Runtime Analysis of Evolutionary Algorithms: Basic Introduction[1]

Per Kristian Lehre
University of Nottingham
Nottingham NG8 1BB, UK
PerKristian.Lehre@nottingham.ac.uk

Pietro S. Oliveto
University of Sheffield
Sheffield S1 4DP, UK
P.Oliveto@sheffield.ac.uk

GECCO'15 Companion, July 11-15, 2015, Madrid, Spain.
ACM 978-1-4503-3488-4/15/07.
http://dx.doi.org/10.1145/2739482.2756588

[1] For the latest version of these slides, see http://www.cs.nott.ac.uk/~pkl/gecco2015.

Bio-sketch - Dr Per Kristian Lehre

- Assistant Professor in the School of Computer Science, at the University of Nottingham.
- MSc and PhD in Computer Science from Norwegian University of Science and Technology (NTNU).
- Research on theoretical aspects of evolutionary algorithms and other randomised search heuristics.
- Editorial board member of Evolutionary Computation. Guest editor for special issues of IEEE Transactions of Evolutionary Computation and Theoretical Computer Science.
- Best paper awards at GECCO 2006, 2009, 2010, 2013, ICSTW 2008, and ISAAC 2014, nominations at CEC 2009, and GECCO 2014.
- Coordinator of 2M euro SAGE EU project unifying population genetics and EC theory.
- Organiser of Dagstuhl seminar on Evolution and Computation.

Bio-sketch - Dr Pietro S. Oliveto

- Vice-Chancellor's Fellow and EPSRC Early Career Fellow in the Department of Computer Science, at the University of sheffield.
- Laurea Degree in Computer Science from the University of Catania, Italy (2005).
- PhD in Computer Science (2006-2009), EPSRC PhD+ Research Fellow (2009-2010), EPSRC Postdoctoral Fellow in Theoretical Computer Science at the University of Birmingham, UK
- Research on theoretical aspects of evolutionary algorithms and other randomised search heuristics.
- Guest editor for special issues of *Evolutionary Computation* (MIT Press, 2015) and *Computer Science and Technology* (Springer, 2012).
- Best paper awards at GECCO 2008, ICARIS 2011, GECCO 2014 and best paper nominations at CEC 2009, ECTA 2011.
- Chair of IEEE CIS Task Force on Theoretical Foundations of Bio-inspired Computation.

Aims and Goals of this Tutorial

- This tutorial will provide an overview of
  - the goals of time complexity analysis of Evolutionary Algorithms (EAs)
  - the most common and effective techniques
- You should attend if you wish to
  - theoretically understand the behaviour and performance of the search algorithms you design
  - familiarise with the techniques used in the time complexity analysis of EAs
  - pursue research in the area
- enable you or enhance your ability to
  1. understand theoretically the behaviour of EAs on different problems
  2. perform time complexity analysis of simple EAs on common toy problems
  3. read and understand research papers on the computational complexity of EAs
  4. have the basic skills to start independent research in the area
  5. follow the other theory tutorials later on today

Evolutionary Algorithms and Computer Science

Goals of design and analysis of algorithms

1. correctness
   "does the algorithm always output the correct solution?"
2. computational complexity
   "how many computational resources are required?"

For Evolutionary Algorithms (General purpose)

1. convergence
   "Does the EA find the solution in finite time?"
2. time complexity
   "how long does it take to find the optimum?"
   (time = n. of fitness function evaluations)

Brief history

Theoretical studies of Evolutionary Algorithms (EAs), albeit few, have always existed since the seventies [Goldberg, 1989];

- Early studies were concerned with explaining the *behaviour* rather than analysing their performance.
- *Schema Theory* was considered fundamental;
  - First proposed to understand the behaviour of the simple GA [Holland, 1992];
  - It cannot explain the performance or limit behaviour of EAs;
  - Building Block Hypothesis was controversial [Reeves and Rowe, 2002];
- No Free Lunch [Wolpert and Macready, 1997]
  - Over all functions...
- Convergence results appeared in the nineties [Rudolph, 1998];
  - Related to the time limit behaviour of EAs.

Convergence

Definition

- Ideally the EA should find the solution in finite steps with probability **1** (visit the global optimum in finite time);
- If the solution is held forever after, then the algorithm converges to the optimum!

Conditions for Convergence ([Rudolph, 1998])

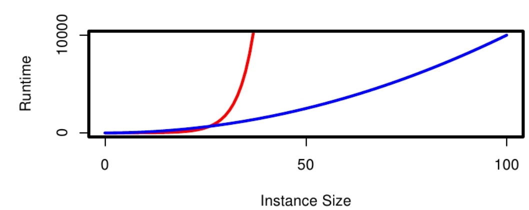
1. *There is a positive probability to reach any point in the search space from any other point*
2. *The best found solution is never removed from the population (elitism)*

- Canonical GAs using mutation, crossover and proportional selection Do Not converge!
- *Elitist* variants Do converge!

In practice, is it interesting that an algorithm converges to the optimum?

- Most EAs visit the global optimum in finite time (RLS does not!)
- How much time?

