

Mutation Rates of the (1+1)-EA on Pseudo-Boolean Functions of Bounded Epistasis

Andrew M. Sutton

L. Darrell Whitley

Adele E. Howe

Department of Computer Science
Colorado State University
Fort Collins, CO 80523
{sutton,whitley,howe}@cs.colostate.edu

ABSTRACT

When the epistasis of the fitness function is bounded by a constant, we show that the expected fitness of an offspring of the (1+1)-EA can be efficiently computed for any point. Moreover, we show that, for any point, it is always possible to efficiently retrieve the “best” mutation rate at that point in the sense that the expected fitness of the resulting offspring is maximized.

On linear functions, it has been shown that a mutation rate of $1/n$ is provably optimal. On functions where epistasis is bounded by a constant k , we show that for sufficiently high fitness, the commonly used mutation rate of $1/n$ is also best, at least in terms of maximizing the expected fitness of the offspring. However, we find for certain ranges of the fitness function, a better mutation rate can be considerably higher, and can be found by solving for the real roots of a degree- k polynomial whose coefficients contain the nonzero Walsh coefficients of the fitness function. Simulation results on maximum k -satisfiability problems and NK-landscapes show that this expectation-maximized mutation rate can cause significant gains early in search.

Categories and Subject Descriptors

I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search

General Terms

Theory, Algorithms, Performance

Keywords

Evolutionary Algorithms, Walsh Analysis

1. INTRODUCTION

Evolutionary Algorithms (EAs) are probabilistic direct search methods that are often applied to the task of func-

tion optimization. It is well-understood that the choice of the *mutation rate* parameter can have a strong impact on the performance of EAs, and a large number of experimental and theoretical investigations have been carried out to determine the optimal mutation rate.

Many experimental studies have suggested a mutation rate between 0.001 and 0.01 [7, 11, 23]. In many cases, however, mutation rates that cause an EA to perform well on one class of functions may produce poor performance on another class of functions. Indeed, Droste et al. [9] have given theoretical evidence that a mutation rate of $1/n$ (where n is the input size) guarantees convergence in $O(n \log n)$ time for the (1+1)-EA applied to linear functions. On the other hand, Jansen and Wegener [17] have introduced a function for which a mutation rate of $1/n$ leads to superpolynomial runtime of the (1+1)-EA with high probability while a mutation rate of $\frac{\log n}{n}$ leads to expected polynomial time convergence on the same function. Furthermore, Doerr et al. [8] present a particular pseudo-Boolean function for which changing the mutation rate of the (1+1)-EA by a constant factor leads to an exponential performance gap. Such results stress the importance of an understanding of the relationship between the mutation rate and the function being searched.

On linear functions, this relationship is well-understood. For instance, in the case of ONEMAX, it is straightforward to derive an analytical expression for the probability of a successful mutation [1]. In the case of general pseudo-Boolean functions, the probability of a successful mutation from any arbitrary point is difficult to know. Furthermore, analytical expressions specifying optimal mutation rates have not previously been derived for epistatically bounded pseudo-Boolean functions.

In this paper, we focus on the simple case of the (1+1)-EA. We show that when the fitness function is a pseudo-Boolean function whose epistasis is bounded by a constant, it is always possible to efficiently compute the *expected fitness* of a mutation from each string for a given rate. We show that it is then always possible to solve for the mutation rate that maximizes the expected fitness of the offspring for any point.

1.1 The (1+1)-EA

We concentrate on the (1+1)-EA applied to the task of maximizing pseudo-Boolean functions, i.e., those of the form $f : \{0, 1\}^n \rightarrow \mathbb{R}$. The (1+1)-EA has been subject to a number of theoretical studies [9, 10, 17, 21, 22]. The algorithm is presented below, parameterized by mutation rate ρ .

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO'11, July 12–16, 2011, Dublin, Ireland.

Copyright 2011 ACM 978-1-4503-0557-0/11/07 ...\$10.00.

(1+1)-EA(ρ)

```

1 Choose  $x \in \{0, 1\}^n$  uniformly at random
2 while stopping criteria not met
3   do
4      $y \leftarrow x$ 
5     Flip each bit of  $y$  independently with prob.  $\rho$ 
6     if  $f(y) \geq f(x)$ 
7       then  $x \leftarrow y$ 

```

The mutation rate parameter ρ controls the degree to which each search point is perturbed to produce the next search point. Often, a constant mutation rate of $\rho = 1/n$ is recommended [3, 21], especially for linear functions. On functions with nonlinearity, there is strong evidence that the optimal mutation rate is time-dependent [1, 4, 15, 16].

For some functions, it is possible to compute the exact probability of a successful mutation as a function of *fitness level* and mutation rate [3, 17]. This is especially useful in the case of runtime analysis because it allows one to bound the expected number of mutations until a successful offspring is produced. However, in the case of general pseudo-Boolean functions, this probability is difficult to compute. When this probability is not known, one solution is to use *self-adaptation* [2] in which each individual is augmented with an encoding of its own mutation rate and the rate is adapted along with the function parameters.

While linear, unimodal functions have a provably optimal mutation rate of $1/n$, Bäck pointed out that when the fitness is multimodal, a search for a dynamically varying mutation rate different from a constant value may be worthwhile to overcome local optima [3]. On the classical LEADINGONES function, Böttcher, Doerr, and Neumann [5] derived an exact expression for the expected runtime of a (1+1)-EA as a function of mutation rate and proved that the standard $1/n$ rate recommended for linear functions is suboptimal. They also showed that a time-varying mutation rate can improve the runtime. Hesser and Männer [15] presented a theoretical argument that suggested the mutation probability in a population-based GA employing crossover should decrease with time. In this paper we also find evidence for this on epistatically k -bounded functions. In fact, we find that each point has its own “expectation-best” mutation rate that maximizes the expected fitness of its offspring. This rate changes in response to the relationship between the fitness of the point itself and the expected fitness of points that lie within Hamming distance k .

