Understanding Landscapes

Elena Popovici

Topics

- What are landscapes? ⇦
- Why talk about them?
  - To try to answer: “What makes search hard/easy?”
- Landscape characterizations

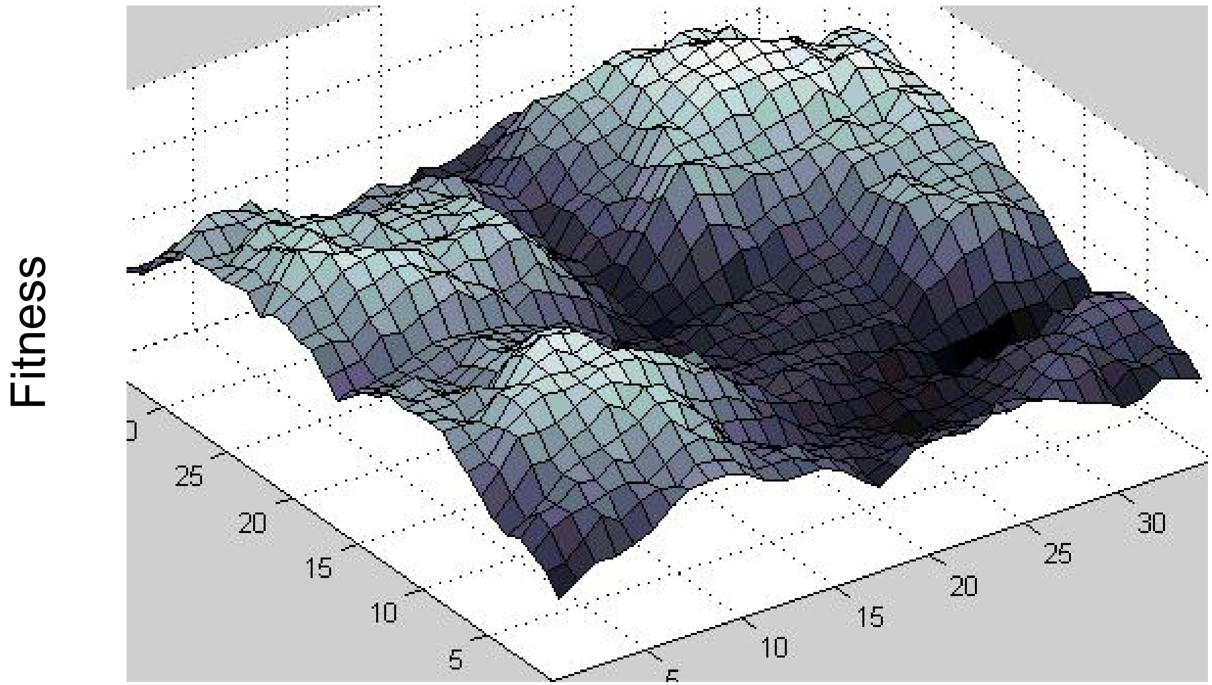
What are landscapes?

- Controversial issue
- Historical perspective
  - Landscapes in other fields
    - Biology (Wright's "surfaces of selective value" → "fitness landscapes")
      - Lack of rigorous definition
    - Physics & Chemistry
      - Some formal definitions
      - Highly customized for specific problems of interest
  - Landscapes in Computer Science - EC

Landscapes in EC

- Two main views
  - Search space + fitness function (+ neighborhood ?)
  - Search space + fitness function + operator (T. Jones)
- Search space can be:
  - Set of solutions (phenotypes / genotypes)
  - Set of sets of solutions (phenotypes / genotypes) i.e. populations (M. Vose)
- One problem – multiple choices of what the search space is and/or what the fitness function is
  - Is a problem hard/easy to solve? →
  - Is a landscape hard/easy to search? →
  - What landscape to construct for a problem to make it easy to solve?

Landscape = Space + Fitness



… + neighborhood

Harder to visualize for multiple dimensions!

… + neighborhood

- assume maximization & discrete search space -

- Plateau: a set of points (at least 2) that is the transitive closure of the neighbor-of-equal-fitness relation
- Peak region: a point or plateau whose fitness is strictly greater than that of all of its neighbors (the neighbors of a plateau = the reunion of the neighbors of the points of the plateau that are not already on the plateau)
- Peak: peak region made of just one point
- Global maximum: peak region of maximum fitness
- Local maxima: peak region that is not a global maxima
- Ridges, Valleys, Hills – intuitive but harder to define formally

- Different neighborhoods can be defined on the same space – structure changes (peaks, plateaus, etc.)

Landscape = Space + Fitness + Operator

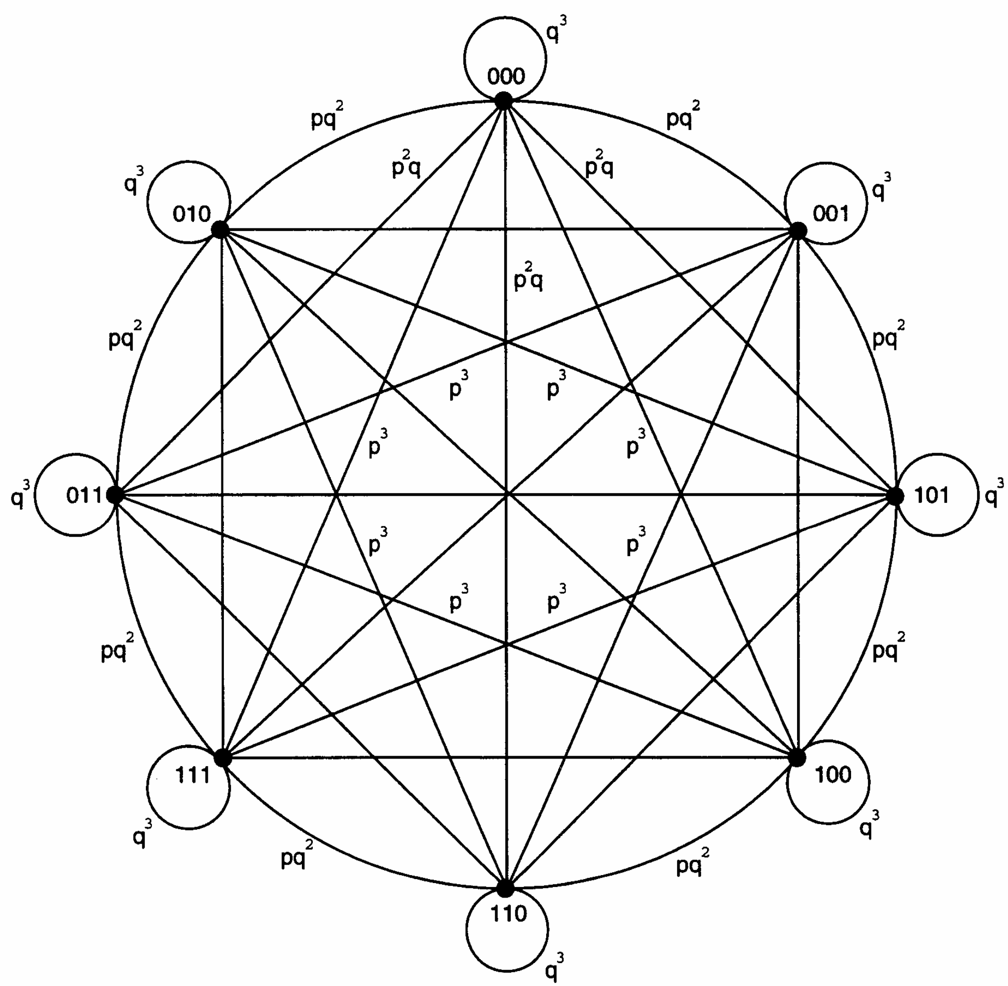


Figure 3. The mutation landscape for binary strings of length three. The mutation probability is p, and $q = 1 - p$. Some edge probabilities are omitted. Edges are bidirectional.

1-point operators

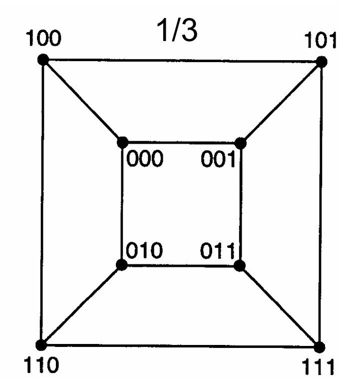


Figure 2. The landscape for the bit-flipping operator (β) on binary strings of length three. Edges are bidirectional and each has probability one-third.