Computational Complexity of EAs



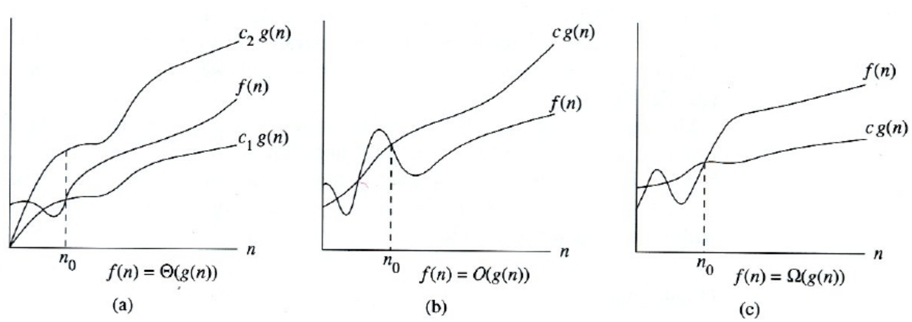
Generally means predicting the resources the algorithm requires:

- Usually the computational time: the number of primitive steps;
- Usually grows with size of the input;
- Usually expressed in asymptotic notation;

Exponential runtime: Inefficient algorithm
Polynomial runtime: "Efficient" algorithm
However (EAs):

1. In practice the time for a fitness function evaluation is much higher than the rest;
2. EAs are **randomised algorithms**
   - They do not perform the same operations even if the input is the same!
   - They do not output the same result if run twice!

Asymptotic notation



$f(n) \in O(g(n)) \iff \exists$ constants $c, n_0 > 0$ st. $0 \leq f(n) \leq cg(n) \quad \forall n \geq n_0$

$f(n) \in \Omega(g(n)) \iff \exists$ constants $c, n_0 > 0$ st. $0 \leq cg(n) \leq f(n) \quad \forall n \geq n_0$

$f(n) \in \Theta(g(n)) \iff f(n) \in O(g(n))$ and $f(n) \in \Omega(g(n))$

$$f(n) \in o(g(n)) \iff \lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$$

Goals

Understand how the runtime depends on:

- parameters of the problem
- parameters of the algorithm

In order to:

- explain the success or the failure of these methods in practical applications,
- understand which problems are optimized (or approximated) efficiently by a given algorithm and which are not
- guide the choice of the best algorithm for the problem at hand,
- determine the optimal parameter settings,
- aid the algorithm design.

Evolutionary Algorithms

(μ+λ) EA

Initialise P_0 with μ individuals chosen uniformly a random from $\{0,1\}^n$
for $t = 0, 1, 2, \ldots$ until stopping condition met **do**
Create λ new individuals by

- choosing $x \in P_t$ uniformly at random
- flipping each bit in x with probability p

Create the new population P_{t+1} by
choosing the best μ individuals out of $\mu + \lambda$.
end for

- If $\mu = \lambda = 1$, then we get the (1+1) EA;
- $p = 1/n$ is generally considered a good parameter setting [Bäck, 1993, Droste et al., 1998];
- By introducing stochastic selection and crossover we obtain a Genetic Algorithm (GA)

(1+1) Evolutionary Algorithm

(1+1) EA

Initialise x uniformly at random from $\{0,1\}^n$.
repeat
Create x' by flipping each bit in x with $p = 1/n$.
if $f(x') \geq f(x)$ **then**
$x \leftarrow x'$.
end if
until stopping condition met.

If only one bit is flipped per iteration: Random Local Search (RLS).

How does it work?

- Given x, how many bits will flip in expectation?

$$E[X] = E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n] =$$

$$(E[X_i] = 1 \cdot 1/n + 0 \cdot (1 - 1/n) = 1 \cdot 1/n = 1/n \quad E(X) = np)$$

$$= \sum_{i=1}^{n} 1 \cdot 1/n = n/n = 1$$

(1+1) EA: 2

How likely is it that exactly one bit flips? $\Pr(X = j) = \binom{n}{j} p^j (1-p)^{n-j}$

- What is the probability of flipping exactly one bit?

$$\Pr(X = 1) = \binom{n}{1}\left(\frac{1}{n}\right)\left(1 - \frac{1}{n}\right)^{n-1} = \left(1 - \frac{1}{n}\right)^{n-1} \geq 1/e \approx 0.37$$

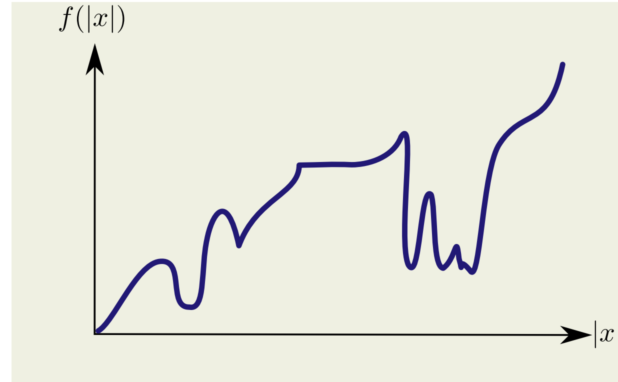
Is flipping two bits more likely than flipping none?

$$\begin{aligned}\Pr(X = 2) &= \binom{n}{2}\left(\frac{1}{n}\right)^2\left(1 - \frac{1}{n}\right)^{n-2} \\ &= \frac{n(n-1)}{2}\left(\frac{1}{n}\right)^2\left(1 - \frac{1}{n}\right)^{n-2} \\ &= \frac{1}{2}\left(1 - \frac{1}{n}\right)^{n-1} \approx 1/(2e)\end{aligned}$$