We now show that on an epistatically k -bounded pseudo-Boolean function, even if we cannot recover the optimal mutation rate (in terms of success probability) for a point, we can at least efficiently compute the mutation rate that maximizes the expected fitness of the offspring.

2. PRELIMINARIES

Let $f : \{0, 1\}^n \rightarrow \mathbb{R}$ be a pseudo-Boolean function. We say f is k -bounded if it can be written as the sum over subfunctions that each depend on at most k bits. In this paper we will assume that the range of f is nonnegative. In the context of function optimization and search, adding an arbitrary constant to satisfy this constraint will not affect the behavior of the algorithm.

For any $x \in \{0, 1\}^n$ we will denote the b^{th} element of the string x as $x[b]$. We denote $|\{b : x[b] = 1\}|$ simply as $|x|$. Any k -bounded pseudo-Boolean function can be written in

its orthogonal Walsh basis expansion

$$f(x) = \sum_{i=0}^{2^n-1} w_i \psi_i(x)$$

where w_i is a real value corresponding to the i^{th} Walsh coefficient and $\psi_i(x)$ is the i^{th} Walsh function defined as

$$\psi_i(x) = \prod_{b=1}^n (-1)^{x[b]i[b]}.$$

Here, the i in the exponent is taken to be in its length- n binary string representation.

Let $x \in \{0, 1\}^n$. The Hamming sphere of radius r around x is the set of points $\{y \in \{0, 1\}^n : \mathcal{H}(x, y) = r\}$. We will write the *expectation* of f over a sphere of radius r around x as

$$\mathbb{S}_x^r = \binom{n}{r}^{-1} \sum_{\{y: \mathcal{H}(x, y) = r\}} f(y),$$

this is interpreted as the expected fitness of a bitstring drawn uniformly at random from points lying at radius r from x .

Sutton et al. [24] showed that

$$\mathbb{S}_x^r = \binom{n}{r}^{-1} \sum_{i: w_i \neq 0} \mathcal{K}_r(|i|, n) w_i \psi_i(x) \quad (1)$$

where $\mathcal{K}_r(|i|, n)$ is the order- r Krawtchouk polynomial [20],

$$\mathcal{K}_r(|i|, n) = \sum_{j=0}^r \binom{|i|}{j} \binom{n-|i|}{r-j} (-1)^j.$$

In the case that f is k -bounded there are at most a polynomial number of nonzero Walsh coefficients w_i (see e.g., Heckendorn et al. [13]); hence \mathbb{S}_x^r can always be efficiently computed. Furthermore, so long that f is k -bounded, if the Walsh coefficients are not known *a priori*, they can be efficiently retrieved deterministically in $O(n^k)$ time [18], or stochastically with negligible error in $O(n^2 \log n)$ time [14]. If there are m nonzero Walsh coefficients, Choi et al. [6] present an $O(m \log n)$ adaptive randomized algorithm for finding all of them with high probability.

3. MUTATION RATES

Applying standard mutation to an element $x \in \{0, 1\}^n$ is analogous to performing n independent Bernoulli trials to determine whether or not to change each bit of x . Thus, the probability that the offspring of x under mutation with rate ρ lies at Hamming distance r from x is distributed binomially: $\binom{n}{r} \rho^r (1-\rho)^{n-r}$. Letting

$$a_r = \sum_{i: w_i \neq 0} \mathcal{K}_r(|i|, n) w_i \psi_i(x), \quad (2)$$

we can immediately compute the expected fitness of the offspring as

$$\begin{aligned} \mathbb{M}_x(\rho) &= \sum_{r=0}^n \binom{n}{r} \rho^r (1-\rho)^{n-r} \mathbb{S}_x^r \\ &= \sum_{r=0}^n a_r \rho^r (1-\rho)^{n-r} \quad \text{by (1)} \end{aligned} \quad (3)$$

We can re-express Equation (3) as a degree- n polynomial in ρ as

$$\mathbb{M}_x(\rho) = A_0 + A_1\rho + A_2\rho^2 + \cdots + A_n\rho^n \quad (4)$$

where

$$A_m = \sum_{\ell=0}^m a_{m-\ell} \binom{n-m+\ell}{\ell} (-1)^\ell. \quad (5)$$

When f is epistatically bounded by a constant, the Walsh coefficients can be found in polynomial time and the coefficients A_m can be efficiently computed. Later, we will also see that it is possible to further bound the degree of this polynomial.

To find the mutation rate ρ which maximizes the expected fitness of the offspring of x we simply need to find

$$\arg \max_{0 \leq \rho \leq 1} \mathbb{M}_x(\rho) = \arg \max_{0 \leq \rho \leq 1} A_0 + A_1\rho + A_2\rho^2 + \cdots + A_n\rho^n.$$

The first and second derivatives of the expected fitness are

$$\frac{d}{d\rho} \mathbb{M}_x(\rho) = A_1 + 2A_2\rho + 3A_3\rho^2 + \cdots + nA_n\rho^{n-1}$$

and

$$\frac{d^2}{d\rho^2} \mathbb{M}_x(\rho) = 2A_2 + 6A_3\rho + 12A_4\rho^2 + \cdots + n(n-1)A_n\rho^{n-2}.$$

It is easy to find the stationary points of $\mathbb{M}_x(\rho)$ by numerically solving for the real roots of $\frac{d}{d\rho} \mathbb{M}_x(\rho)$. Of course, we can use the so-called “second derivative test” to test for concavity and solve for the local maxima point set

$$M = \left\{ \rho : \frac{d}{d\rho} \mathbb{M}_x(\rho) = 0 \text{ and } \frac{d^2}{d\rho^2} \mathbb{M}_x(\rho) < 0 \right\}.$$

The mutation rate that maximizes the expected fitness of the offspring is easily retrieved by finding the point $\rho^* \in (M \cap [0, 1]) \cup \{0, 1\}$ such that $\mathbb{M}_x(\rho^*)$ is maximal.

