Epistasis and Deceptivity

B. Naudts * A. Verschoren

Abstract

Deceptivity and epistasis both contribute to make fitness functions hard to optimize for a genetic algorithm. In this note we examine the relation between these concepts, with particular emphasis on their mutual reinforcement.

Introduction

There are several factors which can make a fitness function f hard to optimize, including deceptivity and high epistasis. By a thorough investigation of the length 2 case, we show that these properties are essentially independent, but may mutually reinforce each other. In particular, epistasis is responsable for the different behavior of the type I and type II minimal deceptive problem.

1 Epistasis

In genetics, a gene or gene pair is said to be epistatic to a gene at another locus, if it masks the (phenotypical) expression of the second one, cf. [8]. In this way, epistasis expresses links between separate genes in a chromosome. The analogous notion in the context of genetic algorithms (GAs) was introduced by Rawlins [7], who defines minimal epistasis to correspond to the situation where every gene (or bit) is independent of every other gene, whereas maximal epistasis arises when no

*Research Assistent of the Fund for Scientific Research (F.W.O.)
Received by the editors July 1997.
Communicated by A. Hoogewijs.
1991 *Mathematics Subject Classification :* 68T05, 68T30.
Key words and phrases : Epistasis, deceptivity, genetic algorithm.

proper subset of genes is independent of any other gene. Actually, this dependency always being defined with respect to some fixed (positively valued) fitness function f, the case of minimal resp. maximal epistasis corresponds to f being expressible as a linear combination of functions, each of which only depends upon a single gene, resp. to f being essentially a random function.

Of course, this description of epistasis is much too vague to be used efficiently as a computable component for GA hardness. In order to remedy this, Davidor [1] suggests the following definition for the epistasis of a length ℓ string $s = s_0 \dots s_{\ell-1}$ in the search space $\Omega = \{0,1\}^\ell$:

$$\varepsilon(s) = f(s) - \sum_{i=0}^{\ell-1} \frac{1}{2^{\ell-1}} \sum_{t\in\Omega, t_i=s_i} f(t) + \frac{\ell-1}{2^\ell} \sum_{t\in\Omega} f(t).$$

In [9, 10], this definition is rewritten as follows. Define

$$\mathbf{e} = \begin{pmatrix} \varepsilon(00\dots0) \\ \varepsilon(00\dots1) \\ \vdots \\ \varepsilon(11\dots1) \end{pmatrix} \text{ resp. } \mathbf{f} = \begin{pmatrix} f_{00\dots0} \\ f_{00\dots1} \\ \vdots \\ f_{11\dots1} \end{pmatrix},$$

where we denote for any $s \in \Omega$ by f_s the fitness value $f(s)$. For any positive integers $0 \le i, j < 2^\ell$, put

$$\mathbf{e}_{ij} = \frac{1}{2^\ell}(\ell + 1 - 2\mathrm{d}_{ij}),$$

where d_{ij} is the Hamming distance between i and j (the number of bits in which the binary representations of i and j differ, cf. [12]). Letting $\mathbf{E}_\ell = (\mathbf{e}_{ij}) \in M_{2^\ell}(\mathbb{Q})$, the rational valued 2^ℓ by 2^ℓ matrices, it is easy to see that

$$\mathbf{e} = \mathbf{f} - \mathbf{E}_\ell \mathbf{f}.$$

This allows us to define the global epistasis of f to be

$$\varepsilon_\ell(f) := \sqrt{\sum_{s\in\Omega} \varepsilon^2(s)} = \|\mathbf{e}\|.$$

Usually, it is easier to work with the matrix $\mathbf{G}_\ell = 2^\ell \mathbf{E}_\ell \in M_{2^\ell}(\mathbb{Z})$ with entries $g_{ij} = \ell + 1 - 2\mathrm{d}_{ij}$ for all $0 \le i, j < 2^\ell$. It is easy to see that

$$\mathbf{G}_\ell = \begin{pmatrix} \mathbf{G}_{\ell-1} + \mathbf{U}_{\ell-1} & \mathbf{G}_{\ell-1} - \mathbf{U}_{\ell-1} \\ \mathbf{G}_{\ell-1} - \mathbf{U}_{\ell-1} & \mathbf{G}_{\ell-1} + \mathbf{U}_{\ell-1} \end{pmatrix},$$

where

$$\mathbf{U}_{\ell-1} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in M_{2^{l-1}}(\mathbb{Z}).$$