Jones PhD 1995

Landscape = Space + Fitness + Operator

multiple-point operators

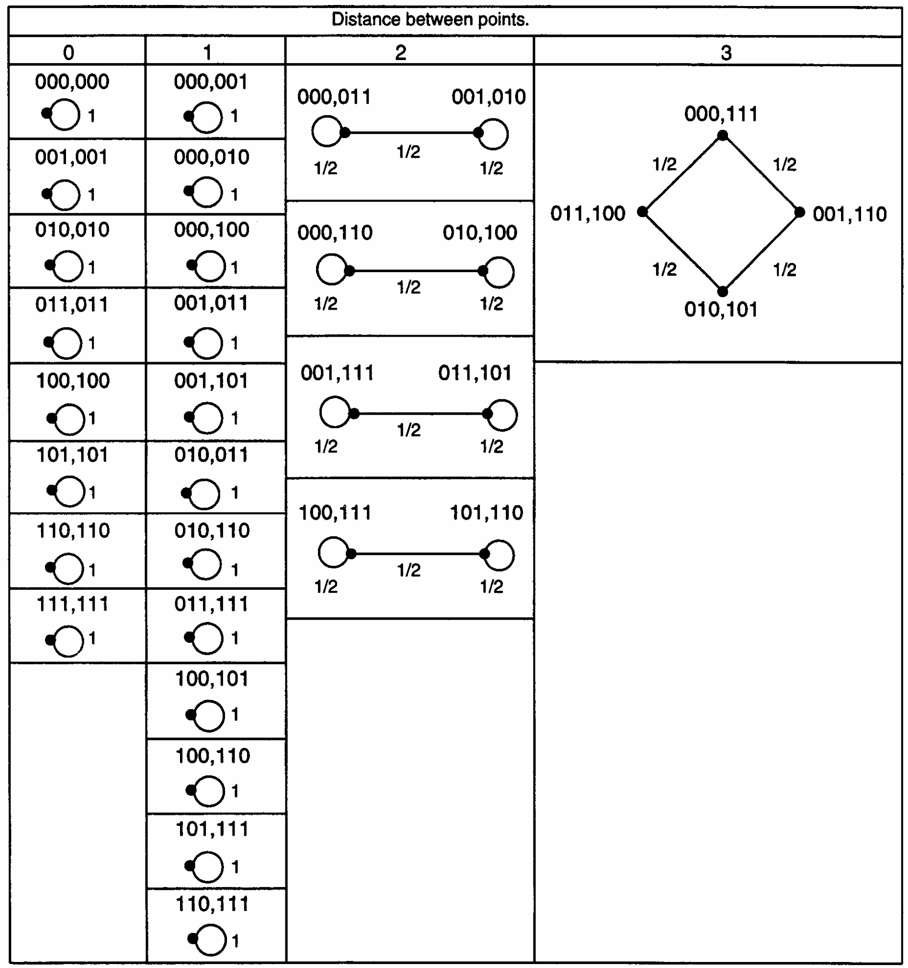
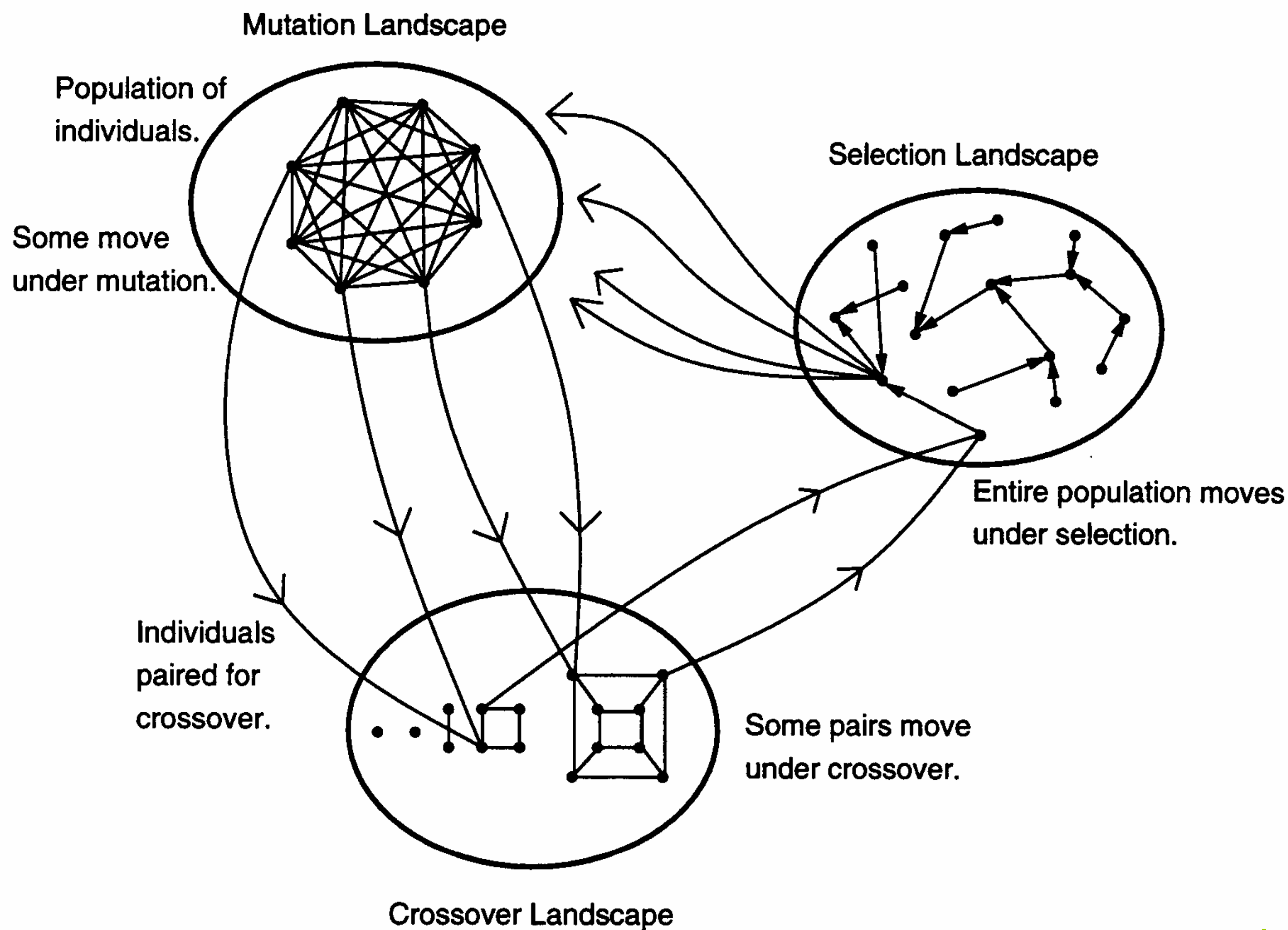


Figure 4. The one-point crossover landscape for binary strings of length three. The crossover operator, $\chi_1^{2\to2}$, produces two offspring from two parents. Edges are bidirectional.

Jones PhD 1995

Landscape = Space + Fitness + Operator



Jones PhD 1995

Topics

- What are landscapes? ✓
- Why talk about them? ⇦
- Landscape characterizations

Why Study Landscapes?