3.1 Degeneracy: when no mutation is “best”

The polynomial defined in (4) always has a maximum in the interval $[0, 1]$. The degenerate case is when $\mathbb{M}_x(\rho)$ is monotonically decreasing and no stationary points lie within the interval. Moreover, it is possible that any maxima lying within the interval have evaluation strictly less than $\mathbb{M}_x(0)$. In this case, the “optimal” value is $\rho^* = 0$. Since $\mathbb{M}_x(0) = f(x)$, this means that there is no possible mutation rate (constant across bitstrings) that will produce an offspring whose expectation is greater than $f(x)$: the fitness of the current point. This corresponds to a local optimum in “expectation space”, that is, any mutation is disimproving in expectation. In Sections 4 and 5 we will find conditions on the fitness of f in which this degeneracy must hold for linear functions and $k > 1$ -bounded functions, respectively.

Choosing a suitable nonzero mutation rate.

When any mutation rate is expected to produce an offspring with lower fitness, the optimal choice to maximize expected fitness is to perform no mutation. Instead, we would like to perform mutations that, in some sense, minimize the expected loss in fitness.

Suppose $\rho^* = 0$. Let $0 < \rho \ll 1$ be any positive value close to zero. Then we know

$$\mathbb{M}_x(\rho) = (1-\rho)^n f(x) + \sum_{r=1}^n \rho^r (1-\rho)^{n-r} \mathbb{S}_x^r < \mathbb{M}_x(0) = f(x).$$

Ignoring the higher order terms we can write

$$(1-\rho)^n f(x) \leq \mathbb{M}_x(\rho) < f(x).$$

Choosing a mutation rate $\rho = k/n$ means that in expectation, k bits will be changed. We can recover the “standard” recommended mutation rate of $\rho = 1/n$ by observing from the above inequality,

$$\left(1 - \frac{k}{n}\right)^n f(x) \leq \mathbb{M}_x\left(\frac{k}{n}\right) < f(x).$$

Asymptotically we have

$$e^{-k} f(x) \leq \mathbb{M}_x\left(\frac{k}{n}\right) < f(x).$$

The lower bound on $\mathbb{M}_x\left(\frac{k}{n}\right)$ is maximized when $k = 1$. Thus, when the offspring is expected to be disimproving, the mutation rate $1/n$ maximizes the lower bound on the expectation of the fitness of the offspring under the constraint that we flip at least 1 bit in expectation. Moreover, in this case we know the expected fitness of the offspring of x is asymptotically bounded below by $e^{-1} f(x)$.

The slope of the $\mathbb{M}_x(\rho)$ polynomial at zero tells us how quickly the expected fitness falls off by choosing close-to-zero mutation rates. Interestingly, this slope is exactly n times the difference between the fitness of the current point and the average fitness over the immediate neighbors at Hamming distance 1. This can be derived easily by observing that

$$\frac{d}{d\rho} \mathbb{M}_x(0) = A_1,$$

so by (5), the slope of the $\mathbb{M}_x(\rho)$ polynomial at zero is equal to $a_1 - na_0$. From (2) it is easy to see that

$$a_1 = \sum_{i:w_i \neq 0} \mathcal{K}_1(|i|, n) w_i \psi_i = n \mathbb{S}_x^1 \quad \text{by (1)}$$

and,

$$a_0 = \sum_{i:w_i \neq 0} \mathcal{K}_0(|i|, n) w_i \psi_i = f(x)$$

so the rate of change at zero is equal to $n(\mathbb{S}_x^1 - f(x))$. Of course this makes intuitive sense: the change in expectation of very small mutations is completely determined by the difference between the current point and its immediate neighbors.

4. LINEAR FUNCTIONS

Many results exist for the linear (or *separable*) case, that is, $k = 1$ -bounded pseudo-Boolean functions. In the case of a linear function, the Walsh basis expansion has only nonzero terms in the zeroth and first order. Therefore,

$$\begin{aligned} a_r &= \mathcal{K}_r(0, n) w_0 + \sum_{i:|i|=1} \mathcal{K}_r(1, n) w_i \psi_i(x) && \text{by (2)} \\ &= \binom{n}{r} w_0 + \sum_{i:|i|=1} \sum_{j=0}^r \binom{1}{j} \binom{n-1}{r-j} w_i \psi_i(x) \\ &= \binom{n}{r} \left(w_0 + \frac{n-2r}{n} (f(x) - w_0) \right) \end{aligned}$$

and since $w_0 = 2^{-n} \sum_{x \in \{0,1\}^n} f(x) = \bar{f}$ is the average fitness over all bitstrings (see e.g., [12]),

$$= \binom{n}{r} \left(\bar{f} + \frac{n-2r}{n} (f(x) - \bar{f}) \right).$$

Substituting a_r in (5) we get

$$A_m = \sum_{\ell=0}^m \left(\bar{f} + \frac{n-2(m-\ell)}{n} (f(x) - \bar{f}) \right) \times \binom{n}{m-\ell} \binom{n-(m-\ell)}{\ell} (-1)^\ell,$$

which simplifies significantly to

$$A_m = \begin{cases} f(x) & \text{if } m = 0; \\ 2(\bar{f} - f(x)) & \text{if } m = 1; \\ 0 & \text{otherwise.} \end{cases}$$

Thus if f is a linear (1-bounded) pseudo-Boolean function, the expected fitness of an offspring using mutation rate ρ is simply

$$\mathbb{M}_x(\rho) = f(x) + 2(\bar{f} - f(x))\rho.$$

So in the case of linear functions, the polynomial terms of order greater than one vanish and the $\mathbb{M}_x(\rho)$ polynomial is always a line with y -intercept $f(x)$ and slope equal to twice the difference between the mean fitness and the fitness of x . Therefore we have recovered the well-known result for linear functions that when $f(x) < \bar{f}$, $\mathbb{M}_x(1)$ is maximal (since the slope is positive) and when $f(x) > \bar{f}$, $\mathbb{M}_x(0)$ is maximal (negative slope). On such functions, large mutations are quickly able to reach the mean value, after which the smallest mutation probability *that still flips at least one bit in expectation*, namely $1/n$ maximizes the expected fitness of the offspring. This agrees somewhat with the result of Droste et al. [9] that on linear functions the (1+1)-EA converges in $O(n \log n)$ steps with this mutation rate, and a constant mutation rate of much larger or much smaller results in provably longer convergence times.

