}^{2^l-1}f(x)\Psi_i(x)$.
Thus:

$$h_j=\frac{1}{2^{l-q}}\sum_{x=0}^{2^l-1}(\sum_{k=0}^{2^l-1}\omega_k\Psi_k(x))H_j(x)$$

Using the expression of H_j in the Walsh basis:

$$h_j=$$
$$\frac{1}{2^{l-q}}\sum_{x=0}^{2^l-1}(\sum_{k=0}^{2^l-1}\omega_k\Psi_k(x))(\frac{1}{2^q}\sum_{v=0}^{2^q-1}(-1)^{\sum_{t=0}^{l-1}m_tv_t}\Psi_{2^{l-q-1}+v2^{l-q}}(x))$$
$$=\frac{1}{2^{l-q}}\frac{1}{2^q}\sum_{v=0}^{2^q-1}(-1)^{\sum_{t=0}^{l-1}m_tv_t}(\sum_{k=0}^{2^l-1}\omega_k\sum_{x=0}^{2^l-1}\Psi_k(x)\Psi_{2^{l-q-1}+v2^{l-q}}(x))$$

The Ψ_j form an othogonal basis:

$$\sum_{x=0}^{2^l-1}\Psi_i(x)\Psi_j(x)=\begin{cases}2^l & \text{if } i=j\\ 0 & \text{else}\end{cases}$$

We obtain:

$$\boxed{h_j=\sum_{v=0}^{2^q-1}(-1)^{\sum_{t=0}^{l-1}m_tv_t}\omega_{2^{l-q-1}+v2^{l-q}}\quad \text{with}\quad j=2^q+m}$$

We now move to the Walsh coefficients:

$$\omega_i=\frac{1}{2^l}\sum_{x=0}^{2^l-1}f(x)\Psi_i(x)\quad\text{with}\quad i=2^{l-q-1}(1+2k)$$

$$\omega_i=\frac{1}{2^l}\sum_{x=0}^{2^l-1}(\sum_{v=0}^{2^l-1}h_vH_v(x))\Psi_i(x)$$

$$\omega_i=\frac{1}{2^l}\sum_{x=0}^{2^l-1}(\sum_{v=0}^{2^l-1}h_vH_v(x))(\sum_{m=0}^{2^q-1}(-1)^{\sum_{t=0}^{q-1}m_tk_t}H_{2^q+m}(x))$$

$$\omega_i=\frac{1}{2^l}(\sum_{m=0}^{2^q-1}(-1)^{\sum_{t=0}^{q-1}m_tk_t}\sum_{v=0}^{2^l-1}h_v\sum_{x=0}^{2^l-1}H_v(x)H_{2^q+m}(x))$$

$$\boxed{\omega_i=\frac{1}{2^q}\sum_{m=0}^{2^q-1}h_{2^q+m}(-1)^{\sum_{t=0}^{q-1}m_tk_t}\quad\text{with}\quad i=2^{l-q-1}(1+2k)}$$

IV. Computation of the Haar adjusted coefficients

Let:

$$f(x)=\sum_{i=0}^{2^l-1}\omega_i\Psi_i(x)=\sum_{j=0}^{2^l-1}h_jH_j(x)$$

and:

$$f'(x)=\sum_{i=0}^{2^l-1}\omega'_i\Psi_i(x)=\sum_{j=0}^{2^l-1}h'_jH_j(x)$$

We write:
$i=2^{l-q-1}(1+2k)$, $q\in[0..l-1]$, $k\in[0..2^q-1]$
$j=2^q+m$, $q\in[0..l-1]$, $m\in[0..2^q-1]$.
Then:

$$\begin{aligned}f'(x)&=\omega'_0+\sum_{q=0}^{l-1}\sum_{k=0}^{2^q-1}\omega'_i\sum_{m=0}^{2^q-1}(-1)^{\sum_{t=0}^{q-1}m_tk_t}H_{2^q+m}(x)\\&=h'_0+\sum_{q=0}^{l-1}\sum_{m=0}^{2^q-1}h'_{2^q+m}H_{2^q+m}(x)\end{aligned}$$

In the following, the subscript t indicates the tth bit of the binary decomposition, i.e. $k=\sum_{t=0}^{t=l-1}k_t2^t$, $m=\sum_{t=0}^{t=l-1}m_t2^t$, etc ...

