Explaining Adaptation in Genetic Algorithms with Uniform Crossover

Keki Burjorjee
keki.burjorjee@gmail.com

Uniform Crossover

- Uniform Crossover first used by Ackley (1987)
- Studied and popularized by Syswerda (1989) and Spears & De Jong (1990, 1991)
- Numerous accounts of optimization in GAs with UX
- In practice frequently outperforms XO with tight linkage (Fogel, 2006)

Optimization in GAs with UX Unexplained

- Cannot be explained within the rubric of the BBH
- No viable alternative has been proposed

Optimization in GAs with UX Unexplained

Hyperclimbing Hypothesis

A scientific explanation for optimization in GAs with UX

What Does "Scientific" Mean?

- Logical Positivism (Proof or Bust)
  - Scientific truth is absolute
  - Emphasis on Verifiability (i.e. mathematical proof)
- The Popperian method
  - Scientific truth is provisional
  - Emphasis on Falsifiability (testable predictions)

Popperian Method

- Logic behind the Popperian Method : Contrapositive
  - $Theory \Rightarrow Phenomenon \quad \Leftrightarrow \quad \neg Phenomenon \Rightarrow \neg Theory$
- Additional tightening by Popper
  - Unexpected Phenomenon → More credence owed to the theory
  - e.g. Gravitational lensing → General theory of relativity

Popperian Method

- Logic behind the Popperian Method : Contrapositive
  - $Theory \Rightarrow Phenomenon \quad \Leftrightarrow \quad \neg Phenomenon \Rightarrow \neg Theory$
- Additional tightening by Popper
  - Unexpected Phenomenon → More credence owed to the theory
  - e.g. Gravitational lensing → General theory of relativity

Additional Requirements in a Science of EC

- Weak assumptions about distribution of fitness
  - This is just Occam's Razor
- Upfront proof of concept
  - Avoid another "Royal Roads moment"
- (Nice to have) Identification of a core computational efficiency

Hyperclimbing Hypothesis

GAs with UX perform optimization by efficiently implementing a global search heuristic called hyperclimbing

Outline

1. Highlight Symmetry of UX
2. Describe Hyperclimbing Heuristic
3. Provide proof-of-concept
   - Show a GA implementing hyperclimbing efficiently
4. Make a prediction and validate it on
   - MAX-3SAT
   - Sherrington Kirkpatrick Spin Glasses problem
5. Outline future work

Variation in GAs

0	1	0	0	1	0	1	0	1	....	0
1	0	0	1	1	0	0	1	1	....	1

Variation in GAs

0	1	0	0	1	0	1	0	1	....	0
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	....	X_ℓ
1	0	0	1	1	0	0	1	1	....	1

Variation in GAs

$$
\begin{array}{cccccccccc}
0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & \ldots\ldots\ldots & 0 \\
\uparrow & \downarrow & \downarrow & \uparrow & \uparrow & \downarrow & \downarrow & \downarrow & \uparrow & \ldots\ldots\ldots & \uparrow \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & \ldots\ldots\ldots & 1 \\
\hline
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & \ldots\ldots\ldots & 0
\end{array}
$$

Variation in GAs

$$
\begin{array}{cccccccccccc}
0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & \ldots\ldots\ldots & 0 \\
\uparrow & \downarrow & \downarrow & \uparrow & \uparrow & \downarrow & \downarrow & \downarrow & \uparrow & \ldots\ldots\ldots & \uparrow \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & \ldots\ldots\ldots & 1 \\
\hline
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & \ldots\ldots\ldots & 0 \\
Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 & Y_7 & Y_8 & Y_9 & \ldots\ldots\ldots & Y_\ell
\end{array}
$$

Variation in GAs

0	1	0	0	1	0	1	0	1	....	0
↑	↓	↓	↑	↑	↓	↓	↓	↑	....	↑
1	0	0	1	1	0	0	1	1	....	1
0	0	0	0	1	0	0	1	1	....	0
⇓	⇓	⇓	✠	⇓	⇓	⇓	✠	⇓	....	⇓
0	0	0	1	1	0	0	0	1	....	0

Variation in GAs

$$X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5 \quad X_6 \quad X_7 \quad X_8 \quad X_9 \quad \ldots\ldots\ldots \quad X_\ell$$

$$Y_1 \quad Y_2 \quad Y_3 \quad Y_4 \quad Y_5 \quad Y_6 \quad Y_7 \quad Y_8 \quad Y_9 \quad \ldots\ldots\ldots \quad Y_\ell$$