Here we also see a weakness in relying solely on the expected fitness of the offspring to choose a mutation rate. In the case of linear functions, when $f(x) = \bar{f}$, the mutation rate that maximizes the *probability of success* is equal to $\frac{n}{2}$ (see e.g., [1]). However, using with the $\mathbb{M}_x(\rho)$ polynomial, when $f(x) = \bar{f}$, the A_1 term vanishes and $\mathbb{M}_x(\rho)$ is a constant function: all mutation rates give the same expectation of $f(x)$.

We can thus conclude that success probability, when available, presents better high-resolution information about the *optimal* mutation rate: i.e., that which maximizes the probability of a successful offspring. However, on general k -bounded pseudo-Boolean functions where that probability is unknown or difficult to compute, the expectation of fitness offers a compromise.

5. FUNCTIONS OF BOUNDED EPISTASIS

Linear functions, while amenable to analysis, are a somewhat restricted class of fitness functions. At the other extreme, the entire set of general pseudo-Boolean functions is rather expansive. The class of pseudo-Boolean functions

whose epistasis is bounded by some constant k contains fitness functions that can be very difficult for evolutionary algorithms, and includes a collection of NP-hard optimization problems such as maximum k -satisfiability and the unrestricted model of NK-landscapes.

In the last section we saw that the $\mathbb{M}_x(\rho)$ polynomial coefficients A_m vanish for $m > 1$. As expected, one can generalize the result to k -bounded pseudo-Boolean functions.

PROPOSITION 1. *Let f be an arbitrary k -bounded pseudo-Boolean function. Consider the $\mathbb{M}_x(\rho)$ polynomial for any $x \in \{0,1\}^n$. If $m > k$ then $A_m = 0$.*

PROOF SKETCH. The proof is based on showing that the sum of the first $k+1$ terms $\ell = 0, \dots, k$ of A_m in (5) is exactly equal to the additive inverse of the sum of the last $m-k$ terms $\ell = (k+1), \dots, m$. This is mostly tedious manipulation and is thus omitted here. \square

It follows from Proposition 1 that in general, when f is epistatically bounded by a constant k , the expected fitness of the offspring is a degree k polynomial in the mutation rate. In order to compute the rate ρ^* that yields the maximal expected fitness, it is enough to solve for the real roots of the degree $k-1$ polynomial $\frac{d}{d\rho} \mathbb{M}_x(\rho)$ as described above.

Proposition 1 also asserts that only the fitness of the points that lie within Hamming distance k of any individual completely determine the expected fitness of the offspring since any A_m contains terms involving \mathbb{S}_x^r only for $r < m$. Thus, as in the linear case, it is enough to compute the mean fitness in Hamming spheres out to radius k . It follows that if the mean fitness in spheres of radius one to k are strictly less than the fitness at x , $\mathbb{M}_x(\rho)$ is degenerate and no mutation rate will produce offspring with expected fitness above the fitness at x . We formalize this argument in the following lemma.

LEMMA 1. *If $\mathbb{S}_x^r < \mathbb{S}_x^0$ for all $0 < r \leq k$, then $\mathbb{M}_x(0)$ is maximal.*

PROOF. Choose $0 < \rho \leq 1$. Then

$$\begin{aligned} \mathbb{M}_x(\rho) &= (1-\rho)^n \mathbb{S}_x^0 + \sum_{r=1}^n \binom{n}{r} \rho^r (1-\rho)^{n-r} \mathbb{S}_x^r \\ &< (1-\rho)^n \mathbb{S}_x^0 + \sum_{r=1}^n \binom{n}{r} \rho^r (1-\rho)^{n-r} \mathbb{S}_x^0 \\ &= \mathbb{S}_x^0 \left(\sum_{r=1}^n \binom{n}{r} \rho^r (1-\rho)^{n-r} \right) \\ &< \mathbb{M}_x(0). \end{aligned}$$

Since the choice of ρ was arbitrary, and, by Proposition 1 all coefficients above k vanish, it follows that $\mathbb{M}_x(0)$ is maximal. \square

Clearly, when $k > 1$, k -bounded functions do not have simple linear $\mathbb{M}_x(\rho)$ polynomials as we saw in the previous section. To illustrate this we plot the $\mathbb{M}_x(\rho)$ polynomials for several random points drawn from various k -bounded functions in Figure 1.

Heckendorn et al. [13] among others proposed using NK-landscapes and k -satisfiability problems as test problem domains for evolutionary algorithms. Since both are representative k -bounded pseudo-Boolean functions we now report

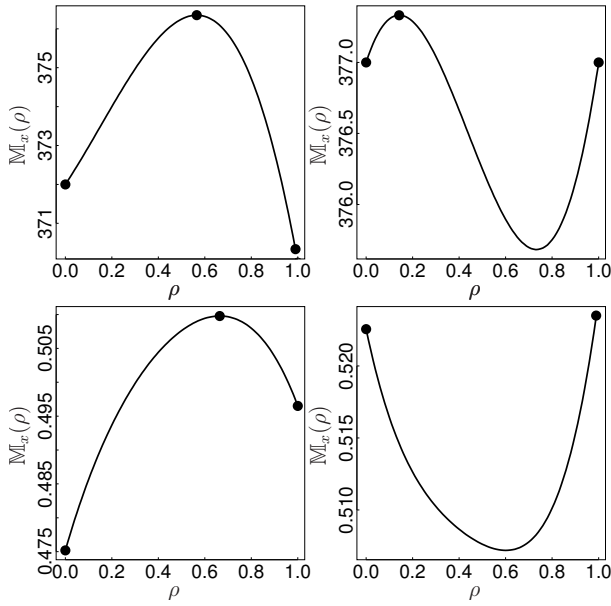


Figure 1: $\mathbb{M}_x(\rho)$ polynomials for random points in the MAX-3-SAT search space ($n = 100$) [TOP] and NK-landscapes ($N = 100, K = 3$) [BOTTOM].

the results of a number of numerical simulations that study the ρ^* mutation rate and compare it to static rates found in the literature.

