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

Andrew M Sutton, Darrell Whitley, Adele Howe

Introduction

Want to maximize $f : \{0,1\}^n \to \mathbb{R}^+$

(1+1)-EA(ρ)

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

Question: how do we choose the mutation rate ρ?

Motivation

Early experimental evidence suggested $0.001 \leq \rho \leq 0.01$ – De Jong (1975), Grefenstette (1986), Schaffer (1989)

Droste et al. (1998): linear functions

$$\rho = 1/n \implies O(n \log n) \text{ expected convergence}$$

Jansen and Wegener (2000): PathToJump

$$\rho = 1/n \implies \text{superpolynomial runtime w.h.p.}$$
$$\rho = \frac{\log n}{n} \implies \text{polytime convergence}$$

Doerr et al. (2010): monotone functions

changing ρ by a constant factor $\implies$ exponential performance gap

Expected offspring fitness under rate ρ

We will study the class of functions $f : \{0, 1\}^n \to \mathbb{R}^+$ whose *epistasis* is bounded by a constant k.

What is the "best" mutation rate?

- Maximizes *probability of improvement* (difficult to know in general)
- Maximizes the *expected fitness of the offspring*

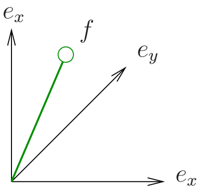
Let $x \in \{0, 1\}^n$. We define

$\mathbb{M}_x(\rho)$ - the expected fitness of the offspring of x under rate ρ.

Expected offspring fitness under rate ρ

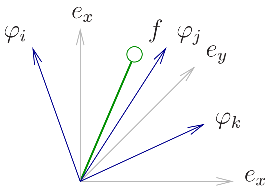
How does one compute $\mathbb{M}_x(\rho)$?
Appeal to a basis function decomposition of the fitness function

Standard and alternative bases



$e_y(x) = \delta_{xy}$

$$f(x) = \sum_y v_y e_y(x)$$



$$f(x) = \sum_i a_i \varphi_i(x)$$

Decomposition provides information about the relationship between the *fitness function* and the *mutation operator*.

Appealing to Walsh decomposition

We can write the fitness function as a Walsh polynomial

$$f(x) = \sum_i w_i \psi_i(x)$$

Expected fitness of a point drawn uniformly at random at Hamming distance r from x (Sutton et al. 2011)

$$\mathbb{S}_x^r = \sum_i \gamma_{i,r} w_i \psi_i(x) \qquad (*)$$

From this we can obtain the expected fitness under rate ρ

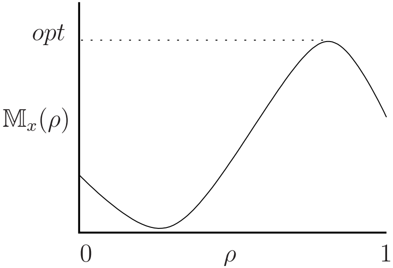
$$\mathbb{M}_x(\rho) = \sum_{r=0}^{n} \binom{n}{r} \rho^r (1-\rho)^{n-r} \mathbb{S}_x^r$$

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

where each A_m is a linear combination of terms from $(*)$.

The expected offspring fitness polynomial

For any pseudo-Boolean function, $\mathbb{M}_x(\rho)$ is a degree at most n polynomial in ρ.



Maximum in the inverval $[0,1]$ gives the best ρ in terms of maximizing expected fitness.

We are interested in exploring some properties of this polynomial.

Degeneracy

Let $\rho^\star \in \arg\max_{\rho \in [0,1]} \mathbb{M}_x(\rho)$.

$\rho^\star \neq 0$: there exists a mutation rate that produces an expected improvement over $f(x)$.

$\rho^\star = 0$: no mutation rate can produce an expected improvement over $f(x)$.

Suppose we insist on flipping $\ell > 0$ bits in expectation. Then $\rho = \ell/n$.

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

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

We can conclude

- $\rho = 1/n$ minimizes expected loss in fitness.
- In this case, expected fitness of offspring is bounded below by $\frac{f(x)}{e}$.

Linear functions

- Expressed as a sum of linear terms (epistasis is $k = 1$)
- Walsh coefficients of order higher than 1 vanish

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

where A_m is a linear combination of terms from Walsh series expansion of f.