UX: $X_1, \ldots, X_\ell$ are independent

- Absence of *positional bias* (Eshelman et al., 1989)
  - Attributes can be arbitrarily permuted
- Suppose $Y_1, \ldots, Y_\ell$ independent and independent of ℓ
  - If locus i immaterial to fitness, can be spliced out

The Hyperclimbing Heuristic

Hyperclimbing: A Stochastic Search Heuristic

The Setting

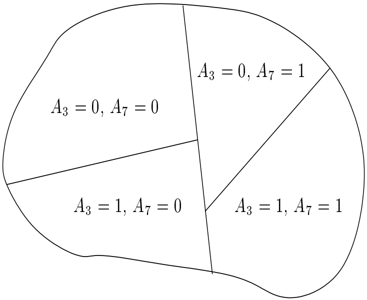
- $\{0,1\}^\ell \xrightarrow{f} \mathbb{R}$
- f may be stochastic

Schema Partitions and Schemata

Given index set $\mathcal{I} \subset \{1, \ldots, \ell\}$, s.t. $|\mathcal{I}| = k$

- $\mathcal{I}$ partitions the search space into 2^k schemata
- Schema partition denoted by $[\![\mathcal{I}]\!]$

E.g. for $\mathcal{I} = \{3, 7\}$



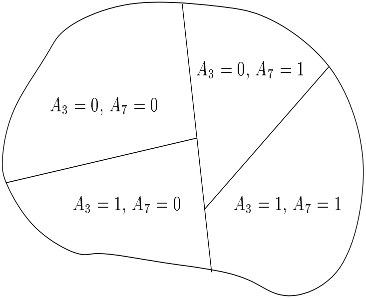
- Coarseness of the partition determined by k (the order)
- $\binom{\ell}{k}$ schema partitions of order k
- For fixed k, $\binom{\ell}{k} \in \Omega(\ell^k)$

Schema Partitions and Schemata

Given index set $\mathcal{I} \subset \{1, \ldots, \ell\}$, s.t. $|\mathcal{I}| = k$

- $\mathcal{I}$ partitions the search space into 2^k schemata
- Schema partition denoted by $[\![\mathcal{I}]\!]$

E.g. for $\mathcal{I} = \{3, 7\}$



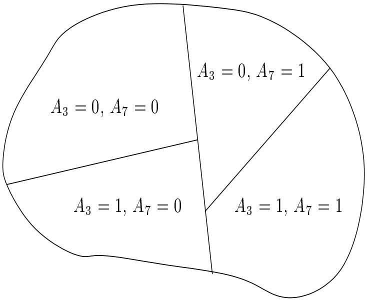
- Coarseness of the partition determined by k (the order)
- $\binom{\ell}{k}$ schema partitions of order k
- For fixed k, $\binom{\ell}{k} \in \Omega(\ell^k)$

Schema Partitions and Schemata

Given index set $\mathcal{I} \subset \{1, \ldots, \ell\}$, s.t. $|\mathcal{I}| = k$

- $\mathcal{I}$ partitions the search space into 2^k schemata
- Schema partition denoted by $[\![\mathcal{I}]\!]$

E.g. for $\mathcal{I} = \{3, 7\}$



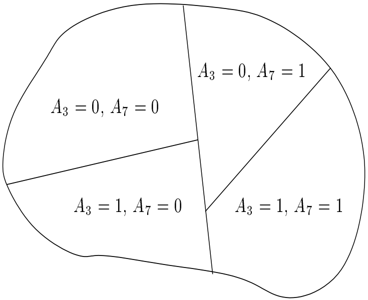
- Coarseness of the partition determined by k (the order)
- $\binom{\ell}{k}$ schema partitions of order k
- For fixed k, $\binom{\ell}{k} \in \Omega(\ell^k)$

Schema Partitions and Schemata

Given index set $\mathcal{I} \subset \{1, \ldots, \ell\}$, s.t. $|\mathcal{I}| = k$

- $\mathcal{I}$ partitions the search space into 2^k schemata
- Schema partition denoted by $[\![\mathcal{I}]\!]$

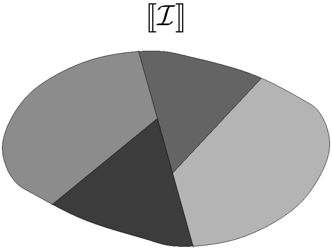
E.g. for $\mathcal{I} = \{3, 7\}$



- Coarseness of the partition determined by k (the order)
- $\binom{\ell}{k}$ schema partitions of order k
- For fixed k, $\binom{\ell}{k} \in \Omega(\ell^k)$