5.1 Unrestricted NK-landscapes

The NK-landscape model [19] is a stochastic method for constructing fitness functions over binary sequences of length n . The model was developed to study how epistasis affects the ruggedness of the fitness landscape. The fitness function for the NK model is defined as

$$f(x) = \frac{1}{n} \sum_{j=1}^n g_j(x[j], x[b_1^{(j)}], x[b_2^{(j)}], \dots, x[b_K^{(j)}])$$

where $g_j : \{0, 1\}^{K+1} \rightarrow [0, 1]$ gives the fitness contribution of the j^{th} bit in x , and K other bits $\{b_i^{(j)}\}$. Typically, the codomain values for g_j are generated uniformly at random and fixed during search.

There are two variants of the NK model. In the *adjacent* model, the set of K bits $\{b_i^{(j)}\}$ that interact epistatically with bit j are adjacent to bit j on the bit string. In the *unrestricted* model (sometimes called the *random* model), the epistatic bit pattern $\{b_i^{(j)}\}$ for the j^{th} bit is drawn randomly (and fixed) from the $n - 1$ remaining bits. Thus for each bit j , there are $\binom{n-1}{K}$ possible selections for the set $\{b_i^{(j)}\}$. Since the fitness function is expressed as the sum of n functions each of which depends only on a single bit and the K bits in its epistatic pattern, the function is epistatically bounded by $K + 1$. Wright et al. [25] proved that the problem of finding the global optimum for the adjacent model is in P by giving a P-time dynamic programming solution. Moreover, they proved that the unrestricted model is NP-hard. In this paper, we concentrate on the unrestricted NK model.

To illustrate the behavior of the optimal rate ρ^* we performed 500 trials of 500 generations each of the (1+1)-EA employing three different mutation rates: 1) the commonly

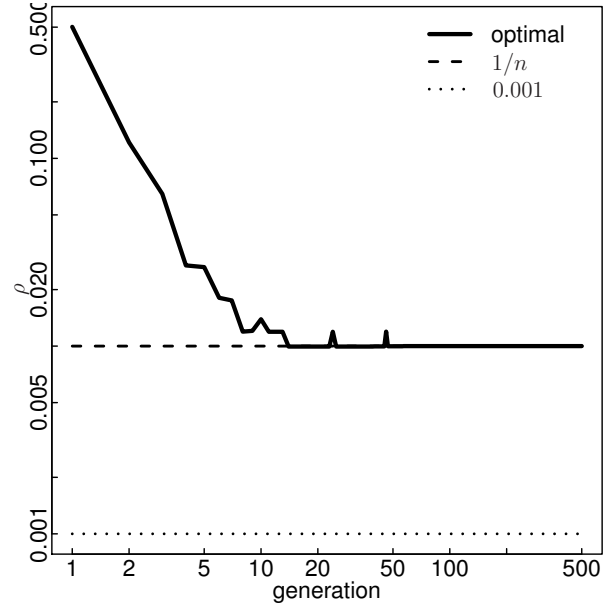


Figure 2: Log-log plot of mean mutation rates for (1+1)-EA on 500 trials of 500 generations each on an unrestricted NK-landscape model with $n = 100, K = 3$.

recommended $1/n$, 2) a “hard-wired” rate of 0.001, and 3) the expectation-optimal rate given by the maximum of $\mathbb{M}_x(\rho)$ at each point. The NK-landscape function parameters were $n = 100$ and $K = 3$. In each case, the extra time necessary to solve for the real roots of $\frac{d}{d\rho}\mathbb{M}_x(\rho)$ was negligible since the degree of the polynomial was so low ($k = 4$). When the expectation maximal rate ρ^* reaches degeneracy (i.e., the maximum is at $\mathbb{M}_x(0)$), we revert to the mutation rate of $1/n$ that we proved in Section 3.1 maximizes the expected fitness of the offspring while imposing the constraint that at least some bits are flipped in expectation.

In Figures 2 and 3 we plot the average mutation rate ρ as a function of generation. As the fitness of the points remain below the average fitness within Hamming radius k , there is a significant increase from the recommended rate of $1/n$. However, very quickly the fitness of the point exceeds the expectation of the fitness within radius k and $\mathbb{M}_x(\rho)$ reaches degeneracy and the rate reverts to $1/n$. Until degeneracy is attained, this trend coincides with claims of others [3, 15] that the best mutation rate tends to decrease during search.

On the NK-landscape with bounded epistasis, the optimal rate ρ^* reverts quickly to the recommended rate of $1/n$. However, we see that it leads to significant gains very early in search. In Figure 4 we plot the mean fitness values over time of the (1+1)-EA for a representative NK-landscape using the three mutation rate schemes. Before the ρ^* rate becomes degenerate, significant gains are made over the static $1/n$ rate.