- Descriptive purposes
  - Understand what makes search hard/easy
- Prescriptive purposes
  - Make predictions
  - Design better algorithms

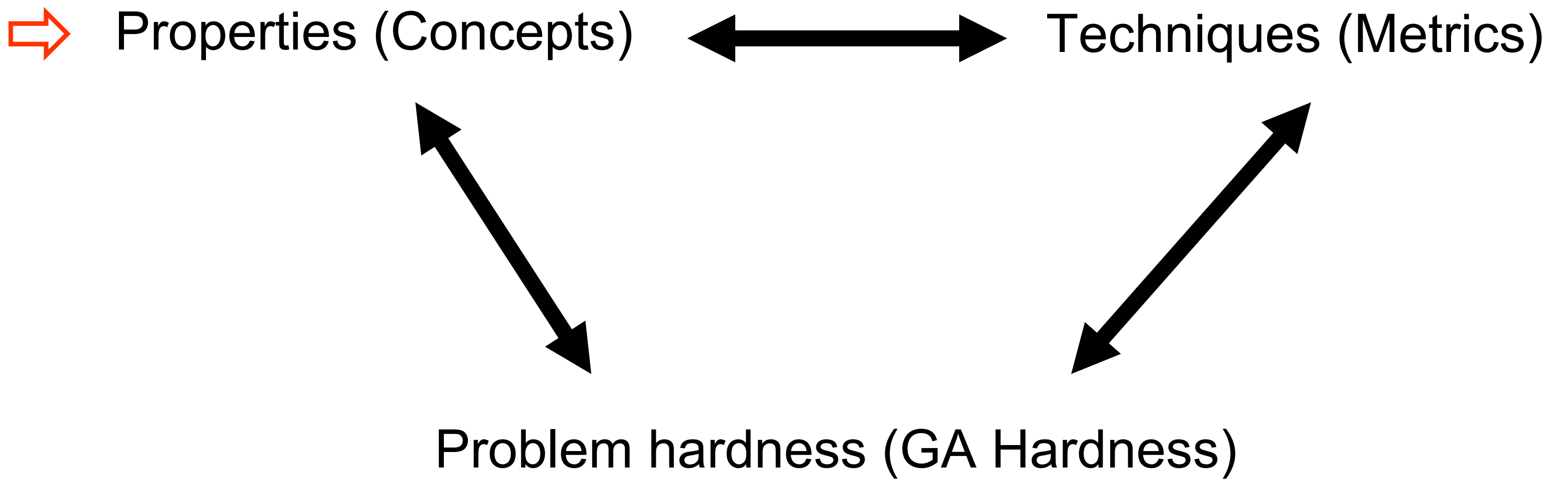
Topics

- What are landscapes? ✓
- Why talk about them? ✓
- Landscape characterizations ⇦

Studying Landscapes

- Approaches:
  - Introducing/identifying properties (concepts)
  - Introducing techniques/measures for analyzing such properties
  - Constructing landscapes with desired properties
  - Assessing the relevance of the properties with respect to the “hardness” of landscapes
  - Assessing the relevance of various landscape-characterizing techniques with respect to certain properties

Studying Landscapes



<u>Goals</u>:

- acquire understanding
- making predictions
- making better EA design choices

<u>Method types</u>:

- qualitative
- quantitative

Properties (Concepts)

- Ruggedness/Modality ⇦
- Deception
- Epistasis

Ruggedness/Modality

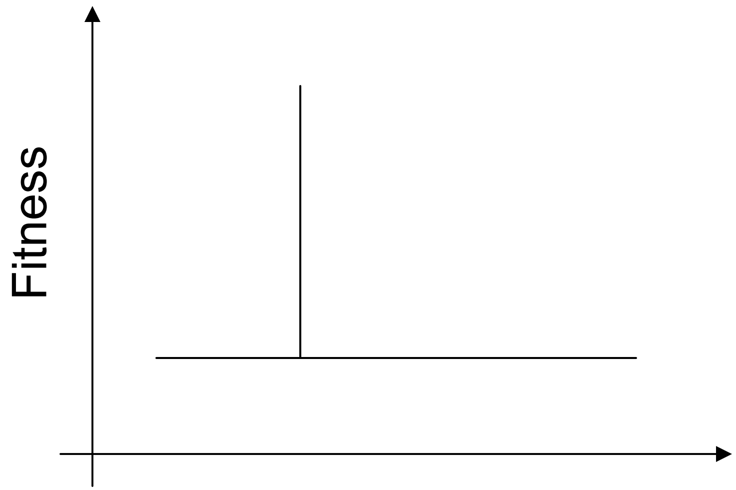
- Uni-modal landscape: a single peak (peak-region)
- Multi-modal: multiple peaks
- High modality: many peaks
- Rugged ≈ highly modal? Not really, ruggedness is a more complex (and vague ☹) concept
  - Modality can be quantified
  - How to quantify ruggedness??

Ruggedness/Modality

- Hypothesis: uni-modal / not rugged is easy

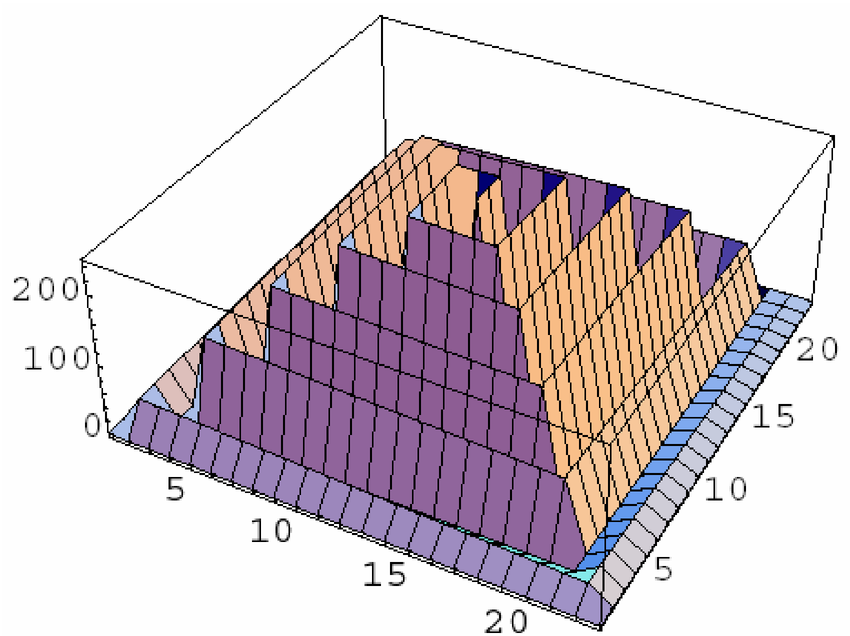
False! Uni-modal *can* be hard.

Needle in a haystack



Search needs a gradient

Long path problem



Easier for GA than for hill-climbing, but hard for GA too

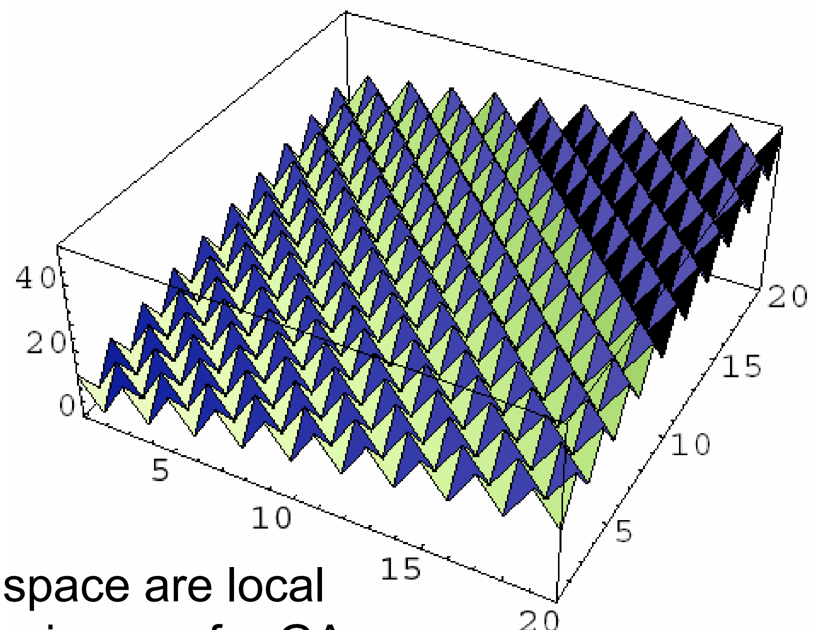
Horn & Goldberg 1995

Ruggedness/Modality

- Hypothesis: highly-modal (rugged) is hard

False! Highly multimodal *can* be easy.

Maximally multimodal problem



Half the points in the space are local optima, yet the problem is easy for GA

Horn & Goldberg 1995

Ruggedness/Modality

- What lies between the extremes??

Ideas:

- Structured ruggedness vs. “random” ruggedness (noise)?

Properties (Concepts)

- Ruggedness/Modality ✓
- Deception ⇦
- Epistasis

Deception

- Some “fit” low order hyperplanes of the search space “guide” the search toward some solution or building block that is not globally competitive
- Fitness of a hyperplane = average of fitnesses of the individuals in the hyperplane
- Fully deceptive

Goldberg 1987

f(0**)>f(1**)	f(00*)>f(11*),f(01*),f(10*)	f(000) = 28	f(001) = 26
f(*0*)>f(*1*)	f(0*0)>f(1*1),f(0*1),f(1*0)	f(010) = 22	f(100) = 14
f(**))>f(**1)	f(*00)>f(*11),f(*01),f(*10)	f(110) = 0	f(011) = 0
f(111)>f(000),f(001),f(010),f(011),f(100),f(101),f(110)		f(101) = 0	f(111) = 30

- Minimally deceptive, consistently deceptive …

Deception

- Two extremes
  - The only challenging problems for GAs are deceptive Whitley 1991
  - Deception is neither necessary nor sufficient for GA-hardness Grefenstette 1992

Deception

- Two extremes
  - The only challenging problems for GAs are deceptive Whitley 1991
  - ⇨ Deception is neither necessary nor sufficient for GA-hardness Grefenstette 1992

Deception

- Work on deception is based on the **Static** Building Block Hypothesis:
  "Given any short, low-order hyperplane partition, a GA is expected to converge to the hyperplane with the best *static* average fitness."