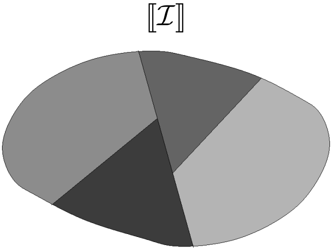
The Effect of a Schema Partition

- Effect of $[\![\mathcal{I}]\!]$: Variance of average fitness of schemata in $[\![\mathcal{I}]\!]$
- If $\mathcal{I} \subset \mathcal{I}'$, then Effect($\mathcal{I}$) $\leq$ Effect($\mathcal{I}'$)



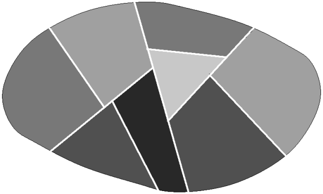
The Effect of a Schema Partition

- Effect of $[\![\mathcal{I}]\!]$: Variance of average fitness of schemata in $[\![\mathcal{I}]\!]$
- If $\mathcal{I} \subset \mathcal{I}'$, then Effect$(\mathcal{I}) \leq$ Effect$(\mathcal{I}')$



How a Hyperclimbing Heuristic Works

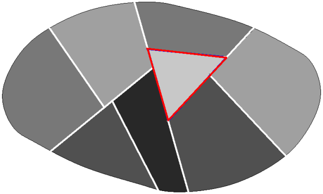
- Finds a coarse schema partition with a detectable effect
- Limits future sampling to a schema with above average fitness



- Raises expected fitness of all future samples

How a Hyperclimbing Heuristic Works

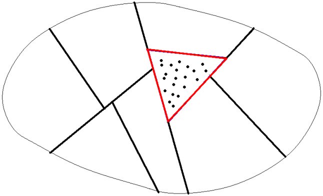
- Finds a coarse schema partition with a detectable effect
- Limits future sampling to a schema with above average fitness



- Raises expected fitness of all future samples

How a Hyperclimbing Heuristic Works

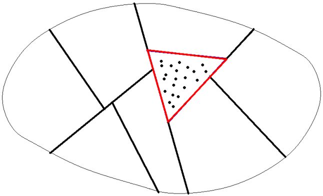
- Finds a coarse schema partition with a detectable effect
- Limits future sampling to a schema with above average fitness



- Raises expected fitness of all future samples

How a Hyperclimbing Heuristic Works

- Finds a coarse schema partition with a detectable effect
- Limits future sampling to a schema with above average fitness



- Raises expected fitness of all future samples

How a Hyperclimbing Heuristic Works

Limit future sampling to some schema

$\Updownarrow$ equivalent to $\Updownarrow$

Fix defining loci of the schema in the population

- Hyperclimbing Heuristic now recurses
  - Search occurs over the unfixed loci
- And so on . . .

Fitness Distribution Assumption

- $\exists$ a small set of unfixed loci, $\mathcal{I}$, such that
  - Conditional effect of $[\![\mathcal{I}]\!]$ is detectable
    - Conditional upon loci that have already been fixed
  - Unconditional effect of $[\![\mathcal{I}]\!]$ may be undetectable
- E.g.
  - Effect of 1*#*0*#* is detectable
  - Effect of **#****#* undetectable
  - ('*' = wildcard, '#'= defined locus)
- Called staggered coarse conditional effects

Fitness Distribution Assumption is Weak

- Staggered coarse conditional effects assumption is weak
  - Weaker than fitness distribution assumption in the BBH
    - BBH assumes unstaggered coarse unconditional effects
- Weaker assumptions are more likely to be satisfied in practice
  - Good b/c we aspire to explain optimization in all GAs with UX

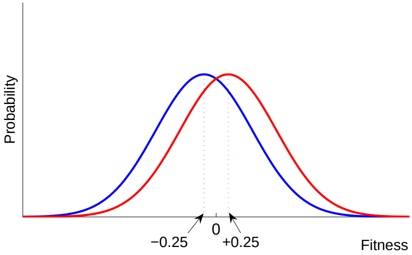
Hyperclimbing Heuristic is in Good Company

- Hyperclimbing an example of a global decimation heuristic
- Global decimation heuristics iteratively reduce the size of a search space
  - Use non-local information to find and fix partial solutions
  - No backtracking
- E.g. Survey propagation (Mérzad et al., 2002)
  - State-of-the-art solver for large, random SAT problems

Proof Of Concept

4-Bit Stochastic Learning Parities Problem

$$fitness(x) \sim \begin{cases} \mathcal{N}(+0.25, 1) \text{ if } x_{i_1} \oplus x_{i_2} \oplus x_{i_3} \oplus x_{i_4} = 1 \\ \mathcal{N}(-0.25, 1) \text{ otherwise} \end{cases}$$