5.2 Maximum k -satisfiability

Another well-studied problem that yields a k -bounded fitness function is the NP-hard maximum k -satisfiability problem (MAX- k -SAT). An instance of MAX- k -SAT is given by a Boolean formula over n variables in conjunctive normal

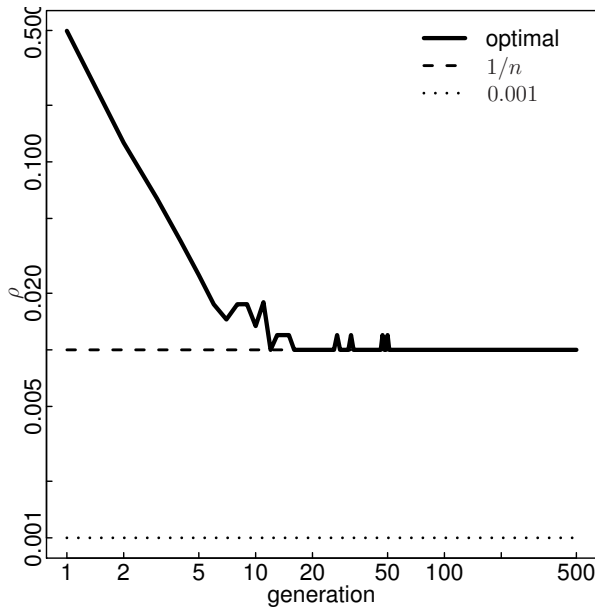


Figure 3: Log-log plot of mean mutation rates for (1+1)-EA on 500 trials of 500 generations each on an unrestricted NK-landscape model with $n = 100$, $K = 2$.

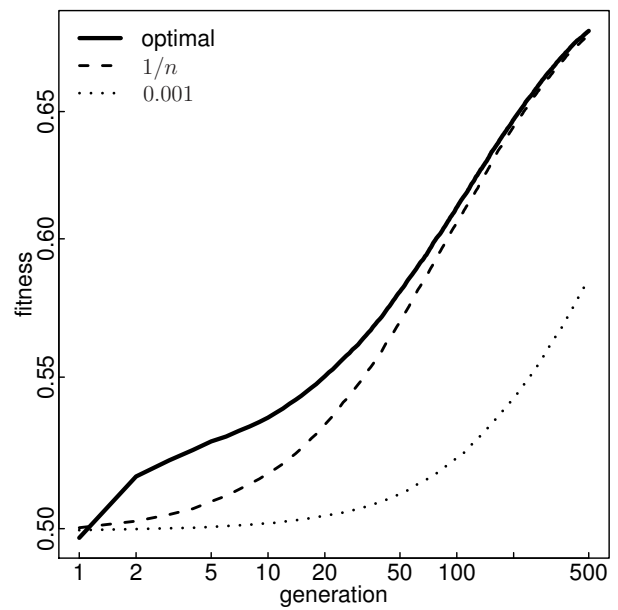


Figure 5: Log-log plot of mean fitness of (1+1)-EA on 500 trials of 500 generations each on an unrestricted NK-landscape model with $n = 100$, $K = 2$.

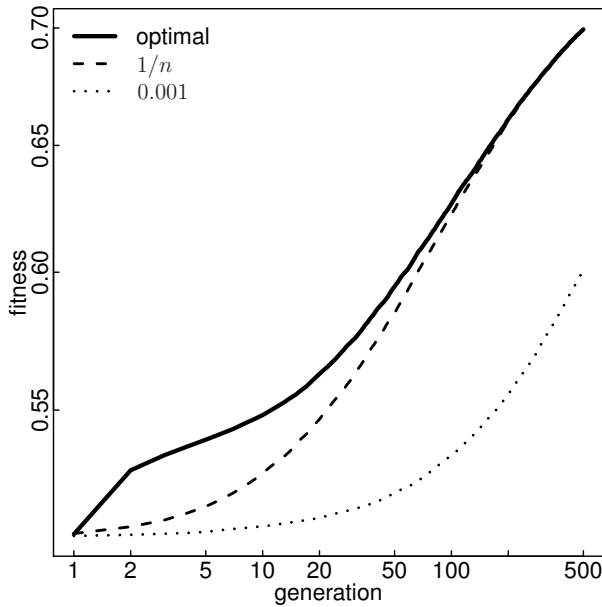


Figure 4: Log-log plot of mean fitness of (1+1)-EA on 500 trials of 500 generations each on an unrestricted NK-landscape model with $n = 100$, $K = 3$.

form where each disjunctive clause is of length at most k . The objective is to find a Boolean assignment to the n variables that maximizes the number of clauses satisfied. Since the set of all Boolean assignments to n variables is isomorphic to $\{0, 1\}^n$, the fitness function can be expressed as a pseudo-Boolean function that counts how many clauses in the formula are satisfied under a corresponding assignment. Supposing the formula has a family of m clauses, the fitness

function can be written as

$$f(x) = \sum_{j=1}^m g_j(x)$$

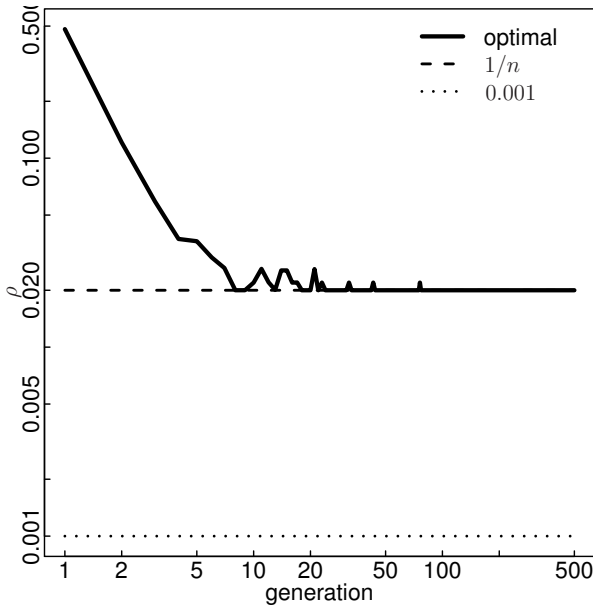
where g_j is a binary indicator function that evaluates to 1 if and only if clause j is satisfied under the assignment corresponding to x . Since g_j only depends on the state of the variables belonging to clause j , and this quantity is bounded by k , the MAX- k -SAT fitness function is epistatically bounded by k .