- The hypothesis ignores the distinction (made by the Schema Theorem) between *observed* and *static* fitness of a hyperplane.

Grefenstette 1992

Deception

- Some deceptive problems are GA-easy

$$\max f(x_1 2, x_2), 0 \le x_i \le 1$$

$$f(x_1, x_2) = \begin{cases} x_1^2 + 10x_2^2, x_2 < 0.995 \\ 2(1 - x_1)^2 + 10x_2^2, x_2 \ge 0.995 \end{cases}$$

- The optimum is (0,1).
- Encoded on 20 bits, the problem has a fully deceptive sub-problem of order 10: (1,#)
- The problem is easily solved by a standard GA.
- Why? *Collateral convergence*. The last 10 bits converge first and alter the observed fitness of the 00..0 schema for the first 10 bits.

Grefenstette 1992

Deception

- Some problems with no deception are GA-hard.

$$x \in [0,1], f(x) = \begin{cases} 2^{L+1}, x = 0 \\ x^2, otherwise \end{cases}, L = \# bits$$

- Any hyperplane H that contains the optimum, 0, has f(H)>2. Any hyperplane H that does not contain 0 has f(H)≤1. → no deception
- Needle in a hay stack problem, hard for GAs.
- Why? Because of high *variance* in the fitnesses of the hyperplanes associated with the optimum. Observed average never reflects static average.

Grefenstette 1992

Properties (Concepts)

- Ruggedness/Modality ✓
- Deception ✓
- Epistasis ⇦

Epistasis

- At a high level: epistasis – the degree of interdependence among genes (with respect to their contribution to fitness).
- How to quantify it?
  - Obvious for some problems (e.g. NK-, SAT-landscapes)
  - Harder to generalize
- Polygeny: the number of genes that influence one particular trait
- Pleitropy: the number of traits influenced by a particular gene
- Constant across genes/traits or variable

Epistasis and GA hardness

- No epistasis should be easy – Separable functions – Line search
- How much epistasis for a landscape to become hard?
- Which operators are more suitable for various degrees of epistasis?
  - As epistasis increases, the relative advantage of crossover over mutation is reduced, but still at high levels De Jong, Potter, Spears 1997

Quantifying Epistasis

- Davidor 1991 – a mathematical formula for epistasis
  - Requires full knowledge of the domain
  - Not quite clear (to me) why the formula should reflect epistasis
  - High variance if formulas are computed over samples (opposite results can be obtained)
  - Support of relationship between epistasis value and GA hardness in the paper is week

Properties (Concepts)

- Ruggedness/Modality ✓
- Deception ✓
- Epistasis ✓

Studying Landscapes

Goals:

- acquire understanding
- making predictions
- making better EA design choices

Method types:

- qualitative
- quantitative

Techniques/Measures

- Fitness-distance relationship
- Auto-correlation & correlation length
- Operator fitness correlation

Fitness-Distance Relationship

- Intuition based on parallel between EAs and general search techniques (e.g. A*)
  - Fitness function ↔ Heuristic function
  - Heuristics try to approximate distance from current point to goal
- Hypothesis: good correlation between fitness and distance (to global optima) should make search amenable
- Quantitative: computing actual correlation figure (standard formula from statistics)
  - Not always good to summarize the relationship between fitness and distance
- Qualitative: fitness vs. distance scatter plots
  - Can reveal structure of the landscape

Fitness-Distance Relationship

- Assessing the FD relationship can be used to predict problem hardness
- Not infallible
- Not generally applicable

FDC by examples

- Confirmation of intuitive results
- Explanation of unintuitive results
- Make comparisons (e.g. among representations)

FDC by examples

- Classify easy as easy, hard as hard

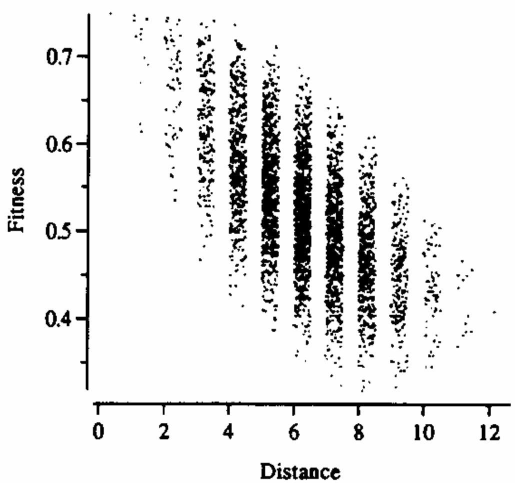


Figure 56. An NK landscape with $N = 12$ and $K = 1$ ($r = -0{\cdot}64$).

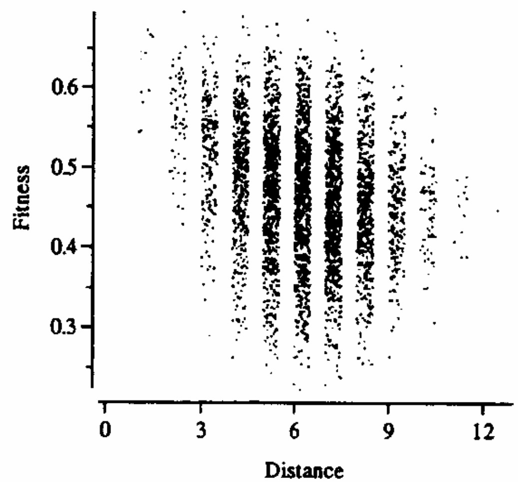


Figure 57. An NK landscape with $N = 12$ and $K = 3$ ($r = -0{\cdot}25$).

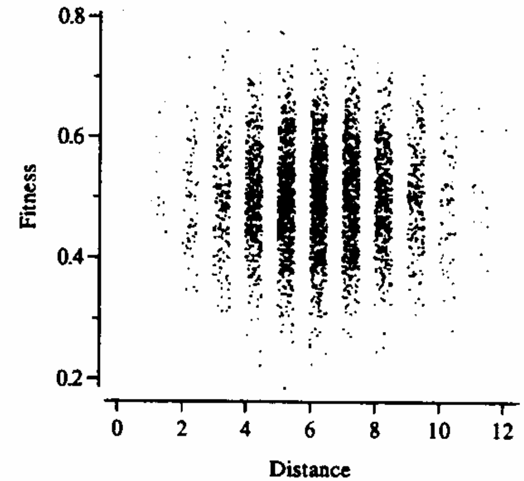


Figure 58. An NK landscape with $N = 12$ and $K = 11$ ($r = -0{\cdot}01$).

Jones PhD 1995

FDC by examples

- Classify easy as easy, hard as hard

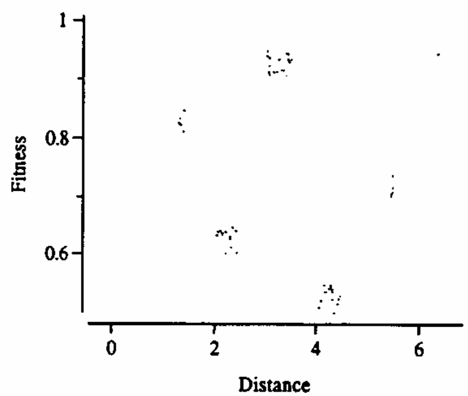


Figure 59. Deb & Goldberg's fully easy 6-bit problem ($r = -0{\cdot}23$).

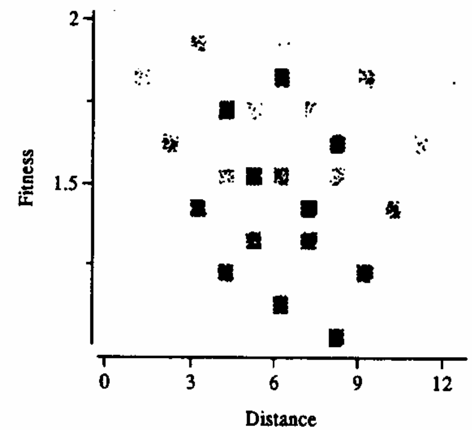


Figure 60. Two copies of Deb & Goldberg's fully easy 6-bit problem ($r = -0{\cdot}23$).

