# Ruggedness and Evolvability - An Evolution's-eye View

Lionel Barnett

Informatics Department & CCNR
University of Sussex

lionelb@sussex.ac.uk

August 2008

CCNR
Centre for Computational Neuroscience and Robotics

University of Sussex

# Outline

Fitness landscapes

- What are they?

# Outline

# Outline

# Outline

# Outline

# Fitness landscapes

- A fitness landscape is a mapping:

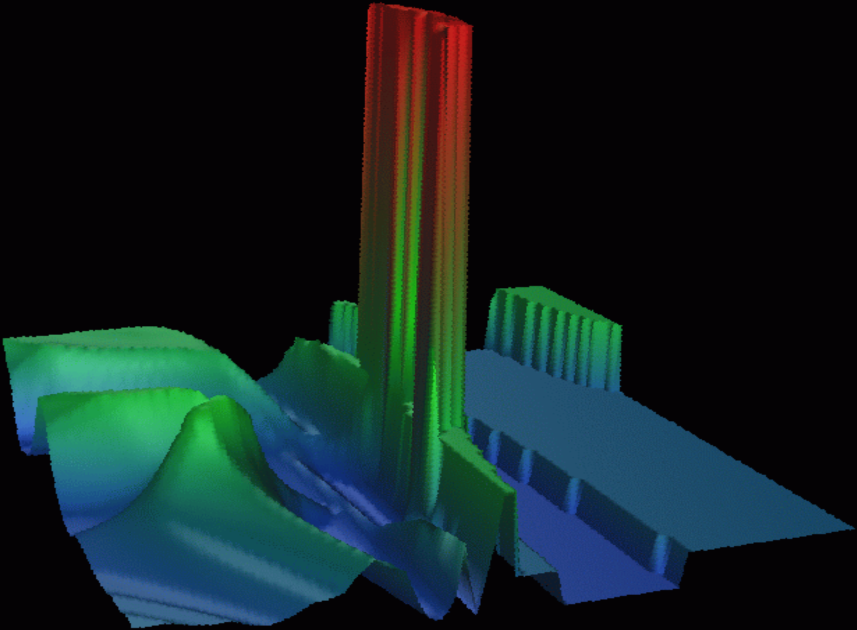
$$genotype \rightarrow phenotype \rightarrow fitness$$

- Genotypes reside in ***genotype space***
- Phenotypes reside in ***phenotype space***
- Fitness is a real number
  - Biology: fitness is the expected number of offspring of a phenotype
  - Artificial evolution: fitness is a measure of the "goodness" of the phenotype - something to be maximised
- Frequently the phenotype is skipped: i.e. a fitness landscape is just a mapping: $genotype \rightarrow fitness$

$$g \mapsto f(g)$$

  where $f(g)$ is the ***fitness function***

- "Fitness Landscape" is a visual/spatial metaphor...



# Landscape correlation: what is it and why might it be useful?

- Commonly viewed as a measure of fitness landscape ***ruggedness***

- Relates to ease of evolving high fitness phenotypes = ***evolvability***
  - Warning: "evolvability" is used in several specific technical senses

- ***What*** does ruggedness have to say about evolvability?

- There are certainly other, interrelated factors which affect evolvability
  - E.g. scale, neutrality, local sub-optima, deceptiveness, ...
  - Ruggedness is not the whole story

# Correlation

- Intuitively, correlation describes how two random measurements relate to each other
- E.g. pick people at random from a population:
  - Height and weight tend to go together: they are ***positively*** correlated
  - Height and the ability to pass under low doorways without bumping one's head are ***negatively*** correlated
  - Height and eye colour don't seem to be related: they are ***uncorrelated***
- Mathematical definition

$$\mathrm{corr}(X, Y) \equiv \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}$$

where $X$, $Y$ are random variables

# Correlated landscapes: the autocorrelation function

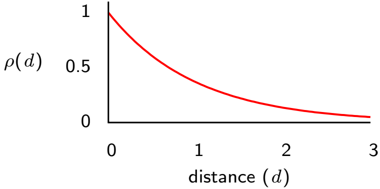
- The ***autocorrelation function*** - correlation between the fitness of two randomly selected genotypes a given distance apart:

  $$\rho(d) \equiv \operatorname{corr}(f(g_1), f(g_2))$$

  with $g_1, g_2$ selected uniformly from all genotypes distance $d$ apart
- A function of distance $d$ - need some notion of “distance”!
- Answers the question: how do the fitnesses of two genotypes at a given distance relate to each other?
- Tends to be $> 0$: small $\rho(d)$ = rugged, large $\rho(d)$ = smooth
- Generally, decays with increasing distance $d$. It is common to find a relation of the approximate form:

$$\rho(d) = \exp\left(-d/\ell\right)$$

The characteristic distance $\ell$ is called ***correlation length***



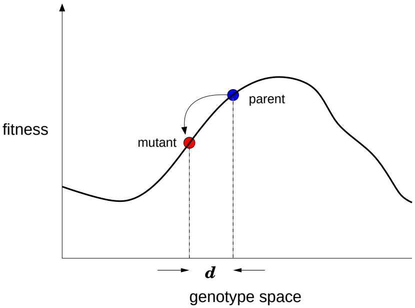
# Ruggedness and evolvability

- ***Mutation*** takes small steps in genotype space; i.e. evolution samples genotypes a small distance apart

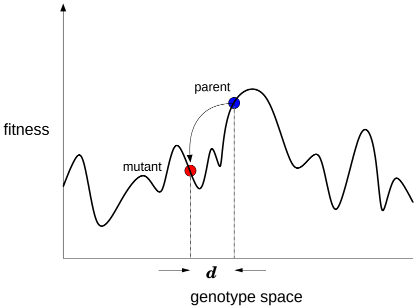
Rugged ⇒ no fitness correlation at small distance
⇒ mutation is a “leap in the dark”
⇒ mutant probably has rubbish fitness

Smooth ⇒ some (positive) fitness correlation at small distance
⇒ mutant fitness is “in same ball-park” as parent fitness
⇒ mutant has a fighting chance of being *fitter* than the parent

# Ruggedness and evolvability



A smooth landscape



A rugged landscape

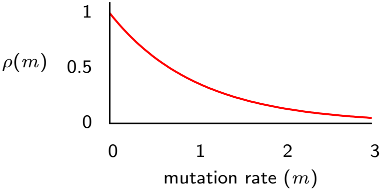
- Note: there is no place for ***recombination*** in this argument. Recombinant genotypes are *not* generally a small distance from the parent genotypes
  - There have been attempts to define “ruggedness for recombination”, but. . . basically, autocorrelation has nothing to say about recombination

# Genotype distance = mutation (rate)

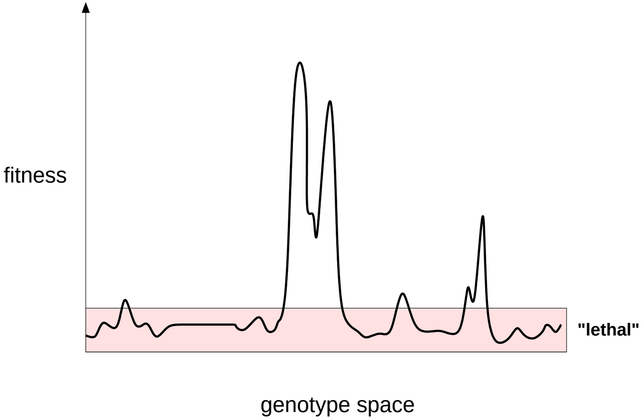
- The autocorrelation function naturally measures correlation between the fitness of a genotype and its ***mutants***:

  $$\rho(\mathcal{M}) \equiv \mathrm{corr}(f(g), f(\mathcal{M}(g)))$$

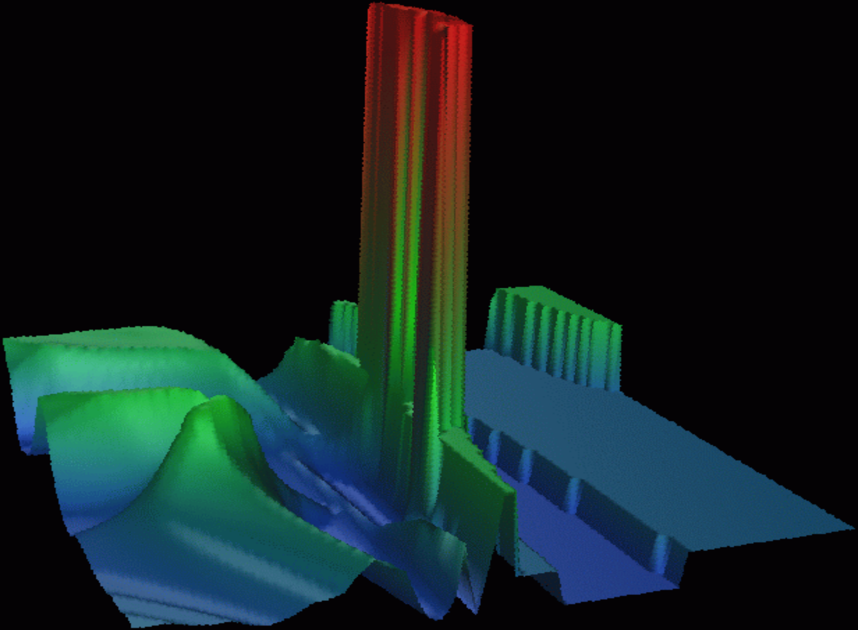
  where $\mathcal{M}$ is a ***mutation operator*** and genotype $g$ is selected uniformly at random from genotype space
- Answers the question: how do the fitnesses of a genotype and its mutant relate to each other? [This is precisely what evolution needs to know...]
- If mutation operator is uniform random over genotypes a fixed distance $d$ from parent, we recover the original definition $\rho(d)$
- May be considered a function of ***mutation rate*** $m$:



# The problem with uniform sampling



Real fitness landscapes look like this



# The problem with uniform sampling

- Uniform sampling is "biased" towards low fitness (lethal?) genotypes
- Evolution most decidely does ***not*** sample genotype space uniformly!
  - It spends (hopefully) as little as time possible sampling low-fitness regions
- So an autocorrelation statistic based on uniform sampling tells us mostly about an uninteresting (to evolution) region of the landscape
- It's worse than that - to estimate $\rho$ by *finite* uniform sampling, we might well end up sampling *only* lethals. . .

# The problem with uniform sampling

- Uniform sampling is “biased” towards low fitness (lethal?) genotypes

- Evolution most decidely does ***not*** sample genotype space uniformly!
  - It spends (hopefully) as little as time possible sampling low-fitness regions

- So an autocorrelation statistic based on uniform sampling tells us mostly about an uninteresting (to evolution) region of the landscape

- It's worse than that - to estimate $\rho$ by *finite* uniform sampling, we might well end up sampling *only* lethals. . .

- ***Solution***: Let evolution do the sampling

# An evolution's-eye view

- Let evolution do the sampling: first pick your evolutionary algorithm (and mutation operator)
- The algorithm proceeds by creating parent-mutant pairs: so use them to calculate autocorrelation:

  $$\rho(\mathcal{M}) \equiv \mathrm{corr}(f(g), f(\mathcal{M}(g)))$$

  where $g, \mathcal{M}(g)$ are sampled from all parent-mutant pairs encountered over the *statistical ensemble* of evolutionary runs
- Finite sampling-friendly: just perform multiple runs of your GA and collate all parent-mutant fitness pairs created during the course of the runs
  - If you like, repeat with varying mutation rate for correlation length
- This autocorrelation-as-evolution-sees-it is no longer "algorithm-agnostic"
  - Is this a good or a bad thing? (discuss)
  - Your landscape may appear more or less rugged depending on the algorithm. Perhaps this could aid in choice of algorithm?

# The mutant fitness distribution

- Taking a step back: we are interested in correlation because we are interested in the distribution of fitness of mutants
  - *cf.* the *transmission function* (Altenberg, Smith & Husbands, . . . )
- The Mutant Fitness Distribution:

$$\mathcal{F}(x) \equiv f(\mathcal{M}(g))\,|f(g) = x$$

- For given $x$, $\mathcal{F}(x)$ is a *random variable* (distribution) - read:

***the fitness of a mutant given that the parent has fitness*** $x$

# The mutant fitness distribution

- We can calculate the moments of $\mathcal{F}(x)$:

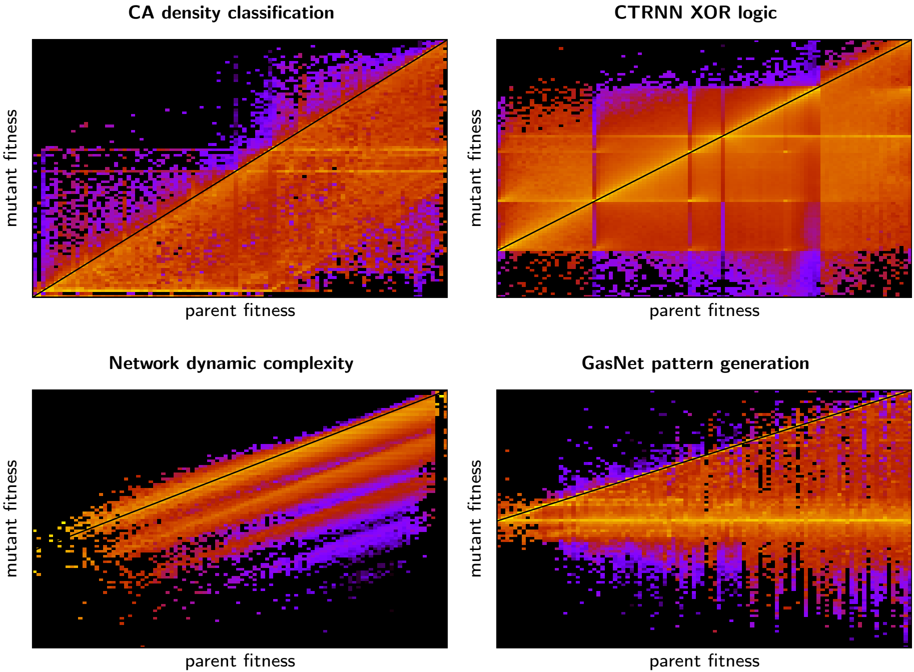
$$\begin{array}{rcll} \mu(x) & \equiv & \mathbf{E}(\mathcal{F}(x)) & \qquad \textit{mean mutant fitness} \\ \sigma^2(x) & \equiv & \mathrm{var}(\mathcal{F}(x)) & \qquad \textit{mutant fitness variance} \end{array}$$

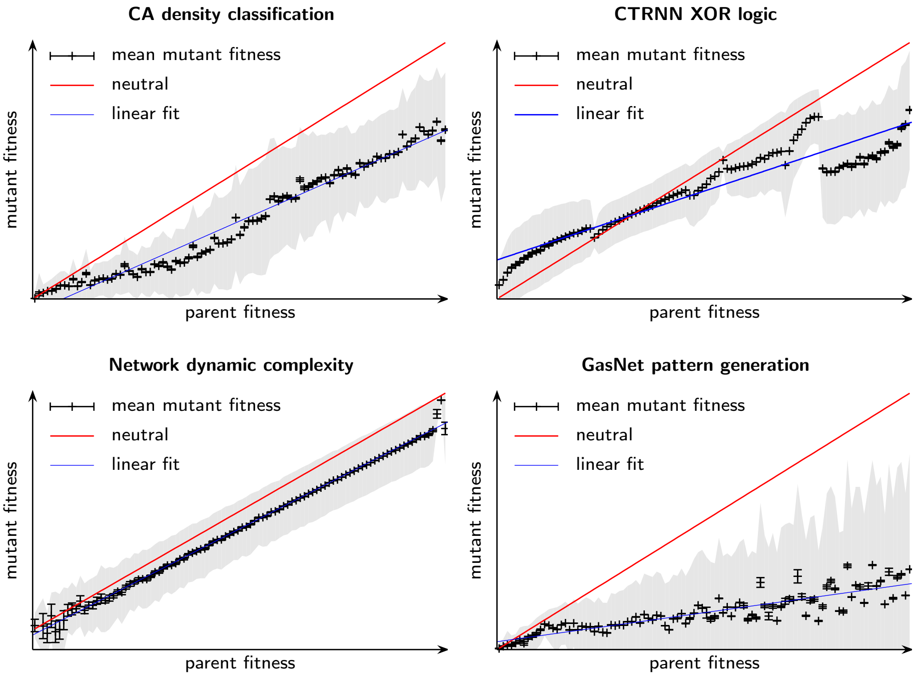
- In fact, knowing just the mean mutant fitness function $\mu(x)$, we can calculate correlation:

$$\rho(\mathcal{M}) = \frac{\mathrm{cov}(f(g), \mu(f(g)))}{\mathrm{var}(f(g))}$$

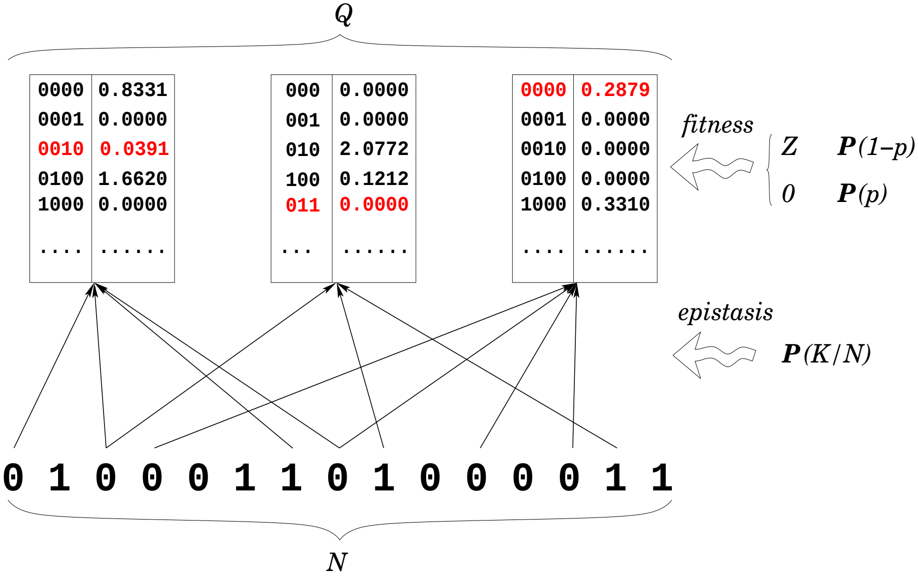
  where genotypes $g$ are sampled from...whatever$^*$ distribution
  - So $\mathcal{F}(x)$ is a *finer-grained* statistic than autocorrelation

- Since $\mathcal{F}(x)$ is conditioned on parent fitness, it doesn't suffer from uniform sampling bias
  - Although it is still obviously problematic to estimate in finite sample
  - But, as for autocorrelation, we can use evolution to do the sampling


CA density classification
mutant fitness
parent fitness
CTRNN XOR logic
mutant fitness
parent fitness
Network dynamic complexity
mutant fitness
parent fitness
GasNet pattern generation
mutant fitness
parent fitness


CA density classification
CTRNN XOR logic
Network dynamic complexity
GasNet pattern generation
mean mutant fitness
neutral
linear fit
mutant fitness
parent fitness

# QNKp landscapes



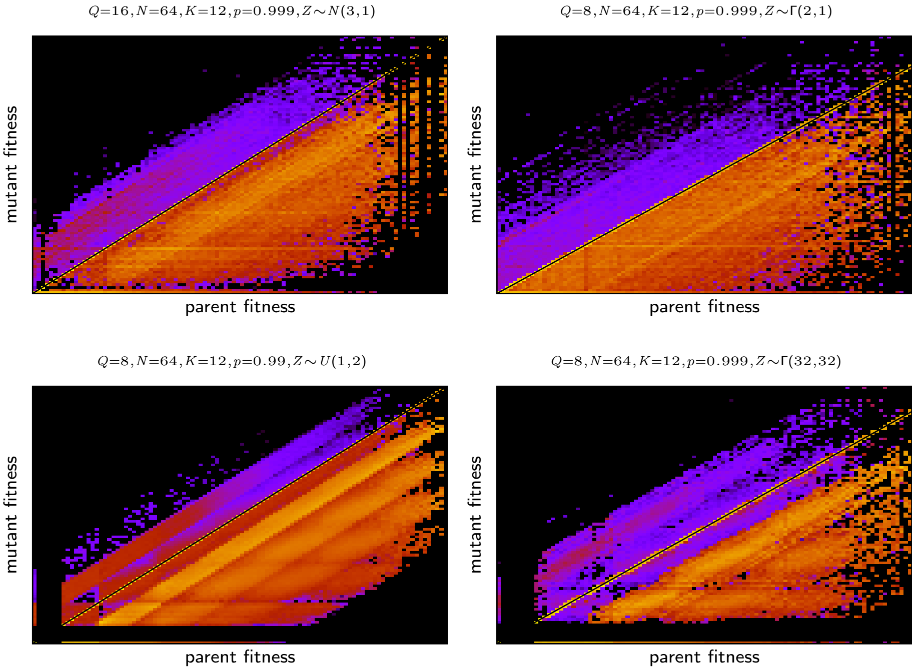
|  |  |
|---|---|
| 0000 | 0.8331 |
| 0001 | 0.0000 |
| 0010 | 0.0391 |
| 0100 | 1.6620 |
| 1000 | 0.0000 |
| .... | ...... |