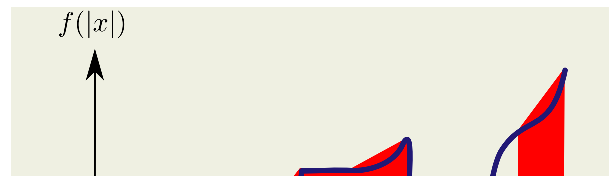
While

$$\Pr(X = 0) = \binom{n}{0}(1/n)^0 \cdot (1 - 1/n)^n \approx 1/e$$

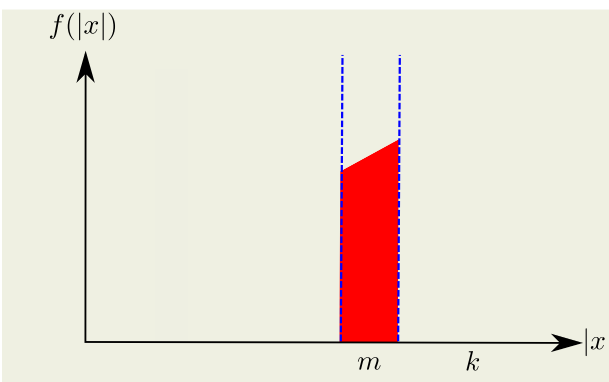
Running Example - Functions of Unitation



$$g(x) = f\left(\sum_{i=1}^{n} x_i\right) \quad \text{where} \quad f : \mathbb{R} \to \mathbb{R}$$

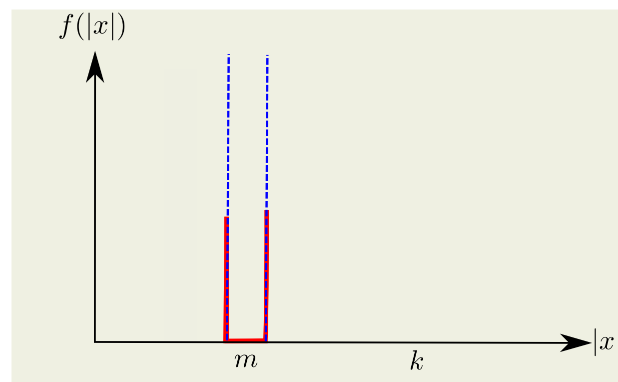


Linear Unitation Block



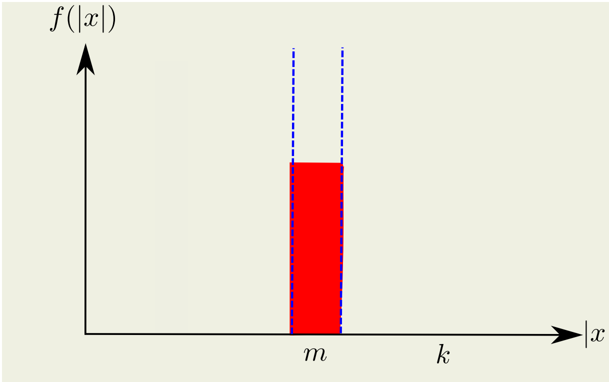
$$f(|x|) = \begin{cases} a|x| + b & \text{if } k < n - |x| \leq k + m \\ 0 & \text{otherwise.} \end{cases}$$

Toy Problem Framework - Gap



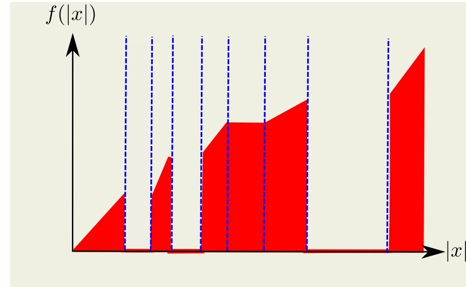
$$f(|x|) = \begin{cases} a & \text{if } n - |x| = k + m \\ 0 & \text{otherwise.} \end{cases}$$

Toy Problem Framework - Plateau



$$f(|x|) = \begin{cases} a & \text{if } k < n - |x| \leq k + m \\ 0 & \text{otherwise.} \end{cases}$$

Upper bound on the total runtime



$$f(x) = \sum_{i=1}^{r} f_i(x)$$

Assumptions

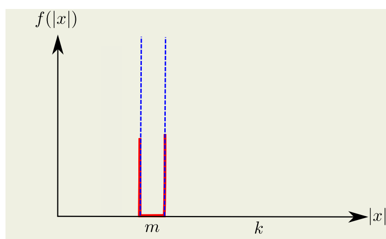
- r sub-functions $f_1, f_2, \ldots, f_r$
- T_i time to optimise sub-function f_i
- the evolutionary algorithm is elitist

By linearity of expectation, an upper bound on the expected runtime is

$$\mathbb{E}[T] \leq \mathbb{E}\left[\sum_{i=1}^{r} T_i\right] = \sum_{i=1}^{r} \mathbb{E}[T_i].$$

Gap block: upper and lower bounds

$$f(|x|) = \begin{cases} a & \text{if } n - |x| = k + m \\ 0 & \text{otherwise.} \end{cases}$$