$$f'(x)=h'_0+\sum_{q=0}^{l-1}\sum_{m=0}^{2^q-1}h'_{2^q+m}H_{2^q+m}(x)$$
$$=\omega'_0+\sum_{q=0}^{l-1}\sum_{m=0}^{2^q-1}[\sum_{k=0}^{2^q-1}\omega'_{2^{l-q-1}(1+2k)}(-1)^{\sum_{t=0}^{q-1}m_tk_t}]$$
$$H_{2^q+m}(x)$$

$$h'_{2^q+m}=\sum_{k=0}^{2^q-1}\omega'_{2^{l-q-1}(1+2k)}(-1)^{\sum_{t=0}^{q-1}m_tk_t}$$

(we have: $\omega'_0=\omega_0=h_0=h'_0$)
Using the expression of the ω'_i (equation (3)):

for $q=0$:

$$\omega'_{2^{l-1}}=\omega_{2^{l-1}}(1-2p_m)=h'_1$$

thus:

$$h'_1=h_1(1-2p_m)$$

for $q > 0$:

$$h'_{2^q+m} = \sum_{k=0}^{2^q-1} \omega_{2^{l-q-1}(1+2k)}(-1)^{\sum_{t=0}^{q-1} m_t k_t} [1 - p_c \frac{\delta(2^{l-q-1}(1+2k))}{l-1} - 2p_m \mathcal{O}(2^{l-q-1}(1+2k))]$$

Then:

$$h'_{2^q+m} = \sum_{k=0}^{2^q-1} \sum_{m'=0}^{2^q-1} h_{2^q+m'} \frac{(-1)^{\sum_{t=0}^{q-1}(m_t+m'_t)k_t}}{2^q} [1 - p_c \frac{\delta(2^{l-q-1}(1+2k))}{l-1} - 2p_m \mathcal{O}(2^{l-q-1}(1+2k))]$$

$$h'_{2^q+m} = \frac{1}{2^q} \sum_{k=0}^{2^q-1} [1 - p_c \frac{\delta(2^{l-q-1}(1+2k))}{l-1} - 2p_m \mathcal{O}(2^{l-q-1}(1+2k))] \sum_{m'=0}^{2^q-1} h_{2^q+m'} (-1)^{\sum_{t=0}^{q-1}(m_t+m'_t)k_t}$$

- It is obvious that: $\delta(2^{l-q-1}(1+2k)) = \delta^*(k)+1$ $\delta^*(k)$ being the position of the last non-zero bit of k. For $k = 0$, we have $\delta(2^{l-q-1}) = 0$. We thus define: $\delta^*(0) = -1$
- We also have: $\mathcal{O}(2^{l-q-1}(1+2k)) = 1 + \mathcal{O}(k)$

h'_{2^q+m} thus becomes:

$$h'_{2^q+m} = \frac{1}{2^q} \sum_{m'=0}^{2^q-1} h_{2^q+m'} \sum_{k=0}^{2^q-1} [1 - p_c \frac{\delta^*(k)+1}{l-1} - 2p_m(1+\mathcal{O}(k))] (-1)^{\sum_{t=0}^{q-1}(m_t+m'_t)k_t}$$

and finally:

$$h'_{2^q+m} = h_{2^q+m}(1 - \frac{p_c}{l-1} - 2p_m) - \frac{p_c}{2^q(l-1)} \sum_{m'=0}^{2^q-1} h_{2^q+m'} \sum_{k=0}^{2^q-1} \delta^*(k)(-1)^{\sum_{t=0}^{q-1}(m_t+m'_t)k_t} - \frac{2p_m}{2^q} \sum_{m'=0}^{2^q-1} h_{2^q+m'} \sum_{k=0}^{2^q-1} \mathcal{O}(k)(-1)^{\sum_{t=0}^{q-1}(m_t+m'_t)k_t}.$$

Let us define $\Delta(m,m')$ and $\mathcal{O}(m,m')$ as:

$$\Delta(m,m') = \sum_{k=0}^{2^q-1} \delta^*(k)(-1)^{\sum_{t=0}^{q-1}(m'_t+m_t)k_t} = -1 + \sum_{k=1}^{2^q-1} \delta^*(k)(-1)^{\sum_{t=0}^{\delta^*(k)}(m'_t+m_t)k_t}$$

and

$$\mathcal{O}(m,m') = \sum_{k=0}^{2^q-1} \mathcal{O}(k)(-1)^{\sum_{t=0}^{q-1}(m'_t+m_t)k_t} = \sum_{k=0}^{2^q-1} [\sum_{t=0}^{q-1} k_t](-1)^{\sum_{t=0}^{q-1}(m'_t+m_t)k_t}$$