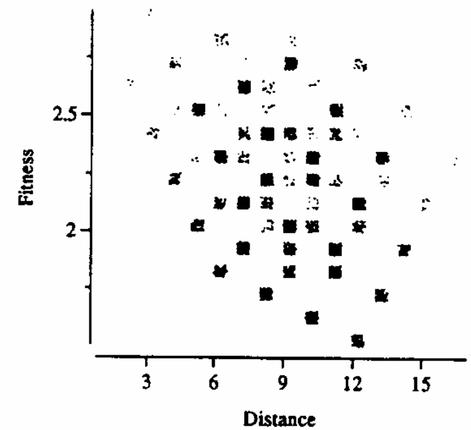


Figure 61. Three of Deb & Goldberg's fully easy 6-bit problems ($r = -0{\cdot}23$, 4000 sampled points).

  - Multiple copies of the same sub-problem don't make the problem harder
  - Separate problem difficulty from algorithm resources

Jones PhD 1995

FDC by examples

- Classify easy as easy, hard as hard

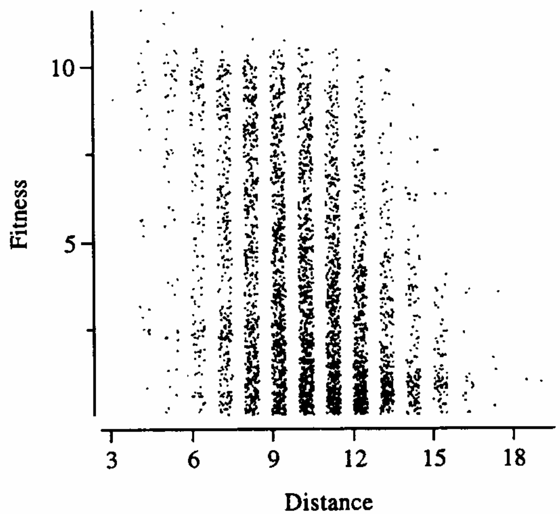


Figure 62. Grefenstette's deceptive but easy 20-bit problem ($r = -0{\cdot}32$, 4000 sampled points).

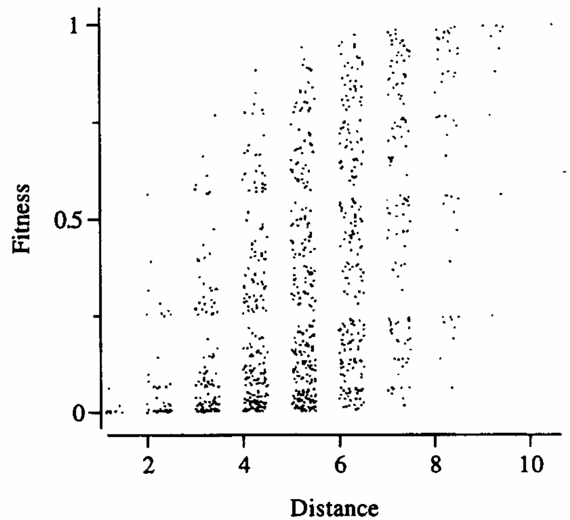


Figure 63. Grefenstette's non-deceptive but hard 10-bit problem. The single point with fitness 2048 is omitted from the plot. When included, $r = -0{\cdot}09$, when excluded, $r = 0{\cdot}53$).

Jones PhD 1995

FDC by examples

- Classify easy as easy, hard as hard

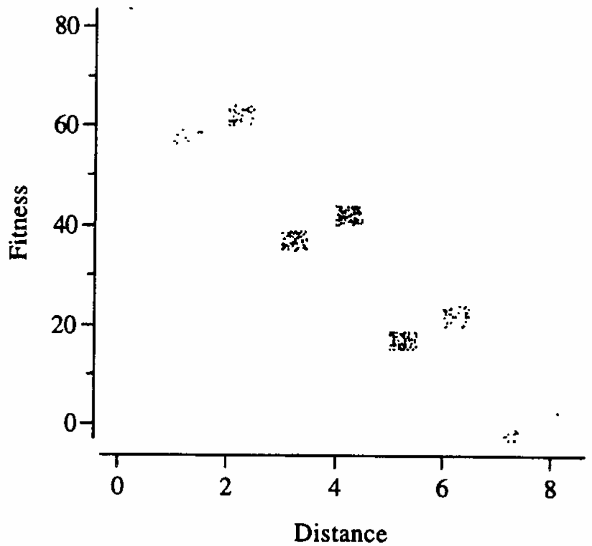


Figure 64. Ackley's porcupine problem on 8 bits ($r = -0{\cdot}88$).

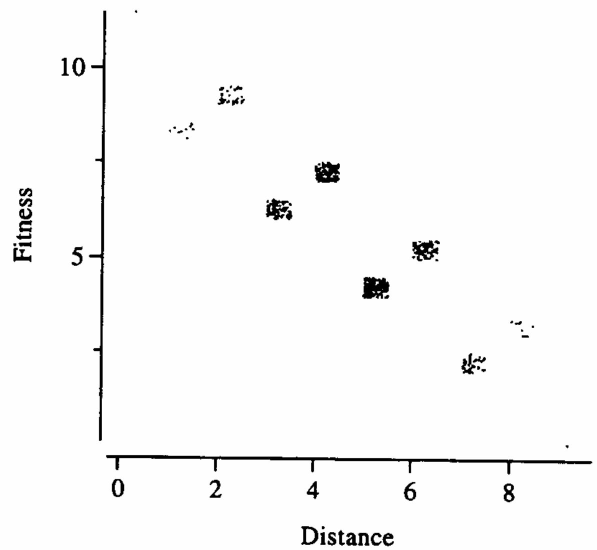


Figure 65. Horn & Goldberg's maximum modality problem on 9 bits ($r = -0{\cdot}83$).

High modality can be easy

Jones PhD 1995

FDC by examples

- Classify easy as easy, hard as hard

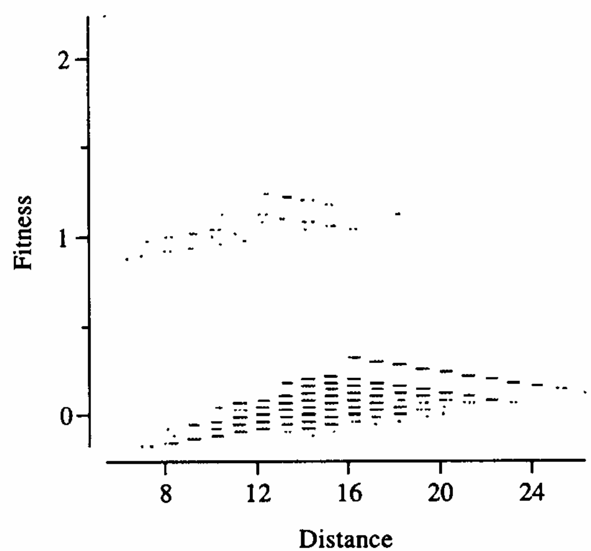


Figure 80. Holland's royal road on 32 bits ($b = 8$, $k = 2$ and $g = 0$), ($r = 0{\cdot}25$, 4000 sampled points).

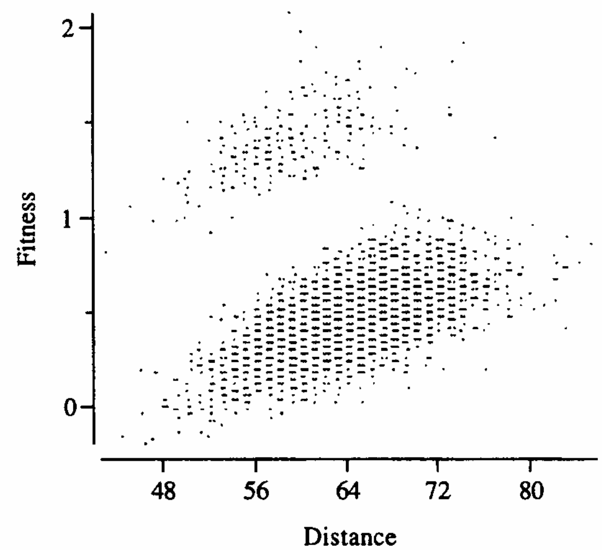


Figure 81. Holland's royal road on 128 bits ($b = 8$, $k = 4$ and $g = 0$), ($r = 0{\cdot}27$, 4000 sampled points).