The probability p of optimising a gap block of length m at position k is

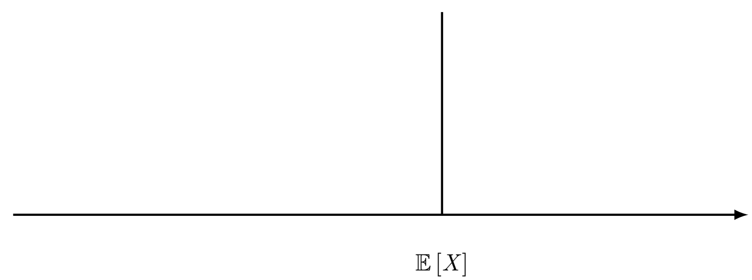
$$\left(\frac{m+k}{nm}\right)^m \frac{1}{e} \leq \binom{m+k}{m}\left(\frac{1}{n}\right)^m \frac{1}{e} \leq p \leq \binom{m+k}{m}\left(\frac{1}{n}\right)^m \leq \left(\frac{(m+k)e}{nm}\right)^m$$

The expected time to optimise the gap block is $1/p$

$$\left(\frac{nm}{(m+k)e}\right)^m \leq \binom{m+k}{m}^{-1} n^m \leq \mathbb{E}[T] \leq en^m \binom{m+k}{m}^{-1} \leq e\left(\frac{nm}{m+k}\right)^m$$

using $\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ for $k \geq 1$.

Tail Inequalities



Tail inequalities:

- The expectation can often be estimated easily.
- Would like to know the probability of deviating far from expectation, i.e., the "tails" of the distribution
- Tail inequalities give bounds on the tails given the expectation.

Markov's Inequality [Motwani and Raghavan, 1995]

A fundamental inequality from which many others are derived.

Theorem (Markov's Inequality)

Let X be a random variable assuming only non-negative values.
Then for all $t \in \mathbb{R}^+$,

$$\Pr(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

Number of bits that are flipped in a mutation step

- If $\mathbb{E}[X] = 1$, then $\Pr(X \geq 2) \leq \mathbb{E}[X]/2 = 1/2$.

Number of one-bits after initialisation

- If $\mathbb{E}[X] = n/2$, then $\Pr(X \geq (2/3)n) \leq \frac{\mathbb{E}[X]}{(2/3)n} = \frac{n/2}{(2/3)n} = 3/4$.

Chernoff Bounds

Let $X_1, X_2, \ldots X_n$ be independent Poisson trials each with probability p_i;
For $X = \sum_{i=1}^{n} X_i$ the expectation is $E(X) = \sum_{i=1}^{n} p_i$.

Theorem (Chernoff Bounds)

1. $\Pr(X \leq (1-\delta)\mathbb{E}[X]) \leq \exp\left(\frac{-\mathbb{E}[X]\delta^2}{2}\right)$ *for* $0 \leq \delta \leq 1$.
2. $\Pr(X > (1+\delta)\mathbb{E}[X]) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}[X]}$ *for* $\delta > 0$.

What is the probability that we have more than $(2/3)n$ one-bits at initialisation?

- $p_i = 1/2$, $\mathbb{E}[X] = n/2$,
  (we fix $\delta = 1/3 \rightarrow (1+\delta)\mathbb{E}[X] = (2/3)n$); then:
- $\Pr(X > (2/3)n) \leq \left(\frac{e^{1/3}}{(4/3)^{4/3}}\right)^{n/2} = c^{-n/2}$

Chernoff Bound Simple Application

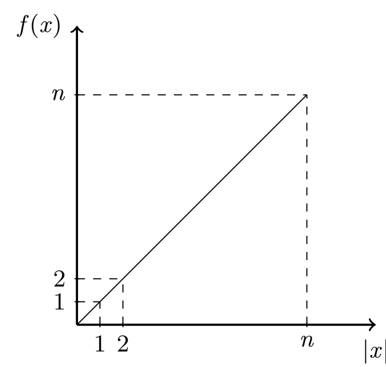
Bitstring of length $n = 100$

$\Pr(X_i) = 1/2$ and $E(X) = np = 100/2 = 50$.
What is the probability to have at least 75 1-bits?

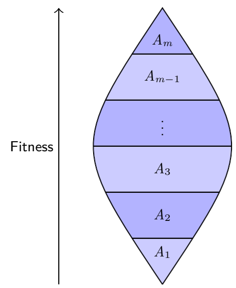
- Markov: $\Pr(X \geq 75) \leq \frac{50}{75} = \frac{2}{3}$
- Chernoff: $\Pr(X \geq (1+1/2)50) \leq \left(\frac{\sqrt{e}}{(3/2)^{3/2}}\right)^{50} < 0.0045$
- Truth: $\Pr(X \geq 75) = \sum_{i=75}^{100} \binom{100}{i} 2^{-100} < 0.000000282$

OneMax

$$\text{OneMax}(x) := x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i$$



Fitness-based Partitions



Definition

A tuple $(A_1, A_2, \dots, A_m)$ is an f-**based partition** of $f : \mathcal{X} \to \mathbb{R}$ if

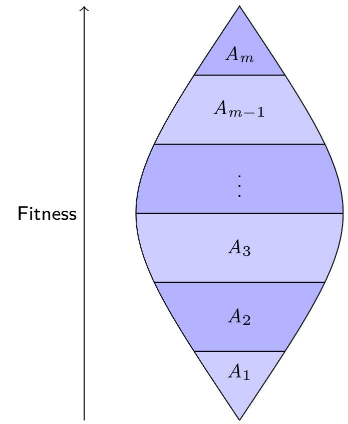
1. $A_1 \cup A_2 \cup \dots \cup A_m = \mathcal{X}$
2. $A_i \cap A_j = \emptyset$ for $i \neq j$
3. $f(A_1) < f(A_2) < \dots < f(A_m)$
4. $f(A_m) = \max_x f(x)$

Example

Partition of OneMax into $n+1$ levels

$$A_j := \{x \in \{0,1\}^n \mid \text{OneMax}(x) = j\}$$

Artificial Fitness Levels - Upper bounds



s_i : prob. of starting in A_i
u_i : prob. of jumping from A_i to any A_j, $i < j$.
T_i : Time to jump from A_i to any A_j, $i < j$.