Write $k = 2^d + b$, or :

$$k = 2^d + \sum_{t=0}^{d-1} b_t 2^t,$$

thus $\delta^*(k) = d$

◇ **Computation of $\Delta(m,m')$:**

$$\Delta(m,m') = -1 + \sum_{d=0}^{q-1} \sum_{b=0}^{2^d-1} d(-1)^{\sum_{t=0}^{q-1}(m'_t+m_t)k_t} = -1 + \sum_{d=0}^{q-1} d(-1)^{m'_d+m_d} \sum_{b=0}^{2^d-1} (-1)^{\sum_{t=0}^{q-1}(m'_t+m_t)b_t}$$

$\sum_{b=0}^{2^d-1}(-1)^{\sum_{t=0}^{q-1}(m'_t+m_t)b_t}$ corresponds to:

$$\sum_{b=0}^{2^d-1} \Psi_m^d(b)\Psi_{m'}^d(b) = \begin{cases} 0 & \text{if the first } d \text{ bits of } m \text{ and } m' \text{ are identical} \\ 2^d & \text{else} \end{cases}$$

where Ψ_m^d is the restriction of the mth Walsh function on d bits, i.e. to the set $[0..2^d-1]$. Thus:

1. if $\forall\, t \in [0..q-1]$ $m_t \neq m'_t$ then $\Delta(m,m') = -1$
2. if $m = m'$ then $\Delta(m,m') = -1 + \sum_{d=0}^{q-1} d2^d = 1 + q2^q - 2^{q+1}$
3. Let us define u such that $\forall t \in [0..u-1]$, $m_t = m'_t$, and $m_u \neq m'_u$ (i.e. $m_u + m'_u = 1$). Then:

$$\Delta(m,m') = -1 + \sum_{d=0}^{q-1} d(-1)^{m_d+m'_d} \sum_{b=0}^{2^d-1} [\Psi_m^d(b)\Psi_{m'}^d(b)]$$

$$\Delta(m,m') = -1 + \sum_{d=0}^{u-1} d(-1)^{m'_d+m_d} [\sum_{b=0}^{2^d-1} \Psi_m^d(b)\Psi_{m'}^d(b)] + u(-1)^{m'_u+m_u} [\sum_{b=0}^{2^u-1} \Psi_m^u(b)\Psi_{m'}^u(b)] + \sum_{d=u+1}^{q-1} d(-1)^{m'_d+m_d} [\sum_{b=0}^{2^d-1} \Psi_m^d(b)\Psi_{m'}^d(b)]$$

$$\Delta(m,m') = -1 + \sum_{d=0}^{u-1} d2^d - u2^u = 1 - 2^{u+1}$$

Finally: denoting u the integer such that $\forall t \in [0..u-1]$, $m_t = m'_t$ and $m_u \neq m'_u$, the three cases above are summarized as:

if $u \in [0..q-1]$	$\longrightarrow$	$\Delta(m,m') = 1 - 2^{u+1}$
if $u = q$ (i.e. $m = m'$)	$\longrightarrow$	$\Delta(m,m') = 1 + q2^q - 2^{q+1}$

◇ **Computation of $\mathcal{O}(m,m')$:**

1. If $m = m'$: $\quad (-1)^{\sum_{t=0}^{q-1}(m'_t+m_t)k_t} = 1$, and:

$$\mathcal{O}(m,m) = \sum_{k=0}^{2^q-1} \mathcal{O}(k)$$

Set $s = \mathcal{O}(k)$, with $k \in [0..2^q-1]$. We obtain:

$$\mathcal{O}(m,m) = \sum_{s=0}^{q} C_q^s s = q2^{q-1}$$

2. If $\forall t \in [0..q-1] \quad m_t \neq m'_t$:

$$\mathcal{O}(m,m') = \sum_{k=0}^{2^q-1} \mathcal{O}(k)(-1)^{\mathcal{O}(k)} = \sum_{s=0}^{q} C_q^s (-1)^s = 0$$

3. In the general case, we have:

$$\mathcal{O}(m,m') = \sum_{t=0}^{q-1}(-1)^{m_t+m'_t} \prod_{v=0,v\neq t}^{q-1} (1+(-1)^{m_v+m'_v}) \quad (8)$$