It easily follows that $\mathbf{G}_\ell^2 = 2^\ell \mathbf{G}_\ell$, hence that $\mathbf{E}_\ell$ is idempotent. Using this, it is clear, taking into account that $\mathbf{E}_\ell$ is symmetric, that $\varepsilon_\ell^2(f) = {}^t\mathbf{f}\, \mathbf{F}_\ell\, \mathbf{f}$, where $\mathbf{F}_\ell = \mathbf{I}_\ell - \mathbf{E}_\ell$, with $\mathbf{I}_\ell$ the unit matrix of dimension 2^ℓ.

It is obvious that for any positive real number $r \in \mathbb{R}$, we have $\varepsilon(rf) = r\varepsilon(f)$, whereas the epistasis of f and rf, viewed as expressing linkage between different bits, should be the same.

This leads us to define the *normalized epistasis* of a fitness function f as

$$\begin{aligned}\varepsilon^*_\ell(f) &= \varepsilon^2_\ell\left(\frac{f}{||\mathbf{f}||}\right) = \frac{\varepsilon^2_\ell(f)}{||\mathbf{f}||} \\ &= \frac{{}^t\mathbf{f}(\mathbf{I}-\mathbf{E}_\ell)\mathbf{f}}{{}^t\mathbf{f}\,\mathbf{f}},\end{aligned}$$

and as $\mathbf{F}_\ell$ is an orthogonal projection (it is both idempotent and symmetric), it thus follows that

$$0 \leq \varepsilon^*_\ell(f) \leq 1.$$

Actually, it is easy to see that $\varepsilon^*_\ell(f) = 0$ if and only if f has minimal epistasis, in the sense of Rawlins [7]. On the other hand, it has been proved in [9] that the maximal value of $\varepsilon^*_\ell(f)$ that may be reached by a (positive valued!) fitness function f is $1 - \frac{1}{2^{\ell-1}}$. In fact, this value is reached precisely by fitness functions f with the property that there exists some $t \in \Omega$ with binary complement $\hat{t}$ and some positive real number α such that $f(t) = f(\hat{t}) = \alpha$ and $f(s) = 0$ for $t \neq s \neq \hat{t}$.

2 Deceptivity

For any length ℓ schema $H \in \{0, 1, \#\}$, let us denote by $f(H)$ the average fitness of (the binary strings represented by) H and let us call the number of fixed bits in H the order of H.

As in [6] for example, two length ℓ schemata

$$H = h_0 \ldots h_{\ell-1}, \quad H' = h'_0 \ldots h'_{\ell-1} \in \{0, 1, \#\}^\ell$$

are said to be *competing*, if they have the same order and if for any $0 \leq i < \ell$ we have $h_i = \#$ if and only if $h'_i = \#$. Let $M(f)$ be the set of global optima of f (usually $M(f)$ consists of a single element). We say that f is *deceptive of order m* if there exists $x \notin M(f)$ such that if H and H' are competing schemata of order at most m and if $x \in H$, then $f(H) > f(H')$. Although the schema H contains the optimum m, the GA is thus lead away from it, as m is more likely expected to belong to competing schemata, in view of their, misleading, higher average fitness.

Let us now work over $\Omega = \{0, 1\}^2$, i.e., let us consider length 2 strings and let us consider a fitness function f on Ω with maximum value f_{11}, i.e., with $M(f) = \{11\}$.

It is easy to see that if both $f_{00} + f_{01} \leq f_{10} + f_{11}$ and $f_{00} + f_{10} \leq f_{01} + f_{11}$, then f is non-deceptive.