4-Bit Stochastic Learning Parities Problem

The Game

10111001011101001111000001	$\mapsto$	$+0.5689$
10001010110110100011110000	$\mapsto$	-0.25565
11101100100010111101101110	$\mapsto$	-0.37747
01110101001111100000110001	$\mapsto$	-0.29589
00000010001000100110011101	$\mapsto$	-1.4751
00001100000010001000110000 0	$\mapsto$	-0.234
01000001111000111100110100	$\mapsto$	$+0.11844$
11010001000101000011000110	$\mapsto$	$+0.31481$
00000011100010001100000111	$\mapsto$	$+1.4435$
00100111000111000001110000	$\mapsto$	-0.35097
10100011011010101010100001	$\mapsto$	$+0.62323$
00011011101010100010100000	$\mapsto$	$+0.79905$
00101010101100101110100000	$\mapsto$	$+0.94089$
01010110000000110110110011	$\mapsto$	-0.99209
11100101000110010110110101	$\mapsto$	$+0.21204$
00011111011101110101000111	$\mapsto$	$+0.23788$
00001111111101011110111010	$\mapsto$	-1.0078
$\vdots$ $\vdots$	$\vdots$	$\vdots$
00001011101111000000111100	$\mapsto$	$+1.0823$
11011100111100010100111101	$\mapsto$	-0.1315

4-Bit Stochastic Learning Parities Problem

The Game

↓ ↓ ↓ ↓		
10111001011101001111000001	$\mapsto$	$+0.5689$
10001010110110100011110000	$\mapsto$	-0.25565
11101100100010111101101110	$\mapsto$	-0.37747
01110101001111100000110001	$\mapsto$	-0.29589
00000010001000100110011101	$\mapsto$	-1.4751
00001100000100010001100000	$\mapsto$	-0.234
01000001111000111100110100	$\mapsto$	$+0.11844$
11010001000101000011000110	$\mapsto$	$+0.31481$
00000011100010001100000111	$\mapsto$	$+1.4435$
00100111000111000001110000	$\mapsto$	-0.35097
10100011011010101010100001	$\mapsto$	$+0.62323$
00011011101010100010100000	$\mapsto$	$+0.79905$
00101010101100101110100000	$\mapsto$	$+0.94089$
01010110000000110110110011	$\mapsto$	-0.99209
11100101000110010110110101	$\mapsto$	$+0.21204$
00011111011101110101000111	$\mapsto$	$+0.23788$
00001111111101011110111010	$\mapsto$	-1.0078
⋮ ⋮	⋮	⋮
00001011101111000000111100	$\mapsto$	$+1.0823$
11011100111100010100111101	$\mapsto$	-0.1315
↑ ↑ ↑ ↑		

Effective Loci

Solving the 4-Bit Stochastic Learning Parities Problem

Naive approach

- Visit loci in groups of four
- Check for differentiation in the multivariate marginal fitness values
- Time complexity is $\Omega(\ell^4)$

x_{i_1}	x_{i_2}	x_{i_3}	x_{i_4}	Expected Marginal Fitness
0	0	0	0	−0.25
0	0	0	1	+0.25
0	0	1	0	+0.25
0	0	1	1	−0.25
0	1	0	0	+0.25
0	1	0	1	−0.25
0	1	1	0	−0.25
0	1	1	1	+0.25
1	0	0	0	+0.25
1	0	0	1	−0.25
1	0	1	0	−0.25
1	0	1	1	+0.25
1	1	0	0	−0.25
1	1	0	1	+0.25
1	1	1	0	+0.25
1	1	1	1	−0.25

Solving the 4-Bit Stochastic Learning Parities Problem

x_j	Expected Marginal Fitness
0	0.0
1	0.0

x_{j_1}	x_{j_2}	Expected Marginal Fitness
0	0	0.0
0	1	0.0
1	0	0.0
1	1	0.0

- Visiting loci in groups of three or less won't work

x_{j_1}	x_{j_2}	x_{j_3}	Expected Marginal Fitness
0	0	0	0.0
0	0	1	0.0
0	1	0	0.0
0	1	1	0.0
1	0	0	0.0
1	0	1	0.0
1	1	0	0.0
1	1	1	0.0

▷
Watch On YouTube

Dynamics of a SGA on the 4-Bit Learning Parities problem
#Loci=200

Symmetry Analytic Conclusion

Expected #generations for the red dots to diverge constant w.r.t.