Proof:
For $q = 1$, this equality is obvious, and we prove the formula by induction: suppose it is true for q, then:

$$\forall m,m' \in [0..2^{q+1}]$$

$$\mathcal{O}^{q+1}(m,m') = \sum_{k=0}^{2^{q+1}-1} \sum_{t=0}^{q} k_t(-1)^{\sum_{t=0}^{q}(m_t+m'_t)k_t}$$

$$\begin{aligned}\mathcal{O}^{q+1}(m,m') &= \sum_{k=0}^{2^q-1} \sum_{t=0}^{q} k_t(-1)^{\sum_{t=0}^{q}(m_t+m'_t)k_t} \\ &+ \sum_{k=2^q}^{2^{q+1}-1} \sum_{t=0}^{q} k_t(-1)^{\sum_{t=0}^{q}(m_t+m'_t)k_t}\end{aligned}$$

In the term $\sum_{k=2^q}^{2^{q+1}-1}\sum_{t=0}^{q} k_t(-1)^{\sum_{t=0}^{q}(m_t+m'_t)k_t}$, let us write $k = 2^q + f$ with $f \in [0..2^q-1]$, i.e. $k_t = f_t \ \ \forall t \in [0..q-1]$.

$$\mathcal{O}^{q+1}(m,m') = \mathcal{O}^q(m,m') + (-1)^{m_q+m'_q} \sum_{f=0}^{2^q-1}(1+\sum_{t=0}^{q-1} f_t)(-1)^{\sum_{t=0}^{q-1}(m_t+m'_t)f_t}$$

$$\begin{aligned}\mathcal{O}^{q+1}(m,m') &= \mathcal{O}^q(m,m') + (-1)^{m_q+m'_q}\mathcal{O}^q(m,m') \\ &+ (-1)^{m_q+m'_q} \sum_{f=0}^{2^q-1}(-1)^{\sum_{t=0}^{q-1}(m_t+m'_t)f_t}\end{aligned}$$

We have to prove that:

$$\sum_{f=0}^{2^q-1}(-1)^{\sum_{t=0}^{q-1}(m_t+m'_t)f_t} = \prod_{v=0}^{q-1}(1+(-1)^{m_v+m'_v})$$

This is obviously true for $q = 1$ and $q = 2$.
Define $S_q = \sum_{f=0}^{2^q-1}(-1)^{\sum_{t=0}^{q-1}(m_t+m'_t)f_t}$. Then

$$\begin{aligned}S_{q+1} &= \sum_{f=0}^{2^{q+1}-1}(-1)^{\sum_{t=0}^{q}(m_t+m'_t)f_t} \\ &= \sum_{f=0}^{2^q-1}(-1)^{\sum_{t=0}^{q}(m_t+m'_t)f_t} \\ &\quad + \sum_{f=2^q}^{2^{q+1}-1}(-1)^{\sum_{t=0}^{q}(m_t+m'_t)f_t} \\ &= S_q + (-1)^{m_q+m'_q}S_q = (1+(-1)^{m_q+m'_q})S_q\end{aligned}$$

We thus obtain for $\mathcal{O}^{q+1}(m,m')$:

$$\begin{aligned}\mathcal{O}^{q+1}(m,m') &= \mathcal{O}^q(m,m')(1+(-1)^{m_q+m'_q}) \\ &\quad +(-1)^{m_q+m'_q}\prod_{v=0}^{q-1}(1+(-1)^{m_v+m'_v}) \\ &= \sum_{t=0}^{q-1}(-1)^{m_t+m'_t}\prod_{v=0,v\neq t}^{q}(1+(-1)^{m_v+m'_v}) \\ &\quad +(-1)^{m_q+m'_q}\prod_{v=0,v\neq t}^{q-1}(1+(-1)^{m_v+m'_v})\end{aligned}$$

□

Now:

- If $m = m'$: $\forall t \quad (-1)^{m_t+m'_t} = 1$ then: $\mathcal{O}(m,m') = \sum_{t=0}^{q-1}\prod_{v=0,v\neq t}^{q-1} 2 = q2^{q-1}$
- Let u be the number of bits where m and m' differ: if $u = 1$, then:
  - $\exists t_0$ such that $m_{t_0} + m'_{t_0} = 1$ and then $(1+(-1)^{m_{t_0}+m'_{t_0}}) = 0$,
  - $\forall t \neq t_0 \quad m_t + m'_t = 0$ or 2 and then all the terms $\prod_{v=0,v\neq t}^{q-1}(1+(-1)^{m_v+m'_v}) = 0$