Jones PhD 1995

FDC by examples

- Classify easy as easy, hard as hard

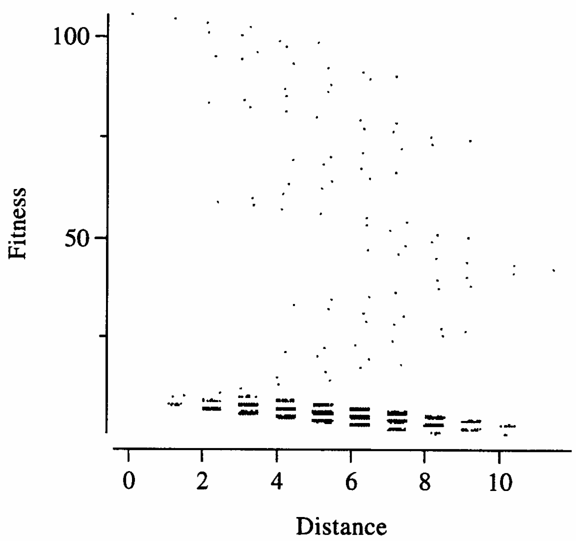


Figure 82. Horn, Goldberg & Deb's long path problem with 11 bits ($r = -0{\cdot}12$). Notice the path.

r_{path} = -0.39

$r_{non\text{-}path}$ = -0.67

Jones PhD 1995

FDC by examples

- Classify easy as hard

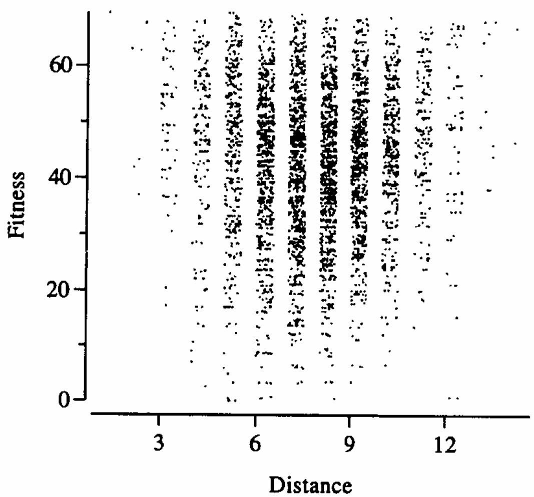


Figure 86. De Jong's F1 binary coded with 15 bits converted to a maximization problem ($r = -0{\cdot}01$, 4000 sampled points).

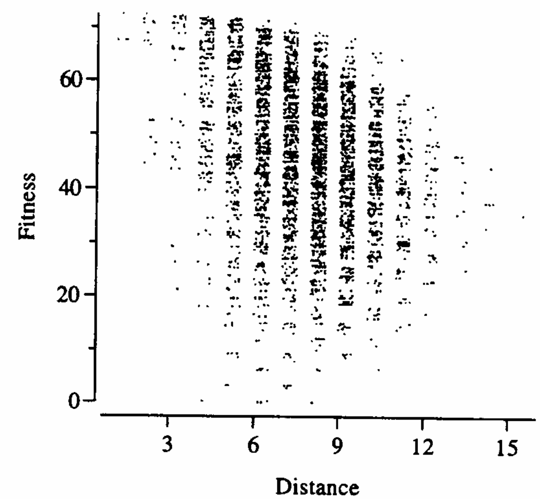


Figure 87. De Jong's F1 Gray coded with 15 bits ($r = -0{\cdot}30$, 4000 sampled points).

High fit points at all distances from the optimum, a GA is expected to have no problems locating one

Correct that Gray encoding is easier for GA on this problem.

FD relation reflects that representation makes a difference

Jones PhD 1995

FDC by examples

- Classify easy as hard

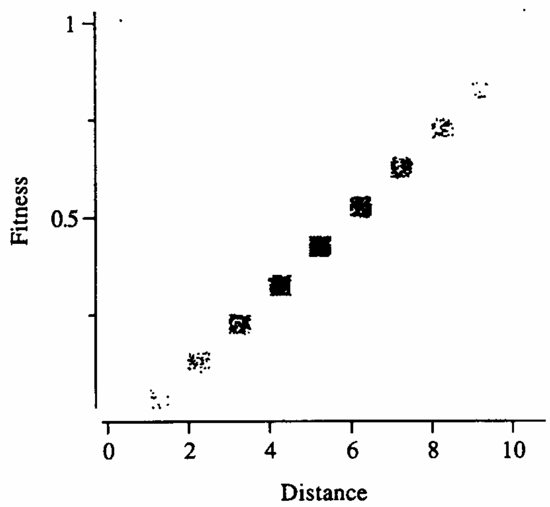


Figure 105. Liepins and Vose's fully deceptive problem on 10 bits ($r = 0{\cdot}98$).

Correctly shows that problem is hard

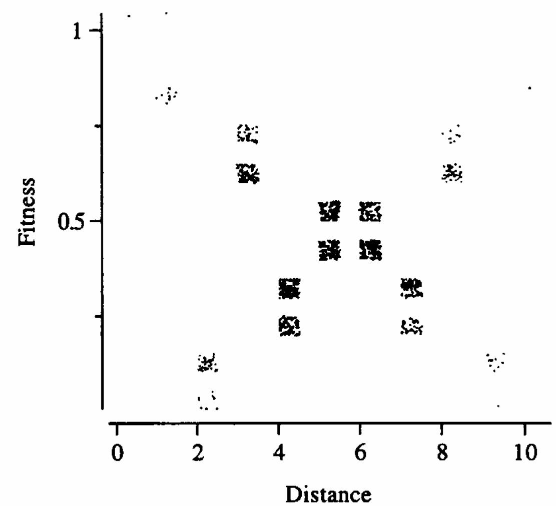


Figure 106. The transform of Liepins and Vose's fully deceptive problem on 10 bits ($r = -0{\cdot}02$). Correlation cannot detect the X structure.

Problem is easy because one of the substructures has FDC $\approx$ -1

FD relation reflects that encoding makes a difference

Jones PhD 1995

FDC by examples

- Classify hard as easy?
  - Can happen when FD relation is determined based on sampling and important points are missed

Jones PhD 1995

Issues with FD relation

- Knows nothing about GAs – good or bad?
  - Indicator of how difficult a problem should be (i.e. general hardness rather than GA hardness
  - FDC = -0.5 but GA does poorly → probably something wrong with the GA
- Requires a distance measure – what makes a good distance?
  - Define distance based on the operators intended for use (i.e. tell it something about GAs) – loose generality
- Says something about problems for which solutions are already known
  - Extend it by determining (local) optima based on a few hill-climbing runs and computing distance from those
    - Sampling can miss important points and give misleading conclusions
  - Can we really use it to make predictions about real world problems??
- Information on small examples of problems does not necessarily generalize to larger instances

Techniques/Measures

- Fitness-distance relationship ✓
- Auto-correlation & correlation length ⇦
- Operator fitness correlation

Auto-correlation and correlation length

- Auto-correlation: for each distance *h*, how correlated are the fitnesses of the points which are separated by that distance *h* from each other
  - Roughly corresponds to the distance one can jump and still have some information about the fitness there given the fitness here
  - Computed usually using pairs of points on some random walk through the space
- Correlation length τ: the distance *h* for which the auto-correlation is ½
- Distance is defined in terms of number of steps taken by some operator → function of the landscape under that operator
- Assumptions: landscape is *isotropic* (statistics of a random walk do not depend on the particular random walk used)
  - Is it really applicable in real world problems?

Manderick, Weger & Spiessens 1991

Auto-correlation and correlation length

- $\tau \approx$ exploratory horizon beyond which genetic search degrades to random search
  - τ is small $\rightarrow$ decrease rates of mutation & crossover

- Relationship between τ and EA performance

K	0	1	2	4	8	16	32	48	95
τ	29.96	24.37	19.51	14.15	7.06	3.90	1.72	1.00	0.52
Imp.	19.80	16.00	15.20	11.60	8.60	6.20	3.80	5.4	5.2

The relation between the correlation length τ of an NK-landscape and the number of improvements *Imp* found by GENITOR during runs of 2048 generations. The dimension of the landscapes is N = 96 and the degree of epistatic interaction K takes values K = 0,1, 2, 4, 8, 16, 32, 48, 95. The results are averaged over 5 runs.