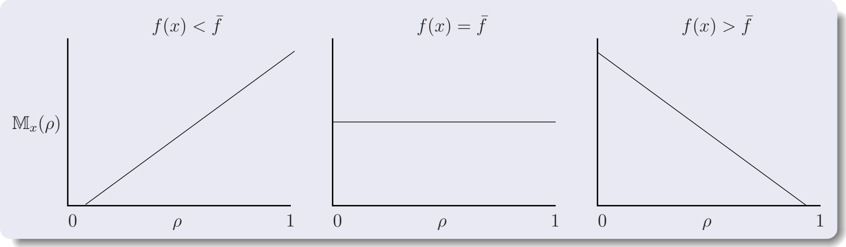
Proposition

If f is a linear function, then

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

Linear functions

When f is linear, $\mathbb{M}_x(\rho) = f(x) + 2\left(\bar{f} - f(x)\right)\rho$



Degenerate when $f(x) > \bar{f}$, in this case $1/n$ maximizes expected offspring s.t. $\ell > 0$ bits flipping in expectation

This illustrates a problem with using expectation: consider when $f(x) = \bar{f}$.

Epistatically bounded functions

$f : \{0,1\}^n \to \mathbb{R}^+$ where the epistasis is bounded by $k = O(1)$.

Max-k-Sat

Boolean formula over a set V of n variables and m clauses consisting of exactly k literals

$$\bigwedge_{i=1}^{m} (\ell_{i,1} \vee \ell_{i,2} \vee \cdots \vee \ell_{i,k}), \text{ where } \ell_{i,j} \in \{v, \neg v : v \in V\}$$

$f : \{0,1\}^n \to \{0,\dots,m\}$ counts clauses satisfied under x.

NK-landscapes

$$f(x) = \frac{1}{n}\sum_{j=1}^{n} g_j\left(x[j], x[b_1^{(j)}], \dots, x[b_K^{(j)}]\right), \text{ where } g_j : \{0,1\}^{K+1} \to [0,1]$$

Epistatically bounded functions

Proposition

When f is epistatically bounded by k, $A_m = 0$ for $m > k$.

Thus for any k-bounded pseudo-Boolean function, $\mathbb{M}_x(\rho)$ is a degree-k polynomial in ρ.

E.g., for $\textsc{Max-2-Sat}$, the mutation polynomial is *quadratic*.

Furthermore, A_m depends only on $\mathbb{S}_x^r$ for $r \leq m$.

Corollary

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

Numerical results

Practically speaking, does solving $\mathbb{M}_x(\rho)$ for the best mutation rate provide any insight?

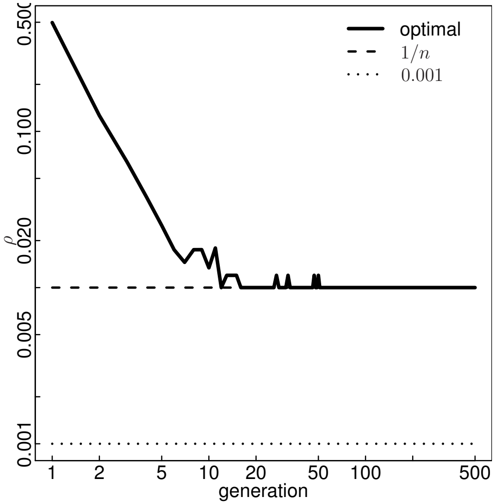
$$\frac{d}{d\rho}\mathbb{M}_x(\rho) = A_1 + 2A_2\rho + 3A_3\rho^2 + \cdots + nA_n\rho^{n-1},$$

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

Numerically finding $\rho^\star$

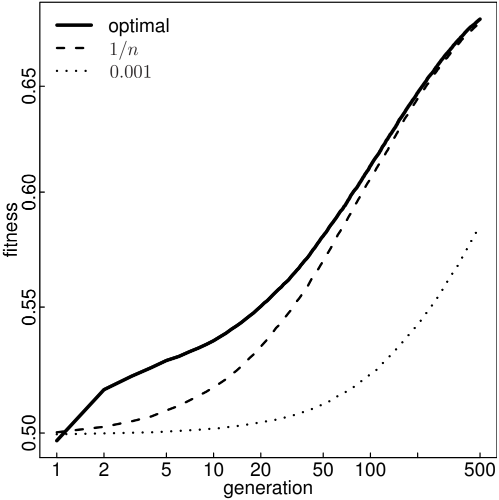
- Find the stationary points of $\mathbb{M}_x(\rho)$ by numerically solving for the real roots of $\frac{d}{d\rho}\mathbb{M}_x(\rho)$.
- Second derivative test for concavity
- $\rho^\star$ is maximum of this set union $\{0, 1\}$