  Thus

$$\begin{aligned}\mathcal{O}(m,m') &= (-1)^{m_{t_0}+m'_{t_0}}\prod_{v=0,v\neq t_0}^{q-1}(1+(-1)^{m_v+m'_v}) \\ &= -2^{q-1}\end{aligned}$$

If $u > 1$, let T_u be the subset of $[0..q-1]$ such that $t \in T_u$ iff $m_t + m'_t = 1$. Then:

$$\begin{aligned}&\text{if } t \in T_u && m_t + m'_t = 1, \\ & && \text{and } [(1+(-1)^{m_t+m'_t})] = 0 \\ &\text{if } t \notin T_u && m_t + m'_t = 0 \text{ or } 2, \\ & && \text{and } [(1+(-1)^{m_t+m'_t})] = 2\end{aligned}$$

Thus $\mathcal{O}(m,m')=0$

Finally:

if m and m' differ by more than 1 bit,	$\mathcal{O}(m,m')=0$,
if m and m' differ by 1 bit,	$\mathcal{O}(m,m')=-2^{q-1}$,
if $m=m'$,	$\mathcal{O}(m,m)=q2^{q-1}$

Recall that h'_{2^q+m} can be written as:

$$\begin{aligned} h'_{2^q+m} &= h_{2^q+m}(1-\frac{p_c}{l-1}-2p_m) \\ &\quad -\frac{p_c}{2^q(l-1)}\sum_{m'=0}^{2^q-1} h_{2^q+m'}\Delta(m,m') \\ &\quad -\frac{2p_m}{2^q}\sum_{m'=0}^{2^q-1} h_{2^q+m'}\mathcal{O}(m,m') \end{aligned}$$

◇ the $\mathcal{O}(m,m')$ term yields:

$$\begin{aligned} \sum_{m'=0}^{2^q-1} h_{2^q+m'}\mathcal{O}(m,m') &= q2^{q-1}h_{2^q+m} \\ &- 2^{q-1}\sum_{m'/\exists u,|m'-m|=2^u} h_{2^q+m'} \end{aligned}$$

since m and m' differ only by 1 bit: $\exists u,\ m'=m+(1-2m_u)2^u$

Thus:

$$\sum_{m'=0}^{2^q-1} h_{2^q+m'}\mathcal{O}(m,m') = q2^{q-1}h_{2^q+m} - 2^{q-1}\sum_{t=0}^{q-1} h_{2^q+m+(1-2m_t)2^t}$$

◇ the $\Delta(m,m')$ term yields:

$$\sum_{m'=0}^{2^q-1} h_{2^q+m'}\Delta(m,m') = [1+(q-2)2^q]h_{2^q+m} + \sum_{m'=0,m'\neq m}^{2^q-1} h_{2^q+m'}\Delta(m,m')$$

for $m'\neq m$, $\exists u/\forall t\in[0..u-1]\quad m_t=m'_t$ and $m_u\neq m'_u$ (i.e. $m'_u=1-m_u$).

We can thus write:

$$m'=\sum_{t=0}^{u-1} m_t2^t+(1-m_u)2^u+\sum_{t=u+1}^{q-1} m'_t2^t \quad u\in[0..q-1]$$

And:

$$\sum_{m'=0}^{2^q-1} h_{2^q+m'}\Delta(m,m') = [1+(q-2)2^q]h_{2^q+m}+ \sum_{u=0}^{q-1}(1-2^{u+1})\sum_{r=0}^{2^{q-u-2}} h_{2^q+\sum_{t=0}^{u-1} m_t2^t+(1-m_u)2^u+r2^{u+1}}$$

Finally, h'_{2^q+m} can be written as:

$$\begin{aligned} &h'_{2^q+m} = \\ &h_{2^q+m}[1-\frac{p_c}{l-1}(1+\frac{1+(q-2)2^q}{2^q})-2p_m(1+\frac{q}{2})] \\ &-\frac{p_c}{2^q(l-1)}\sum_{u=0}^{q-1}(1-2^{u+1})\sum_{r=0}^{2^{q-u-2}} h_{2^q+\sum_{t=0}^{u-1} m_t2^t+(1-m_u)2^u+r2^{u+1}} \\ &-p_m\sum_{t=0}^{q-1} h_{2^q+m+(1-2m_t)2^t} \end{aligned}$$