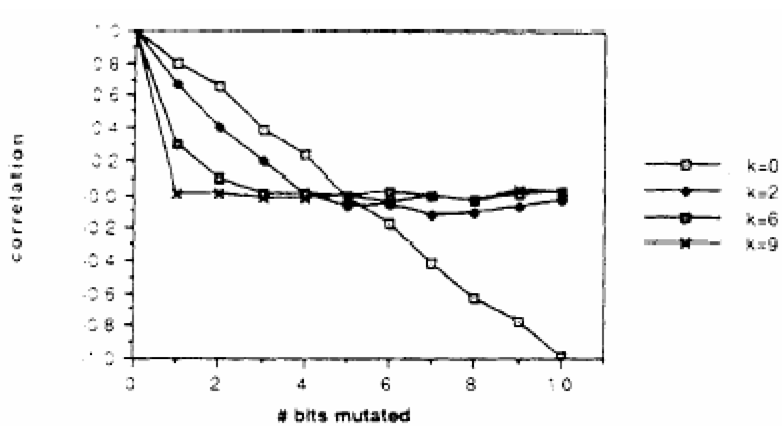
Manderick, Weger & Spiessens 1991

Techniques/Measures

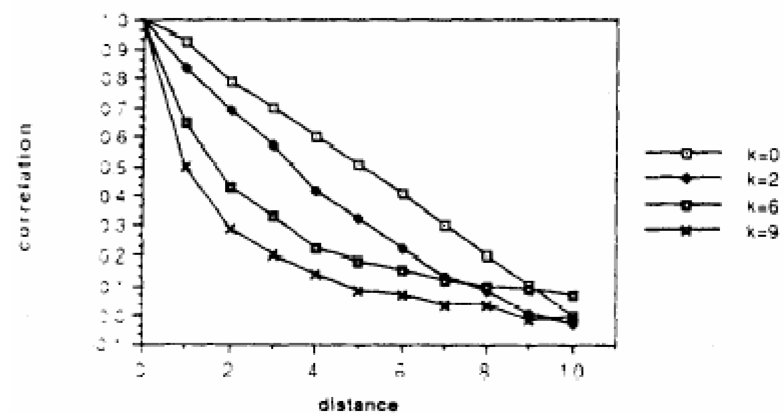
- Fitness-distance relationship ✓
- Auto-correlation & correlation length ✓
- Operator fitness correlation ⇦

Operator fitness correlation

- Correlation between fitness of parents and fitness of children
- Same isotropy assumption must hold.
- Results on NK landscapes



mutation



1-point crossover

Manderick, Weger & Spiessens 1991

Operator fitness correlation

- Results on TSP

$\rho_{OX} = 0.72$	$\rho_{Reverse} = 0.86$
$\rho_{PMX} = 0.61$	$\rho_{Remove\text{-}and\text{-}Reinsert} = 0.80$
$\rho_{CX} = 0.57$	$\rho_{Swap} = 0.77$
$\rho_{EX} = 0.90$	

- Hypothesis: high correlation means high EA performance
- Empirical results seem to support hypothesis, however, only 5 runs were performed

Manderick, Weger & Spiessens 1991

Operator fitness correlation

- Predictive models using operator fitness correlation
- Try to fit a linear model to the dependency between the fitness of the parents and the fitness of the children
- Static estimation vs. dynamic estimation
  - Dynamic estimates have a lot more variance
- Crossover: less linear & more variance
- Linearity and variance also depend on the problem and the operator rates