|  |  |
|---|---|
| 000 | 0.0000 |
| 001 | 0.0000 |
| 010 | 2.0772 |
| 100 | 0.1212 |
| 011 | 0.0000 |
| ... | ...... |

|  |  |
|---|---|
| 0000 | 0.2879 |
| 0001 | 0.0000 |
| 0010 | 0.0000 |
| 0100 | 0.0000 |
| 1000 | 0.3310 |
| .... | ...... |

# QNKp landscapes

- Parameters:
  - $Q$ ***quantitative traits***
  - Genotype length $N$
  - Mean epistasis $K$
  - Fitness distribution $Z$
  - Neutrality $p$
- Each trait links independently to each locus with probability $K/N$
- Fitness tables filled independently: $0$ with probability $p$, else from $Z$
- Properties:
  - Tunably rugged, tunably neutral
  - Correlation length $\ell = N/K$ - does not depend on $Q, Z, p$
  - So in particular, ***ruggedness may be varied independently of neutrality***
  - $\mu(x)$ is ***linear*** in fitness $x$
  - $\sigma^2(x)$ is ***quadratic*** in fitness $x$


$Q{=}16, N{=}64, K{=}12, p{=}0.999, Z \sim N(3,1)$
mutant fitness
parent fitness
$Q{=}8, N{=}64, K{=}12, p{=}0.999, Z \sim \Gamma(2,1)$
mutant fitness
parent fitness
$Q{=}8, N{=}64, K{=}12, p{=}0.99, Z \sim U(1,2)$
mutant fitness
parent fitness
$Q{=}8, N{=}64, K{=}12, p{=}0.999, Z \sim \Gamma(32,32)$
mutant fitness
parent fitness

# Why QNKp landscapes might be a useful model

- $Q$ parametrises (actual or notional) quantitative traits in the phenotype
  - $Q$ may well be $\ll N$
  - Conventional NK models associate one trait per locus. . . why?
- Linearity of $\mu(x)$ is a consequence of independence of epistasis per trait
- Large neutrality parameter $p$ implies high proportion of lethals and decreasing neutrality with increasing fitness
  - Horizontal and diagonal "banding" reflects (near-)***neutral networks***

# Why QNKp landscapes might be a useful model

- $Q$ parametrises (actual or notional) quantitative traits in the phenotype
  - $Q$ may well be $\ll N$
  - Conventional NK models associate one trait per locus. . . why?

- Linearity of $\mu(x)$ is a consequence of independence of epistasis per trait

- Large neutrality parameter $p$ implies high proportion of lethals and decreasing neutrality with increasing fitness
  - Horizontal and diagonal "banding" reflects (near-)***neutral networks***

- *Remark*: conventional GAs perform rather poorly on QNKp landscapes. Random mutation hill-climbers and in particular ***simulated annealing*** fare rather better

# Take home message (decreasing credibility order)

# Take home message (decreasing credibility order)

- Standard autocorrelation is useless for realistic fitness landscapes

# Take home message (decreasing credibility order)

- Standard autocorrelation is useless for realistic fitness landscapes
- All realistic fitness landscapes are like QNKp landscapes

## Take home message (decreasing credibility order)

- Standard autocorrelation is useless for realistic fitness landscapes
- All realistic fitness landscapes are like QNKp landscapes
- Forget the GA, use simulated annealing