- Location of the red dots
- Number of blue dots

Dynamics of a blue dot invariant to

- Location of the red dots
- Number of blue dots

▷

Watch On YouTube

Dynamics of a SGA on the 4-Bit Learning Parities problem
#Loci=1000

▷

Watch On YouTube

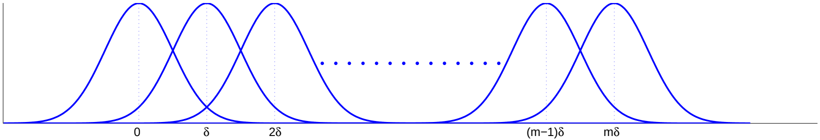
Dynamics of a SGA on the 4-Bit Learning Parities problem
#Loci=10000

Staircase Functions

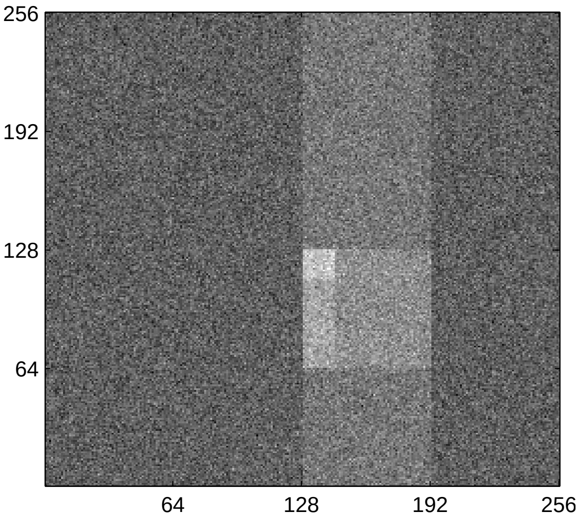
- Staircase function descriptor: $(m, n, \delta, \sigma, \ell, L, V)$

- $L = \underbrace{\begin{bmatrix} 3 & 87 & \cdots & 93 \\ 42 & 58 & \cdots & 72 \\ \vdots & \vdots & \ddots & \vdots \\ 67 & 73 & \cdots & 81 \end{bmatrix}}_{n} \Bigg\} m, \quad V = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 1 \end{bmatrix}}_{n} \Bigg\} m$

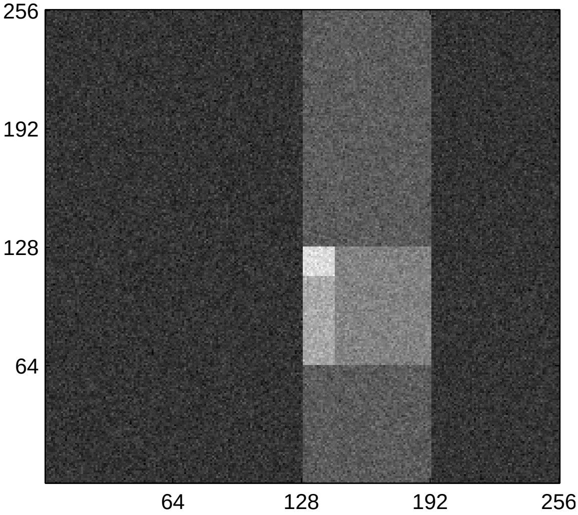
- Staircase function is a stochastic function $f : \{0,1\}^\ell \to \mathbb{R}$