We performed 500 trials of 500 generations each of the (1+1)-EA on two randomly generated Boolean formulas with 50 variables and 218 clauses, and 100 variables and 430 clauses. The algorithm employed the three different mutation rates discussed above. Again, when the optimal rate reaches degeneracy, we revert to the rate that maximizes expected fitness while enforcing bits to be flipped in expectation.

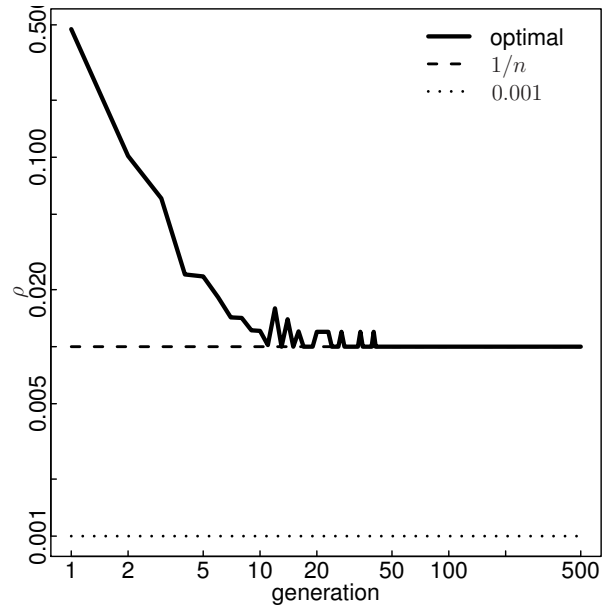
In Figure 6 we report the ρ values found during search. Again, due to the simplicity of the $\mathbb{M}_x(\rho)$ polynomial, the solution time is negligible to compute ρ^* in each case. In both instances, we see again the decrease of the expectation-optimal rate quickly to the degenerate rate where it then reverts to the standard $1/n$ mutation rate around generation 20 to 50. The initially higher expectation-optimal rate shown in Figure 6 translates to early gains in search as reported in Figure 7 when compared to standard and hard-wired mutation rates.

6. CONCLUSION

We have shown that when the epistasis of the fitness function is bounded by a constant k , it is possible to efficiently compute the expected fitness of a mutation in the (1+1)-EA for any given mutation rate. Moreover, it is also possible to efficiently compute for any point the mutation rate that re-

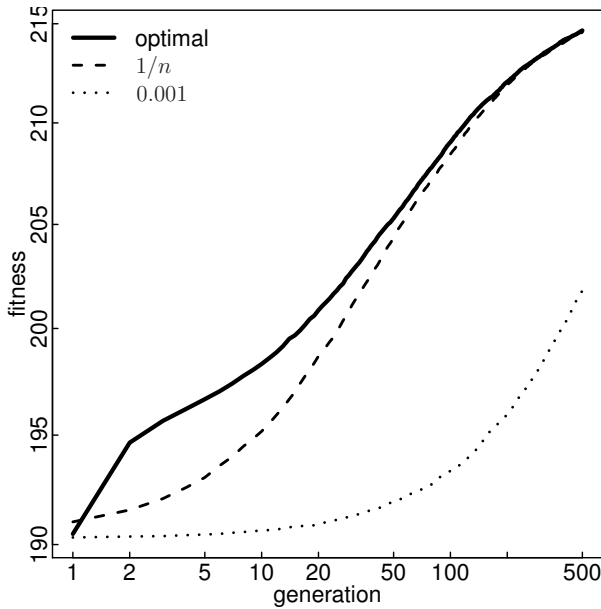


(a) 50 variables and 218 clauses.

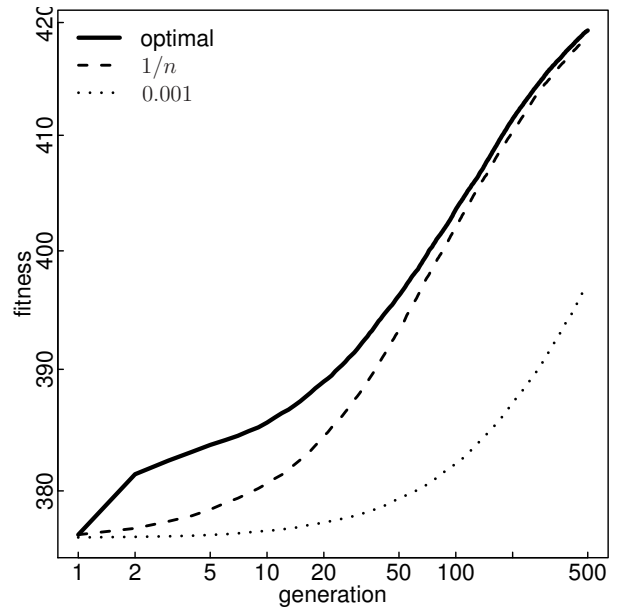


(b) 100 variables and 430 clauses.

Figure 6: Log-log plot of mean mutation rates for the (1+1)-EA on 500 trials of 500 generations each on MAX-3-SAT instances.



(a) 50 variables and 218 clauses.



(b) 100 variables and 430 clauses.

Figure 7: Log-log plot of mean fitness values for the (1+1)-EA on 500 trials of 500 generations each on MAX-3-SAT instances.

sults in the highest possible expected fitness (supposing that the mutation rate must be constant across the string).

We have also proved that, for strings with fitness higher than the expectation in Hamming spheres up to radius k , the frequently recommended rate of $1/n$ yields the maximal expected fitness of offspring while imposing the constraint that some bits are flipped in expectation.