However, if we assume $f_{00} + f_{01} > f_{10} + f_{11}$ for example, then f is deceptive of order 1. Indeed, the string $x = 01$ belongs to the two order-1 schemata $H_1 = 0\#$ and $H_2 = \#1$, with (unique) corresponding competing schemata $H'_1 = 1\#$ resp. $H'_2 = \#0$. We have

$$f(H_1) = \frac{1}{2}(f_{00} + f_{01}) > \frac{1}{2}(f_{10} + f_{11}) = f(H'_1).$$

On the other hand, we also have

$$f(H_2) = \frac{1}{2}(f_{01} + f_{11}) > \frac{1}{2}(f_{00} + f_{10}) = f(H_2').$$

Indeed, otherwise $f_{01} + f_{11} \leq f_{00} + f_{10}$, which, combined with $f_{00} + f_{01} > f_{10} + f_{11}$ would lead to $f_{11} - f_{00} < f_{00} - f_{11}$, a contradiction.

In the deceptive situation described in the preceding paragraph, depending on whether $f_{01} > f_{00}$ or $f_{00} \geq f_{01}$, one speaks, as in [3], of the *minimal deceptive problem of type I* resp. *of type II*. It appears that, although both are deceptive of order 1, these functions exhibit a very different convergence behaviour with respect to genetic algorithms. Indeed, although in the type I case deceptivity initially leads the GA away from the global maximum, after a sufficiently large number of generations, the GA is still capable of discovering the real maximum, whereas this is not necessary so in the type II case . We refer to [3] for full details and examples of typical runs.

In view of the previous remarks, deceptivity, which is of order 1 in both cases, does not permit to distinguish the (very different) type I and type II behavior.

3 Epistasis versus Deceptivity

In this section, we take a detailed look at the link between epistasis and deceptivity in the special case $\ell = 2$. As one easily verifies, in this case the matrix $\mathbf{G} = \mathbf{G}_2$ equals

$$\mathbf{G} = \begin{pmatrix} 3 & 1 & 1 & -1 \\ 1 & 3 & -1 & 1 \\ 1 & -1 & 3 & 1 \\ -1 & 1 & 1 & 3 \end{pmatrix}.$$

It follows that $\mathbf{F} = \mathbf{F}_2 = \mathbf{I}_2 - \mathbf{E}_2$ is given by

$$\mathbf{F} = \frac{1}{4}\begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

Write

$$\tilde{f} = f_{00} + f_{11} - f_{01} - f_{10},$$

then

$${}^t\mathbf{f}\,\mathbf{F}\,\mathbf{f} = \frac{1}{4}\begin{pmatrix} f_{00} & f_{01} & f_{10} & f_{11} \end{pmatrix}\begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}\begin{pmatrix} f_{00} \\ f_{01} \\ f_{10} \\ f_{11} \end{pmatrix} = \frac{1}{4}\tilde{f}^2.$$

The normalized epistasis of f is given by

$$\varepsilon_2^*(f) = \frac{{}^t\mathbf{f}\,\mathbf{F}\,\mathbf{f}}{{}^t\mathbf{f}\,\mathbf{f}},$$

so, restricting (as we may) to the case $||\mathbf{f}|| = 1$, we have that $\varepsilon^*(f)$ is proportional to

$$\tilde{f}^2 = (f_{11} + f_{00} - f_{01} - f_{10})^2.$$

Let us assume throughout f_{11} to be the maximum value of f. The maximal value of f is then reached for

$$\mathbf{f} = \begin{pmatrix} \alpha \\ 0 \\ 0 \\ \alpha \end{pmatrix}$$

with $\alpha = \sqrt{2}/2$. Here, $\tilde{f} = 2\alpha = \sqrt{2}$, so $\varepsilon_2^*(f) = 1/2$ (as ${}^t\mathbf{f}\,\mathbf{f} = 2\alpha^2 = 1$). Let us fix the set $\{a, b, c, d\}$ of values of f and assume that $a = f_{11} > b > c > d$, then it is clear that $\tilde{f} = f_{11} + f_{00} - f_{01} - f_{10}$ can take three values:

1. if
$$\mathbf{f} = \begin{pmatrix} b \\ c \\ d \\ a \end{pmatrix} \text{ or } \mathbf{f} = \begin{pmatrix} b \\ d \\ c \\ a \end{pmatrix}$$
then $\tilde{f} = \alpha_1 = a + b - c - d$;

2. if
$$\mathbf{f} = \begin{pmatrix} c \\ b \\ d \\ a \end{pmatrix} \text{ or } \mathbf{f} = \begin{pmatrix} c \\ d \\ b \\ a \end{pmatrix}$$
then $\tilde{f} = \alpha_2 = a + c - b - d$;

3. if
$$\mathbf{f} = \begin{pmatrix} d \\ b \\ c \\ a \end{pmatrix} \text{ or } \mathbf{f} = \begin{pmatrix} d \\ c \\ b \\ a \end{pmatrix}$$
then $\tilde{f} = \alpha_3 = a + d - b - c$.

Since $a > b > c > d$, it is clear that $\alpha_1 > 0$ and $\alpha_2 > 0$. Moreover, $b - c > 0$ (and $c - b < 0$), so $\alpha_1 > \alpha_2$.

On the other hand, if $a + d > b + c$, then $a_3 > 0$ and if $a + d < b + c$, then $\alpha_3 < 0$. Clearly, $\alpha_2 > \alpha_3$, as $c - d > 0$ (and $d - c < 0$) and $\alpha_2 > -\alpha_3$, as $a - b > 0$ (and $b - a < 0$).

We have thus shown that

$$\alpha_1^2 > \alpha_2^2 > \alpha_3^2.$$

Since $\varepsilon_2^*(f)$ is proportional to $\tilde{f}^2$, this yields:

Proposition 3.1. *With a fixed set of values $f_{00}, f_{01}, f_{10} < f_{11}$, the normalized epistasis $\varepsilon_2^*(f)$ can take three different values $\varepsilon_1 < \varepsilon_2 < \varepsilon_3$, where*

1. *ε_1 corresponds to $f_{01} > f_{10} > f_{00}$ or $f_{10} > f_{01} > f_{00}$* (low epistasis)*;*

2. *ε_2 corresponds to $f_{01} > f_{00} > f_{10}$ or $f_{10} > f_{00} > f_{01}$* (medium epistasis)*;*

3. *ε_3 corresponds to $f_{00} > f_{01} > f_{10}$ or $f_{00} > f_{10} > f_{01}$* (high epistasis).

Let us now relate the previous result to deceptivity.

If f has *low* epistasis $\varepsilon_2^*(f) = \varepsilon_1$, then $f_{00}+f_{01} < f_{10}+f_{11}$ and $f_{00}+f_{10} < f_{01}+f_{11}$, so f is non-deceptive. So, if f is deceptive, it has necessarily medium or high epistasis.

If f is deceptive and has *medium* epistasis, i.e., if $\varepsilon_2^*(f) = \varepsilon_2$, then f is deceptive of type I. Indeed, if we assume $f_{00} + f_{01} > f_{10} + f_{11}$ for example, then we are necessarily in the situation $f_{01} > f_{00} > f_{10}$ (otherwise $f_{10} > f_{00} > f_{01}$, hence $f_{10} + f_{11} > f_{00} + f_{01}$, a contradiction). So f is deceptive of type I, as $f_{00} < f_{01}$.

If f is deceptive and has *high* epistasis, i.e., if $\varepsilon_2^*(f) = \varepsilon_3$, then f is deceptive of type II, as in each of the possibilities $f_{00} > f_{01} > f_{10}$ and $f_{00} > f_{10} > f_{01}$, we have $f_{01} < f_{00}$.

Conclusion 3.2. *We have proved:*

1. *deceptivity cannot occur in the low epistasis case;*

2. *if deceptivity occurs, epistasis allows to differentiate between type I deceptivity (medium epistasis) and type II deceptivity (high epistasis).*

In particular, the fact that deceptive fitness functions of type II are much harder to optimize than their type I analogues is explained by the extra difficulty implied by their epistatic behaviour.