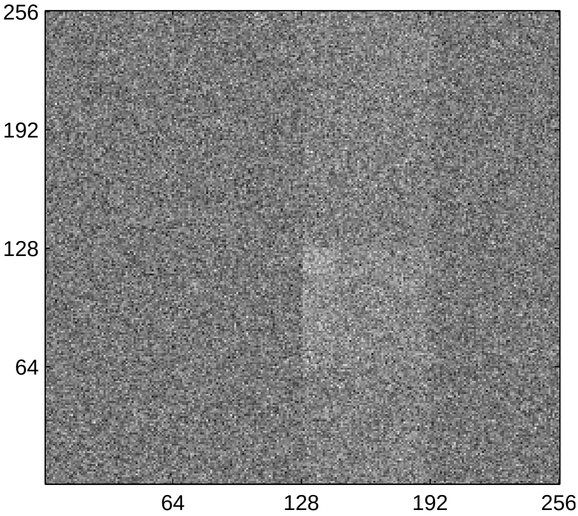
Staircase Function



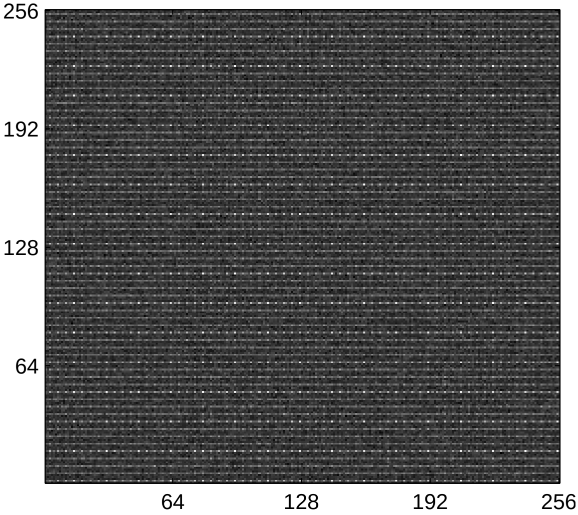
Staircase Function



Staircase Function

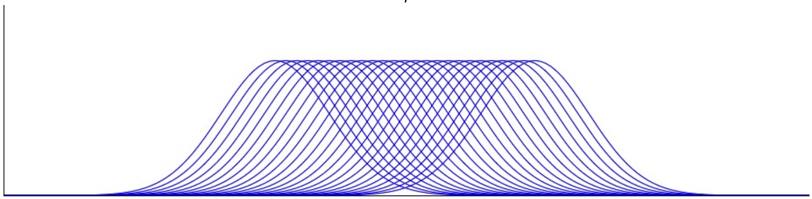


Staircase Function



Staircase Functions

$\delta = 0.2,\ \sigma = 1$



$\ell = 50$

$m = 5$

$n = 4$

$\delta = 0.2$

$\sigma = 1$

$$L = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \\ 17 & 18 & 19 & 20 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

▷
Watch On YouTube

$\ell = 50$

$m = 5$

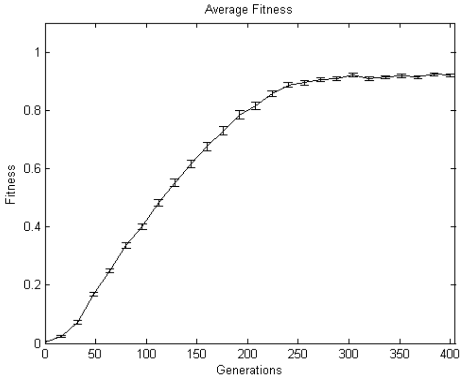
$n = 4$

$\delta = 0.2$

$\sigma = 1$

$$L = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \\ 17 & 18 & 19 & 20 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$



#trials=100. Error bars show standard error.

$\ell = 50$

$m = 5$

$n = 4$

$\delta = 0.2$

$\sigma = 1$

$$L = \begin{bmatrix} 35 & 25 & 37 & 46 \\ 31 & 5 & 32 & 20 \\ 21 & 33 & 50 & 42 \\ 40 & 27 & 30 & 3 \\ 12 & 14 & 48 & 2 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

▷
Watch On YouTube

$\ell = 5000$

$m = 5$

$n = 4$

$\delta = 0.2$

$\sigma = 1$

$$L = \begin{bmatrix} 2050 & 4659 & 1931 & 3284 \\ 2130 & 2404 & 205 & 4418 \\ 143 & 1284 & 53 & 437 \\ 2061 & 904 & 2650 & 3080 \\ 3503 & 724 & 4822 & 4444 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

▷
Watch On YouTube

$\ell = 5000$

$m = 5$

$n = 4$

$\delta = 0.2$

$\sigma = 1$

$$L = \begin{bmatrix} 2050 & 4659 & 1931 & 3284 \\ 2130 & 2404 & 205 & 4418 \\ 143 & 1284 & 53 & 437 \\ 2061 & 904 & 2650 & 3080 \\ 3503 & 724 & 4822 & 4444 \end{bmatrix}$$

$$V = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

▷
Watch On YouTube

Prediction & Validation

$\ell = 500$

$m = 125$

$n = 4$

$\delta = 0.2$

$\sigma = 1$

$L =$

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ \vdots & \vdots & \vdots & \vdots \\ 496 & 497 & 498 & 500 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

▷

Watch On YouTube

$\ell = 500$

$m = 125$

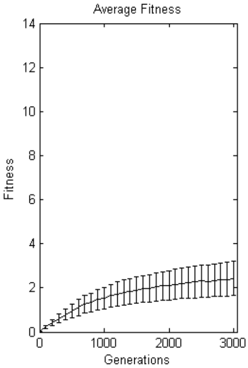
$n = 4$

$\delta = 0.2$

$\sigma = 1$

$$L = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ \vdots & \vdots & \vdots & \vdots \\ 496 & 497 & 498 & 500 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{bmatrix}$$



#trials=10. Error bars show standard error.