Acknowledgements

This research was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number FA9550-08-1-0422. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

7. REFERENCES

- [1] Thomas Bäck. The interaction of mutation rate, selection, and self-adaptation within a genetic algorithm. In Reinhard Männer and Bernard Manderick, editors, *Parallel Problem Solving from Nature 2*, pages 85–94, Amsterdam, 1992. Elsevier.
- [2] Thomas Bäck. Self-adaptation in genetic algorithms. In Francisco J. Varela and Paul Bourgnine, editors, *Proceedings of the First European Conference on Artificial Life*, pages 263–271, Cambridge, MA, 1992. The MIT Press.
- [3] Thomas Bäck. Optimal mutation rates in genetic search. In Stephanie Forrest, editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pages 2–8, San Mateo, CA, 1993. Morgan Kaufmann.
- [4] Thomas Bäck and Martin Schütz. Intelligent mutation rate control in canonical genetic algorithms. In Zbigniew Ras and Maciek Michalewicz, editors, *Foundations of Intelligent Systems*, volume 1079 of *Lecture Notes in Computer Science*, pages 158–167. Springer Berlin / Heidelberg, 1996.
- [5] Süntje Böttcher, Benjamin Doerr, and Frank Neumann. Optimal fixed and adaptive mutation rates for the LeadingOnes problem. In Robert Schaefer, Carlos Cotta, Joanna Kolodziej, and Günter Rudolph, editors, *Parallel Problem Solving from Nature – PPSN XI*, volume 6238 of *Lecture Notes in Computer Science*, pages 1–10. Springer Berlin / Heidelberg, 2010.
- [6] Sung-Soon Choi, Kyomin Jung, and Jeong Han Kim. Almost tight upper bound for finding Fourier coefficients of bounded pseudo-Boolean functions. In Rocco A. Servedio and Tong Zhang, editors, *Proceedings of the Twenty-first Conference on Learning Theory (COLT 2008)*, pages 123–134, Helsinki, Finland, July 2008. Omnipress.
- [7] Kenneth A. De Jong. *An analysis of the behavior of a class of genetic adaptive systems*. PhD thesis, University of Michigan, 1975.
- [8] Benjamin Doerr, Thomas Jansen, Dirk Sudholt, Carola Winzen, and Christine Zarges. Optimizing monotone functions can be difficult. In Robert Schaefer, Carlos Cotta, Joanna Kolodziej, and Günter Rudolph, editors, *Parallel Problem Solving from Nature – PPSN XI*, volume 6238 of *Lecture Notes in Computer Science*, pages 42–51. Springer Berlin / Heidelberg, 2010.
- [9] Stefan Droste, Thomas Jansen, and Ingo Wegener. A rigorous complexity analysis of the $(1+1)$ evolutionary algorithm for separable functions with Boolean inputs. *Evolutionary Computation*, 6(2):185–196, 1998.
- [10] Josselin Garnier, Leila Kallel, and Marc Schoenauer. Rigorous hitting times for binary mutations. *Evolutionary Computation*, 7(2):167–203, 1999.
- [11] John J. Grefenstette. Optimization of control parameters for genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, 16(1):122–128, 1986.
- [12] Robert B. Heckendorn. *Walsh Analysis, Epistasis, and Optimization Problem Difficulty for Evolutionary Algorithms*. PhD thesis, Colorado State University, Fort Collins, CO, July 1999.
- [13] Robert B. Heckendorn, Soraya B. Rana, and L. Darrell Whitley. Test function generators as embedded landscapes. In Wolfgang Banzhaf and Colin R. Reeves, editors, *Foundations of Genetic Algorithms V*, pages 183–198. Morgan Kaufmann, 1998.
- [14] Robert B. Heckendorn and Alden H. Wright. Efficient linkage discovery by limited probing. *Evolutionary Computation*, 12:517–545, 2004.
- [15] Jürgen Hesser and Reinhard Männer. Towards an optimal mutation probability for genetic algorithms. In Hans-Paul Schwefel and Reinhard Männer, editors, *Parallel Problem Solving from Nature*, volume 496 of *Lecture Notes in Computer Science*, pages 23–32. Springer Berlin / Heidelberg, 1991.
- [16] John H. Holland. *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, 1975.
- [17] Thomas Jansen and Ingo Wegener. On the choice of the mutation probability for the $(1+1)$ EA. In Marc Schoenauer, Kalyanmoy Deb, Günther Rudolph, Xin Yao, Evelyne Lutton, Juan Merelo, and Hans-Paul Schwefel, editors, *Parallel Problem Solving from Nature VI*, volume 1917 of *Lecture Notes in Computer Science*, pages 89–98. Springer Berlin / Heidelberg, 2000.
- [18] Hillol Kargupta and Byung-hoon Park. Gene expression and fast construction of distributed evolutionary representation. *Evolutionary Computation*, 9(1):43–69, 2001.
- [19] Stuart Kauffman and Simon Levin. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128:11–45, 1987.
- [20] Mikhail Kravchuk. Sur une généralisation des polynômes d’Hermite. *Comptes rendus de l’Académie des sciences*, 189(17):620–622, 1929.
- [21] Heinz Mühlenbein. How genetic algorithms really work: I. mutation and hillclimbing. In Reinhard Männer and Bernard Manderick, editors, *Parallel Problem Solving from Nature 2*, pages 15–25. Elsevier, Amsterdam, 1992.
- [22] Günter Rudolph. *Convergence Properties of Evolutionary Algorithms*. Verlag Dr Kovač, 1997.
- [23] J. David Schaffer, Richard A. Caruana, Larry J. Eshelman, and Rajarshi Das. A study of control parameters affecting online performance of genetic algorithms for function optimization. In *Proceedings of the Third International Conference on Genetic Algorithms*, pages 51–60, San Francisco, CA, USA, 1989. Morgan Kaufmann Publishers Inc.
- [24] Andrew M. Sutton, L. Darrell Whitley, and Adele E. Howe. Computing the moments of k -bounded pseudo-Boolean functions over Hamming spheres of arbitrary radius in polynomial time. *Theoretical Computer Science*, 2011, doi:10.1016/j.tcs.2011.02.006 (In press).
- [25] Alden H. Wright, Richard K. Thompson, and Jian Zhang. The computational complexity of N - K fitness functions. *Evolutionary Computation*, 4(4):373–379, 2000.