Grefenstette 1995

Operator fitness correlation

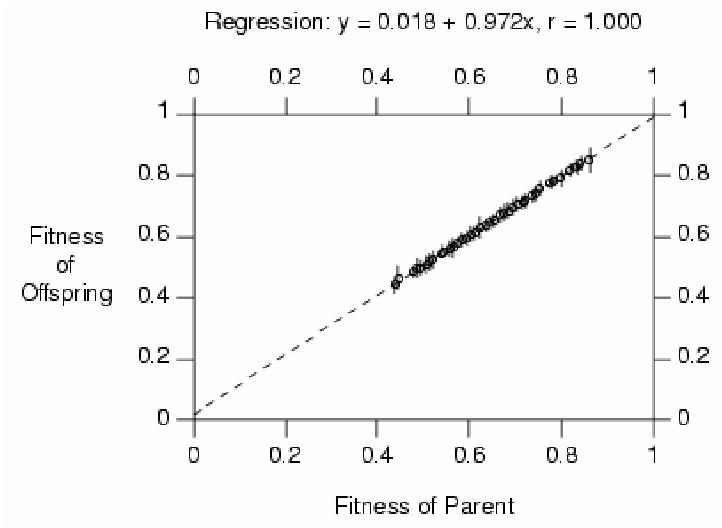


Figure 2: Static Estimates for Mutation (0.01) on f_1

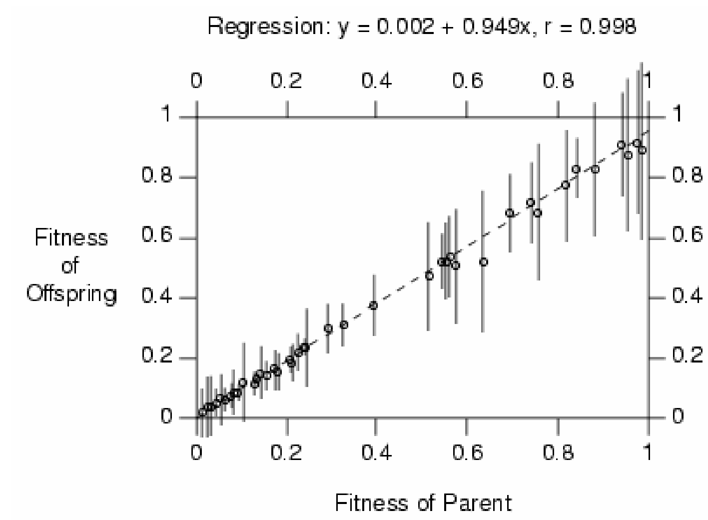


Figure 6: Static Estimates for Mutation (0.01) on f_5

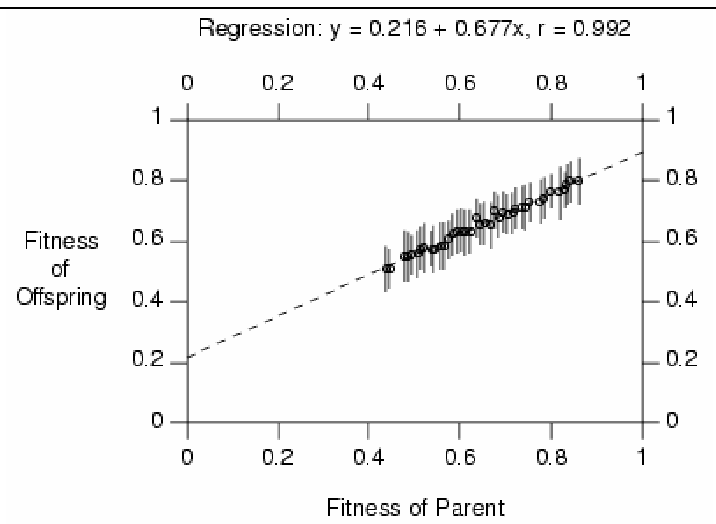


Figure 3: Static Estimates for Mutation (0.10) on f_1

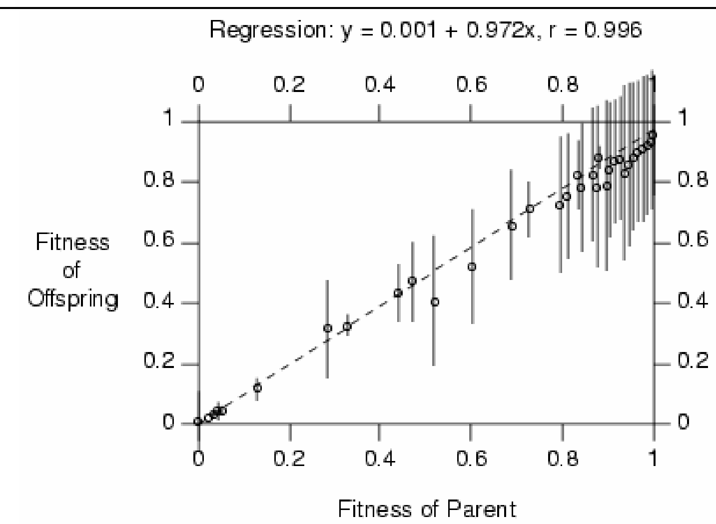


Figure 7: Dynamic Estimates for Mutation (0.01) on f_5

M
U
T
A
T
I
O
N

Grefenstette 1995

Operator fitness correlation

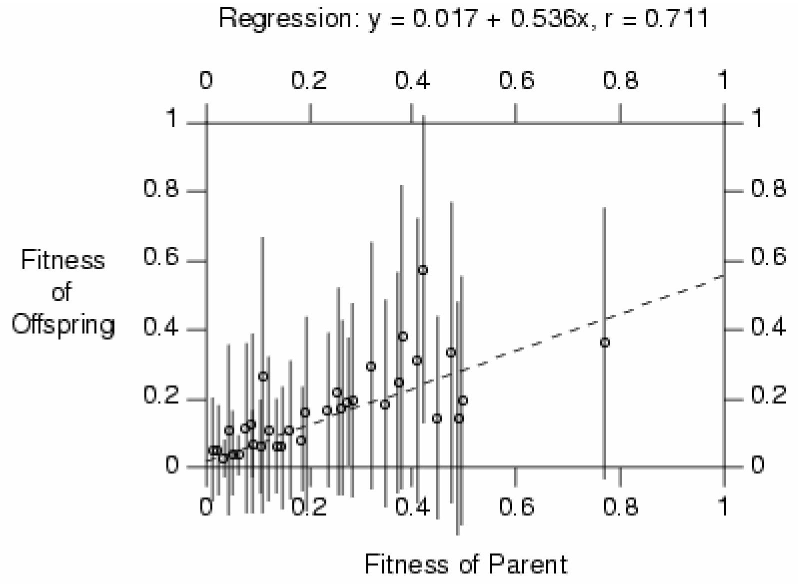


Figure 10: Static Estimates for 2pt on f_5

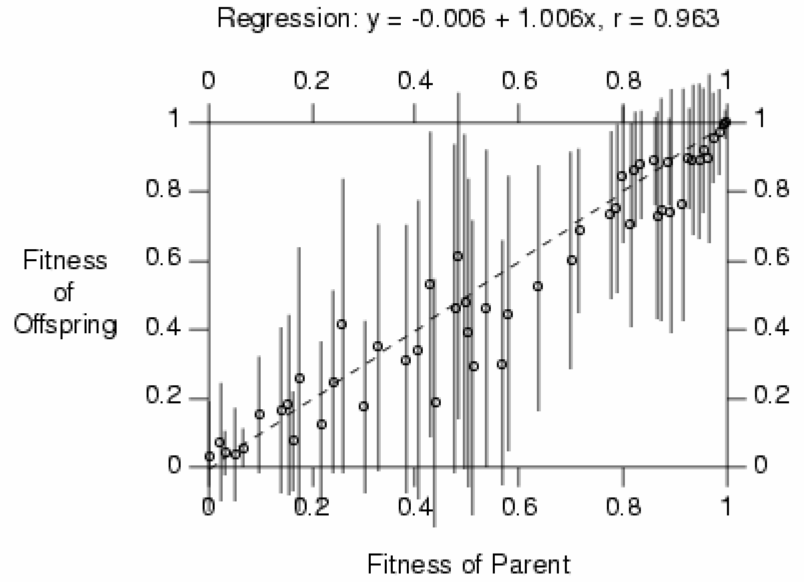


Figure 11: Dynamic Estimates for 2pt on f_5

CROSSOVER

Grefenstette 1995

Operator fitness correlation

- Iterate the model to predict change in average fitness of the population over time

Computations can get nasty when extending the methodology to more types of EAs

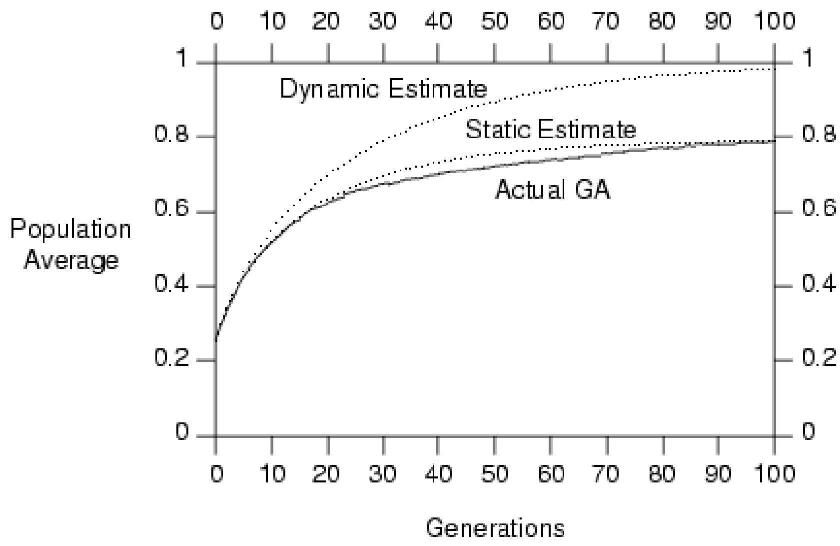


Figure 13: Predicted vs. Actual Population Fitness on f_7

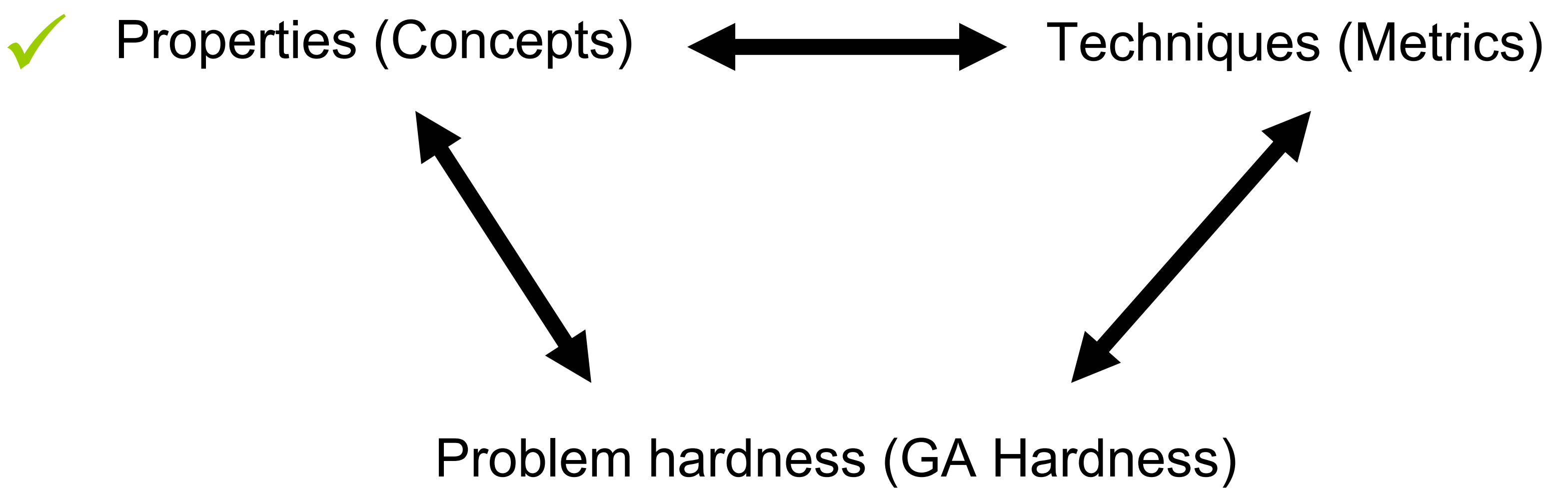
- As average fitness rises over time, smaller improvements result from same amount of variation
  - Scale fitness for average fitness to increase faster

Grefenstette 1995

Techniques/Measures

- Fitness-distance relationship ✓
- Auto-correlation & correlation length ✓
- Operator fitness correlation ✓

Studying Landscapes



Goals:

- acquire understanding
- making predictions
- making better EA design choices

Method types:

- qualitative
- quantitative

Topics

- What are landscapes? ✓
- Why talk about them? ✓
- Landscape characterizations ✓

Conclusions

- No single perfect way to look at the big picture
- Must consider all angles collectively to get the view
- Must extend work from analyzing hand-crafted, well known problems to real-world, unknown ones
- Identify additional, more relevant properties of landscapes to be used for problem-to-EA matching