Expected runtime

$$\begin{aligned}\mathbb{E}[T] &\leq \sum_{i=1}^{m-1} s_i \mathbb{E}\left[\sum_{j=i}^{m-1} T_j\right] \\ &= \sum_{i=1}^{m-1} s_i \sum_{j=i}^{m-1} \mathbb{E}[T_j] \\ &= \sum_{i=1}^{m-1} s_i \sum_{j=i}^{m-1} 1/u_j \leq \sum_{j=i}^{m-1} 1/u_j.\end{aligned}$$

(1+1) EA on OneMax

Theorem

The expected runtime of (1+1) EA on OneMax *is* $O(n \ln n)$.

Proof

- The current solution is in level A_j if it has j ones (hence $n-j$ zeroes).
- To reach a higher fitness level it is sufficient to flip a zero into a one and leave the other bits unchanged, which occurs with probability

$$u_j \geq (n-j)\frac{1}{n}\left(1 - \frac{1}{n}\right)^{n-1} \geq \frac{n-j}{en}$$

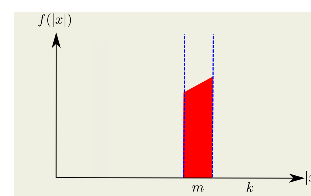
- Then by Artificial Fitness Levels

$$\mathbb{E}[T] \leq \sum_{j=0}^{m-1} 1/u_j \leq \sum_{j=0}^{n-1} \frac{en}{n-j} = en \sum_{i=1}^{n} \frac{1}{i} \leq en(\ln n + 1) = O(n \ln n)$$

Linear Unitation Block: Upper bound

Theorem

The expected runtime of the (1+1)-EA for a linear block is $O(n \ln((m+k)/k))$.

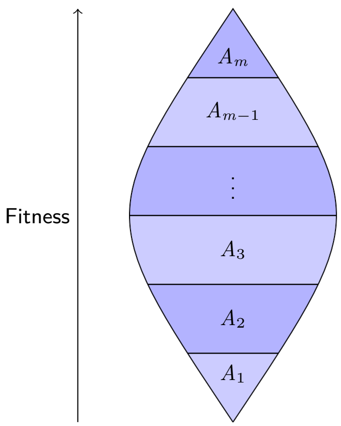


Proof

- Let $i := n - j$ be the number of 0-bits in block A_j
- The probability is $u_i \geq i \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right)^{n-1} \geq \left(\frac{i}{en}\right)$
- Hence, $\left(\frac{1}{u_i}\right) \leq \left(\frac{en}{i}\right)$
- Then (Artificial Fitness Levels):

$$E(T) \leq \sum_{i=k+1}^{k+m} \frac{en}{i} \leq en \sum_{i=k+1}^{k+m} \frac{1}{i} \leq en \left(\sum_{i=1}^{k+m} \frac{1}{i} - \sum_{i=1}^{k} \frac{1}{i}\right) \leq en \ln\left(\frac{m+k}{k}\right)$$

Artificial Fitness Levels - Lower bounds[2]



Theorem ([Sudholt, 2010])

Let

s_i : *prob. of starting in* A_i

u_i : *prob. of leaving* A_i, *and*

p_{ij} : *prob. of jumping from* A_i *to* A_j.

If there exists a $\chi \in [0,1)$ *st. for* $\forall i < j$

$$p_{ij} \geq \chi \sum_{k=j}^{m-1} p_{ik},$$

then

$$\mathbb{E}[T] \geq \chi \sum_{i=1}^{m-1} s_i \sum_{j=i}^{m-1} \frac{1}{u_j}.$$

[2] A different version of the theorem is presented.

(1+1) EA lower bound for OneMax

Fitness level $A_i := \{x \in \{0,1\}^n \mid \text{OneMax}(x) = i\}$

$$x = \overbrace{1111111111111111111111111}^{i}\overbrace{00000000000000000000}^{n-i} \in A_i$$

Probability p_{ij} of jumping to level $j > i$ and beyond

$$p_{ij} \geq \binom{n-i}{j-i}\left(\frac{1}{n}\right)^{j-i}\left(1-\frac{1}{n}\right)^{n-(j-i)}$$

$$\sum_{k=j}^{n-1} p_{ik} \leq \binom{n-i}{j-i}\left(\frac{1}{n}\right)^{j-i}$$

Hence, for $\chi = 1/e$

$$p_{ij} \geq \left(1-\frac{1}{n}\right)^{n-(j-i)} \sum_{k=j}^{n-1} p_{ik} \geq \chi \sum_{k=j}^{n-1} p_{ik}$$

(1+1) EA lower bound for OneMax

Theorem

The expected runtime of the (1+1) EA for OneMax *is* $\Omega(n \ln n)$.