$\ell = 500$

$m = 125$

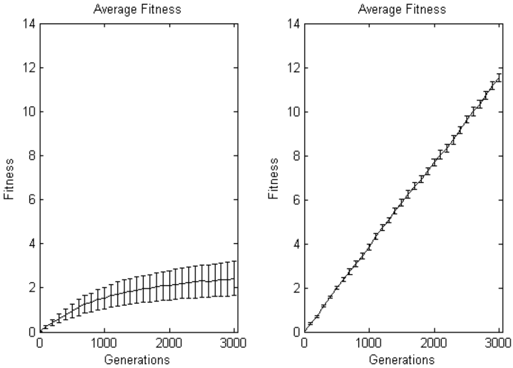
$n = 4$

$\delta = 0.2$

$\sigma = 1$

$$L = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ \vdots & \vdots & \vdots & \vdots \\ 496 & 497 & 498 & 500 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{bmatrix}$$



#trials=10. Error bars show standard error.

$\ell = 500$

$m = 125$

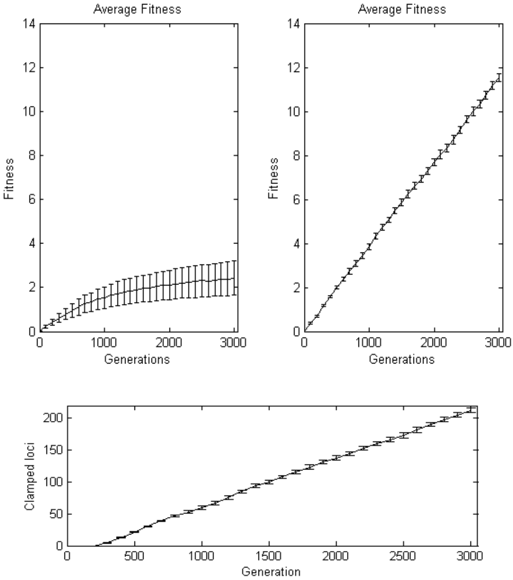
$n = 4$

$\delta = 0.2$

$\sigma = 1$

$$L = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ \vdots & \vdots & \vdots & \vdots \\ 496 & 497 & 498 & 500 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{bmatrix}$$



#trials=10. Error bars show standard error.

$\ell = 500$

$m = 125$

$n = 4$

$\delta = 0.2$

$\sigma = 1$

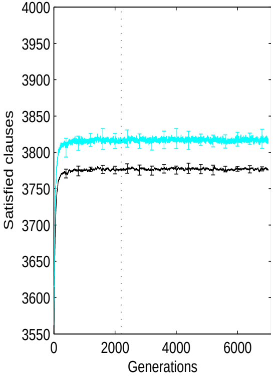
$$L = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ \vdots & \vdots & \vdots & \vdots \\ 496 & 497 & 498 & 500 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

▷
Watch On YouTube

Uniform Random MAX-3SAT

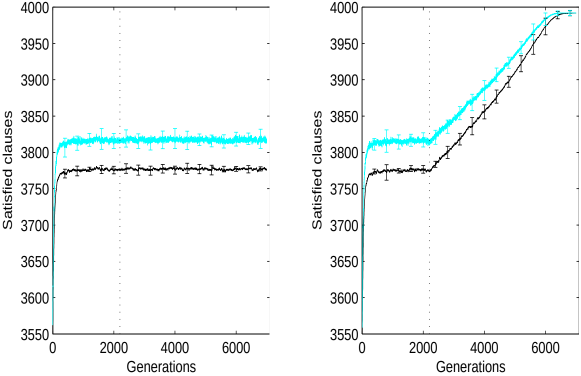
#variables=1000, #clauses=4000



#trials=10. Error bars show standard error.

Uniform Random MAX-3SAT

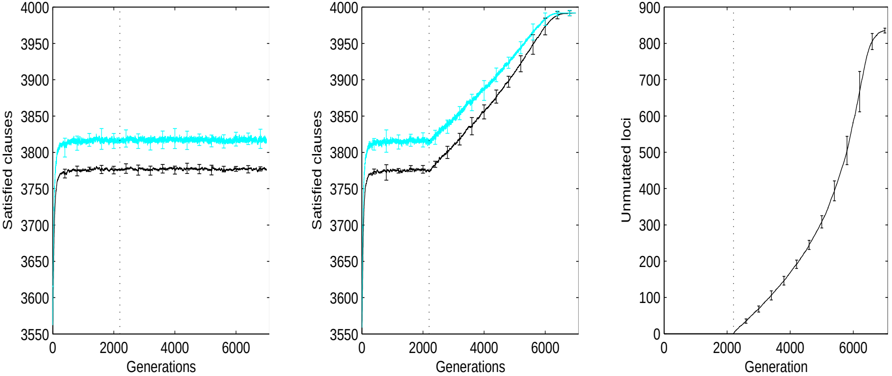
#variables=1000, #clauses=4000



#trials=10. Error bars show standard error.