Note 3.3. Let us note that although deceptive functions cannot have low epistasis, the converse is not necessarily true, i.e., even fitness functions with high epistasis are not necessarily deceptive (but remain difficult to optimize precisely due to this high epistasis).

Indeed, consider the fitness function f given by

$$\mathbf{f} = \begin{pmatrix} \alpha - 2\varepsilon \\ \varepsilon \\ 0 \\ \alpha \end{pmatrix},$$

where $\alpha = \sqrt{2}/2$ and $\varepsilon > 0$. As

$$f_{00} + f_{01} = \alpha - \varepsilon < \alpha = f_{10} + f_{11}$$

and

$$\alpha - 2\varepsilon = f_{00} + f_{10} < f_{01} + f_{11} = \alpha$$

it follows that f is non-deceptive. However, if we choose $\varepsilon < \alpha/3$, then

$$f_{11} > f_{00} > f_{01} > f_{10},$$

so, clearly, f has high epistasis. Actually, it is easy to see that

$$\varepsilon_2^*(f) = \frac{1}{4}\frac{(2\alpha - 3\varepsilon)^2}{(\alpha - 2\varepsilon)^2 + \varepsilon^2 + \alpha^2},$$

so, for small values of ε, we find that

$$\varepsilon_2^*(f) \approx \frac{1}{4}\frac{4\alpha^2}{2\alpha^2} = \frac{1}{2},$$

which is the highest possible value that $\varepsilon_2^*(f)$ may reach. It thus follows that even maximally epistatic fitness functions are not necessarily deceptive.

Note 3.4. A similar type of analysis may be done for higher dimensions, in particular, by using Walsh transforms as in [4, 5]. On the other hand, one should realize that for $\ell \geq 3$, the situation becomes much more complex. This is mainly due to the higher dimensionality of the function space, which eliminates some of the constraints which lead to the tight links between epistasis and deceptivity in the 2-bit case. In particular, in higher dimensions no intrinsic classification of deceptivity (type I versus type II) is available, and other factors than high epistasis or deceptivity may account for the GA hardness of a fitness function.

References

[1] Y. Davidor, *Epistasis Variance: A Viewpoint on GA-Hardness*, in: Foundations of Genetic Algorithms 1, ed. G. J. E. Rawlins, Morgan Kaufmann Publishers, San Mateo, 1991.

[2] S. Forrest and M. Mitchell, *Relative Building Block Fitness and the Building Block Hypothesis*, in: Foundations of Genetic Algorithms 2, ed. L. D. Whitley, Morgan Kaufmann Publishers, San Mateo, 1993.

[3] D. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, 1989,

[4] D. Goldberg, *Genetic Algorithms and Walsh Functions: Part I. A Gentle Introduction*, Complex Systems **3** (1989) 129-152.

[5] D. Goldberg, *Genetic Algorithms and Walsh Functions: Part II. Deception and Its Analysis*, Complex Systems **3** (1989) 153-171.

[6] G. Liepins and M. Vose, *Deceptiveness and Genetic Algorithm Dynamics*, in: Foundations of Genetic Algorithms 1, ed. L. D. Whitley, Morgan Kaufmann Publishers, San Mateo, 1991.

[7] G. J. E. Rawlins, *Foundations of Genetic Algorithms,* Morgan Kaufmann Publishers, San Mateo, 1991.

[8] M. M. Stickberger, *Genetics,* Collier-McMillan Ltd, London, 1968.

[9] D. Suys and A. Verschoren, *Extreme Epistasis*, International Conference on Intelligent Technologies in Human-Related Sciences (ITHURS'96) Vol. II, 251-258, León, 1996.

[10] H. Van Hove and A. Verschoren, *On Epistasis*, Computers and AI, **14** (1994) 271-277.

[11] H. Van Hove and A. Verschoren, *What is Epistasis?*, Bull. Belg. Math. Soc. - Simon Stevin, **5** (1998) 69-77.

[12] J. H. Van Lint, *Introduction to Coding Theory,* Springer Verlag, New York, 1982.

University of Antwerp, RUCA
Department of Mathematics and Computer Science
Groenenborgerlaan 171, 2020 Antwerpen, Belgium