Probability u_i of any improvement

$$u_i \leq \frac{n-i}{n}$$

We have already seen that $\sum_{i=(2/3)n}^{n} s_i \leq 3/4$, hence

$$\mathbb{E}[T] \geq \left(\frac{1}{e}\right)\sum_{i=0}^{n-1} s_i \sum_{j=i}^{n-1}\frac{1}{u_j}$$

$$> \left(\frac{1}{e}\right)\left(\sum_{i=0}^{(2/3)n} s_i\right)\left(\sum_{j=(2/3)n}^{n-1}\frac{1}{u_j}\right)$$

$$> \left(\frac{n}{e}\right)(1-3/4)\left(\sum_{j=1}^{n/3}\frac{1}{j}\right) = \Omega(n \ln n)$$

Linear Block: Lower Bound

Theorem

The expected runtime to finish a linear block of length m *starting at* $k+m$ *0-bits is* $\Omega(n \ln((m+k)/k))$.

For $0 \leq i \leq m$, define $A_i := \{x : n - |x| = k + m - i\}$. Note that

$$p_{ij} = \binom{k+m-i}{j-i}\left(\frac{1}{n}\right)^{j-i}\left(1-\frac{1}{n}\right)^{n-(j-i)}$$

$$\sum_{k=j}^{m-1} p_{ik} \leq \binom{k+m-i}{j-i}\left(\frac{1}{n}\right)^{j-i}$$

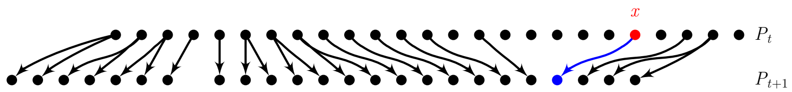
Therefore,

$$p_{ij} \geq \left(1-\frac{1}{n}\right)^{n-(j-i)}\sum_{k=j}^{m-1} p_{ik} \geq \left(\frac{1}{e}\right)\sum_{k=j}^{m-1} p_{ik}$$

and assuming that $s_0 = 1$, we get

$$\mathbb{E}[T] \geq \left(\frac{1}{e}\right)\sum_{i=0}^{m-1}\frac{1}{u_i} \geq \left(\frac{1}{e}\right)\sum_{i=0}^{m-1}\frac{n}{m+k-i} = \left(\frac{n}{e}\right)\left(\sum_{i=1}^{m+k}\frac{1}{i} - \sum_{i=1}^{k}\frac{1}{i}\right)$$

Advanced: Fitness levels for non-elitist populations

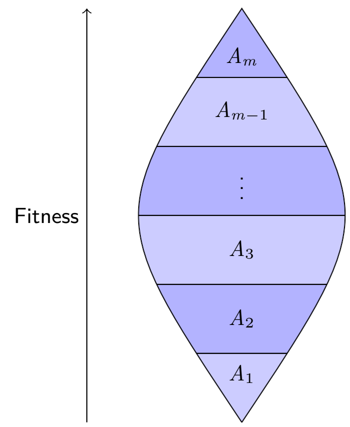


for $t = 0, 1, 2, \ldots$ until termination condition **do**
  for $i = 1$ to λ **do**
    Sample i-th parent x according to $p_{\mathrm{sel}}(P_t, f)$
    Sample i-th offspring $P_{t+1}(i)$ according to $p_{\mathrm{var}}(x)$
  end for
end for

A general algorithmic scheme for non-elitistic EAs

- $f : \mathcal{X} \to \mathbb{R}$ fitness function over arbitrary finite search space $\mathcal{X}$
- p_{sel} selection mechanism (e.g. (μ, λ)-selection)
- p_{var} variation operator (e.g. mutation)

Advanced: Fitness Levels for non-Elitist Populations[3]



Theorem ([Lehre, 2011a])

If exists $\delta, \gamma_*, s_1, \ldots, s_m, s_*, p_0 \in (0,1)$ *st.*

(C1) $p_{\mathrm{var}}\left(y \in A_j^+ \mid x \in A_j\right) \geq s_j \geq s_*$
upgrade probability s_j

(C2) $p_{\mathrm{var}}\left(y \in A_j \cup A_j^+ \mid x \in A_j\right) \geq p_0$
resting probability p_0

(C3) $\beta(\gamma) > \gamma(1+\delta)/p_0$ *for all* $\gamma < \gamma_*$
"high" selective pressure

(C4) $\lambda > c' \ln(m/s_*)$ *for some const.* c'
"large" population size

then for a constant $c > 0$

$$\mathbb{E}[T] \leq c\left(m\lambda^2 + \sum_{j=1}^{m-1} \frac{1}{s_j}\right)$$

[3] See this year's GECCO theory track for an improved version! [Dang and Lehre, 2014].

Example: (μ, λ) EA on LeadingOnes

$$x = \overbrace{1111111111111111}^{\text{Leading 1-bits.}}0\overbrace{***************}^{\text{Random bitstring.}}.$$

First 0-bit.

$$\textsc{LeadingOnes}(x) = \sum_{i=1}^{n} \prod_{j=1}^{i} x_i$$

Theorem

If $\lambda/\mu > e$ *and* $\lambda > c \ln n$*, then the expected runtime of* (μ,λ) *EA on* LeadingOnes *is* $O(n\lambda^2 + n^2)$.