Unrestricted NK-landscape, $n = 100$, $k = 2$



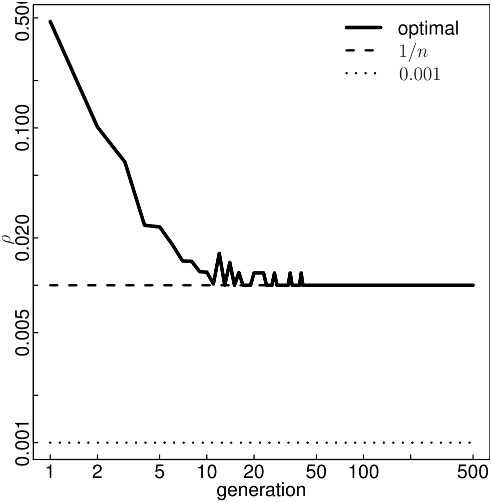
Numerical results

Unrestricted NK-landscape, $n = 100$, $k = 2$



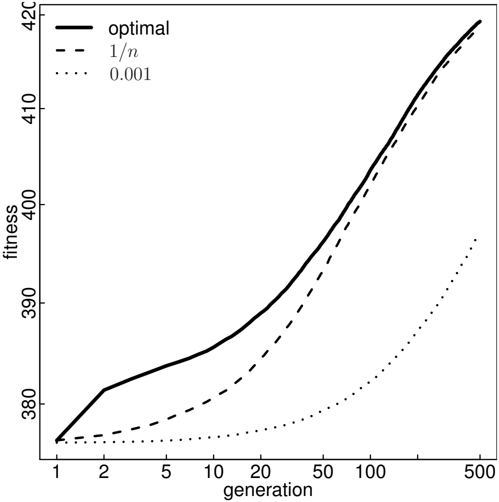
Numerical results

MAX-3-SAT, $n = 100$



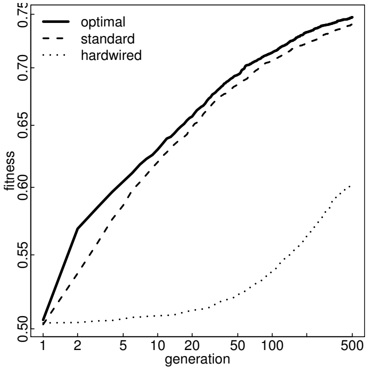
Numerical results

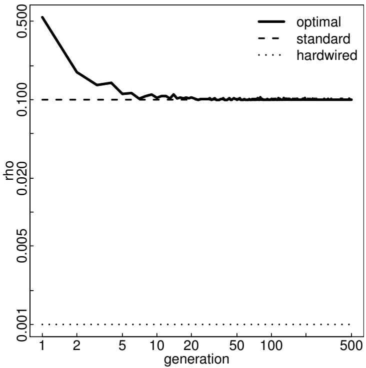
MAX-3-SAT, $n = 100$



Numerical results (high epistasis)

NK-landscape, N=10, K=9





Research directions

Analyzing w_i for certain specific problems (e.g., $\textsc{Max}$-k-$\textsc{Sat}$). Provide more precise statements about the expected fitness in general.

Working with higher moments of the random variable distribution

Connection to runtime analysis... might be possible if we can discover bounds for the higher moments of the distribution

Generalization to broader problem classes (Fourier decomposition)

Conclusion

As long as epistasis is bounded by k,

- it is possible to efficiently compute the expected fitness of a mutation for any rate.
- it is possible to efficiently find the rate that results in the highest possible expected fitness.

For strings with fitness higher than expectation in spheres up to radius k, $1/n$ yields maximal expected fitness of the offspring while imposing the constraint that $\ell > 0$ bits are flipped in expectation.