Uniform Random MAX-3SAT

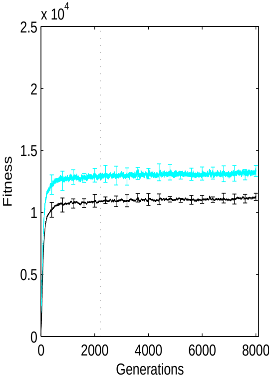
#variables=1000, #clauses=4000



#trials=10. Error bars show standard error.

SK Spin Glasses System

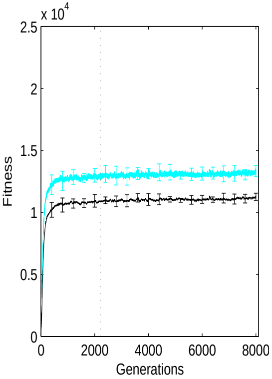
#spins=1000 drawn from $\{-1, +1\}$, $\text{fitness}(\sigma) = \sum_{1 \le i < j \le \ell} J_{ij}\sigma_i\sigma_j$, J_{ij} drawn from $\mathcal{N}(0,1)$

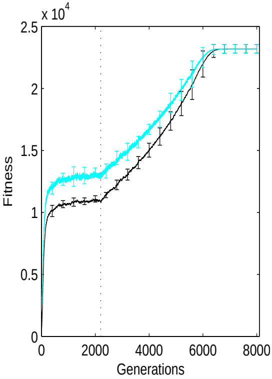


#trials=10. Error bars show standard error.

SK Spin Glasses System

#spins=1000 drawn from $\{-1,+1\}$, $\quad \text{fitness}(\sigma) = \sum_{1\leq i<j\leq \ell} J_{ij}\sigma_i\sigma_j, \quad J_{ij}$ drawn from $\mathcal{N}(0,1)$

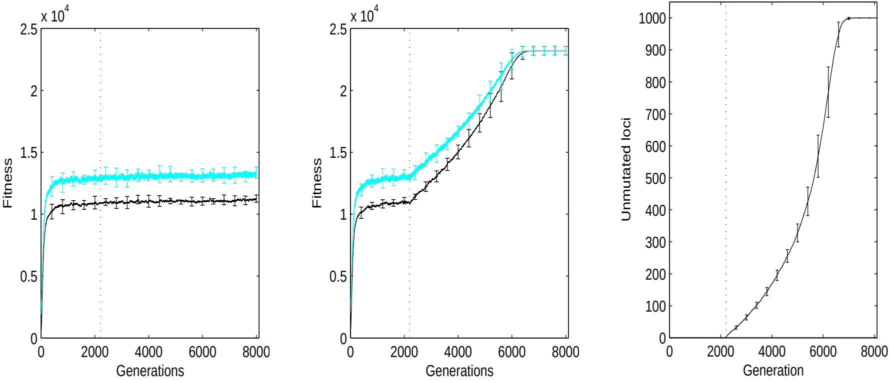




#trials=10. Error bars show standard error.

SK Spin Glasses System

#spins=1000 drawn from $\{-1,+1\}$, $\text{fitness}(\sigma) = \sum_{1\leq i<j\leq \ell} J_{ij}\sigma_i\sigma_j$, J_{ij} drawn from $\mathcal{N}(0,1)$



#trials=10. Error bars show standard error.

Conclusion

Summary

Proposed the Hyperclimbing Hypothesis

- Weak assumptions about the distribution of fitness
- Based on a computational efficiency of GAs with UX
- Upfront proof of concept
- Testable predictions + Validation on
  - Uniform Random MAX-3SAT
  - SK Spin Glasses Problem

Corollaries

- Implicit Parallelism is real
  - Not your grandmother's implicit parallelism
  - *Effect of coarse schema partitions* evaluated in parallel
    - NOT the average fitness of short schemata
  - Defining length of a schema partition is immaterial
    - New Implicit parallelism more powerful
- Think in terms of Hyperscapes not Landscapes

What Now?

- Flesh out the Hypothesis
  - What does a deceptive hyperscape look like?
- More Testable Predictions + Validation
  - Are staggered coarse conditional effects indeed common?
- Generalization
  - Explain optimization in other EAs
- Outreach
  - Identify and efficiently solve problems familiar to theoretical computer scientists
- Exploitation
  - Construct better representations
  - Choose better parameters for existing EAs
  - Construct better EAs (e.g. backtracking EAs)

Questions?

follow me @evohackr