Measuring Selective Pressure

Definition

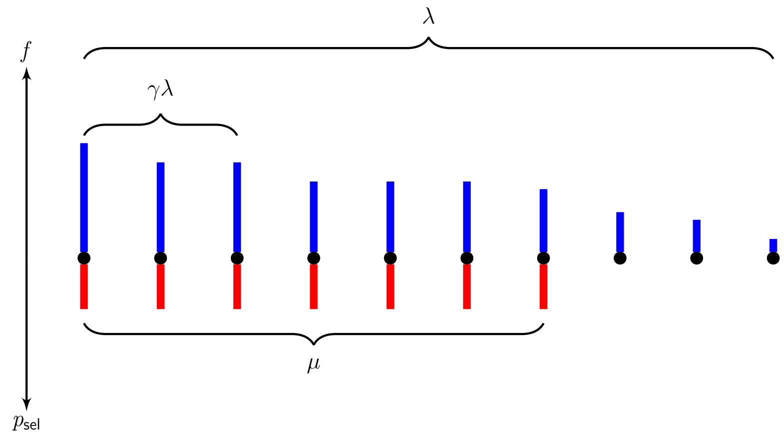
Let $x^{(1)}, x^{(2)}, \ldots, x^{(\lambda)}$ be the individuals in a population $P \in \mathcal{X}^\lambda$, sorted according to a fitness function $f : \mathcal{X} \to \mathbb{R}$, i.e.

$$f\left(x^{(1)}\right) \geq f\left(x^{(2)}\right) \geq \cdots \geq f\left(x^{(\lambda)}\right).$$

For any $\gamma \in (0,1)$, the **cumulative selection probability** of p_{sel} is

$$\beta(\gamma) := \Pr\left(f(y) \geq f\left(x^{(\gamma\lambda)}\right) \mid y \text{ is sampled from } p_{\mathrm{sel}}(P, f)\right)$$

Cumulative Selection Prob. - Example (μ, λ)-selection



$$\beta(\gamma) = \Pr\left(f(y) \geq f\left(x^{(\gamma\lambda)}\right) \mid y \text{ is sampled from } p_{\text{sel}}(P, f) \right)$$
$$\geq \frac{\gamma\lambda}{\mu} \quad \text{if } \gamma\lambda \leq \mu$$

Example Application[4]

(μ,λ) EA with bit-wise mutation rate χ/n on LeadingOnes

Partition of fitness function into $m := n + 1$ levels

$$A_j := \{x \in \{0,1\}^n \mid x_1 = x_2 = \cdots = x_{j-1} = 1 \wedge x_j = 0\}$$

If $\lambda/\mu > e^\chi$ and $\lambda > c'' \ln(n)$ then

(C1) $p_{\text{var}}\left(y \in A_j^+ \mid x \in A_j\right) = \Omega(1/n) \quad =: \; s_j =: s_*$

(C2) $p_{\text{var}}\left(y \in A_j \cup A_j^+ \mid x \in A_j\right) \approx e^{-\chi} \quad =: \; p_0$

(C3) $\beta(\gamma) \geq \gamma\lambda/\mu > \gamma e^\chi \quad = \; \gamma/p_0$

(C4) $\lambda > c'' \ln(n) \quad > \; c \ln(m/s^*)$

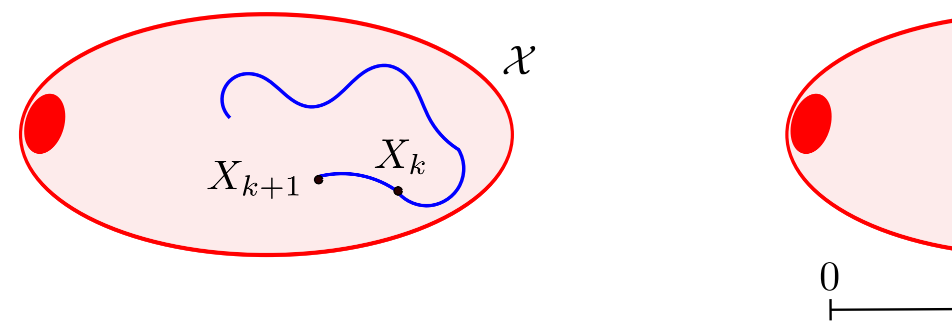
then $\mathbb{E}[T] = O(m\lambda^2 + \sum_{j=1}^{m} s_j^{-1}) = O(n\lambda^2 + n^2)$

[4] Calculations on this slide are approximate. See [Lehre, 2011a] for exact calculations.

Artificial Fitness Levels: Conclusions

- It's a powerful general method to obtain (often) tight upper bounds on the runtime of simple EAs;
- For offspring populations tight bounds can often be achieved with the general method;
- There exists a variant of artificial fitness levels for populations [Lehre, 2011b].

What is Drift[5] Analysis?

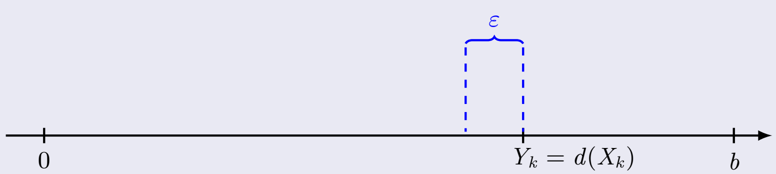


- Prediction of the long term behaviour of a process X
  - hitting time, stability, occupancy time etc.

  from properties of Δ.

[5] NB! (Stochastic) drift is a different concept than *genetic drift* in population genetics.

Additive Drift Theorem



(C1+) $\forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > 0\right] \le -\varepsilon$

(C1−) $\forall k \quad \mathbb{E}\left[Y_{k+1} - Y_k \mid Y_k > 0\right] \ge -\varepsilon$

Theorem ([He and Yao, 2001, Jägersküpper, 2007, Jägersküpper, 2008])

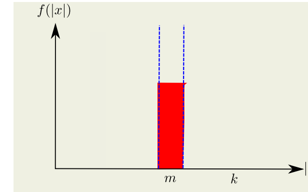
Given a stochastic process $Y_1, Y_2, \dots$ over an interval $[0, b] \subset \mathbb{R}$. Define $T := \min\{k \ge 0 \mid Y_k = 0\}$, and assume $\mathbb{E}[T] < \infty$.

- *If (C1+) holds for an $\varepsilon > 0$, then $\mathbb{E}[T \mid Y_0] \le b/\varepsilon$.*
- *If (C1−) holds for an $\varepsilon > 0$, then $\mathbb{E}[T \mid Y_0] \ge Y_0/\varepsilon$.*

Plateau Block Function: Upper Bound

Let $k > n/2 + \epsilon n$.

$$PlateauBlock_\ell(|x|) = \begin{cases} a & \text{if } k \le n - |x| \le k + m \\ 0 & \text{otherwise.} \end{cases}$$



Theorem

The expected time for the (1+1)-EA to optimise the Plateau function is $O(m)$.

Proof

Let X_t be the number of 0-bits at time t. Then the drift is

$$E(\Delta(t) \ge \frac{X_t}{n} - \frac{n - X_t}{n} = \frac{2X_t}{n} - 1 \ge \frac{2k}{n} - 1$$

Hence, by drift analysis

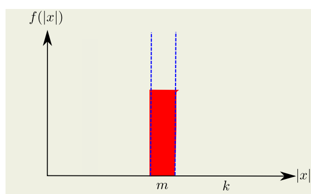
$$E[T] \le \frac{m}{(2k)/n - 1} = \frac{mn}{2k - n} = O(m)$$

where the last equality holds as long as $k > n/2 + \epsilon n$

Plateau Block Function: Lower Bound

Let $k > n/2 + \epsilon n$.

$$PlateauBlock_\ell(|x|) = \begin{cases} a & \text{if } k \le n - |x| \le k + m \\ 0 & \text{otherwise.} \end{cases}$$



Theorem

The expected time for the (1+1)-EA to optimise the Plateau function is $\Theta(m)$.

Proof

Let X_t be the number of 0-bits at time t. Then the drift is

$$E(\Delta(t) = \frac{X_t}{n} - \frac{n - X_t}{n} = \frac{2X_t}{n} - 1 \le \frac{2(m+k)}{n} - 1$$

Hence, by drift analysis

$$E[T] \ge \frac{m}{2(m+k)/n - 1} = \frac{mn}{2(m+k) - n} = \Omega(m)$$

where the last equality holds as long as $k > n/2 + \epsilon n$

Drift Analysis for OneMax

Lets calculate the runtime of the (1+1)-EA using the additive Drift Theorem.

1. Let $d(X_t) = i$ where i is the number of zeroes in the bitstring;
2. Note that $d(X_t) - d(X_{t+1}) \ge 0$ for all t;
3. The distance decreases by 1 as long as a 0 is flipped and the ones remain unchanged:

$$E(\Delta(t)) = E[d(X_t) - d(X_{t+1}) \mid X_t] \ge 1 \cdot \frac{i}{n}\left(1 - \frac{1}{n}\right)^{n-1} \ge \frac{i}{en} \ge \frac{1}{en} =: \delta$$

4. The expected initial distance is $E(d(X_0)) = n/2$

The expected runtime is (i.e. Eq. (**??**)):

$$E(T \mid d(X_0) > 0) \le \frac{E[(d(X_0)]}{\delta} \le \frac{n/2}{1/(en)} = e/2 \cdot n^2 = O(n^2)$$

We need a different distance function!

Drift Analysis for ONEMAX

1. Let $d(X_t) = \ln(i+1)$ where i is the number of zeroes in the bitstring;
2. For $x \geq 1$, it holds that $\ln(1+1/x) \geq 1/x - 1/(2x^2) \geq 1/(2x)$.
3. The distance decreases as long as a 0 is flipped and the ones remain unchanged

$$\begin{aligned}\mathbb{E}[\Delta(t)] &= \mathbb{E}[d(X_t) - d(X_{t+1}) \mid d(X_t) = i \geq 1] \\ &\geq \frac{i}{en}(\ln(i+1) - \ln(i)) = \frac{i}{en}\ln\left(1+\frac{1}{i}\right) \\ &\geq \frac{i}{en}\frac{1}{2i} = \frac{1}{2en} =: \delta.\end{aligned}$$

4. The initial distance is $d(X_0) \leq \ln(n+1)$

The expected runtime is (i.e. Eq. (**??**)):

$$E(T \mid d(X_0) > 0) \leq \frac{d(X_0)}{\delta} \leq \frac{\ln(n+1)}{1/(2en)} = O(n \ln n)$$

If the amount of progress depends on the distance from the optimum we need to use a logarithmic distance!

Multiplicative Drift Theorem

Theorem (Multiplicative Drift, [Doerr et al., 2010a])

Let $\{X_t\}_{t\in\mathbb{N}_0}$ be random variables describing a Markov process over a finite state space $S \subseteq \mathbb{R}$. Let T be the random variable that denotes the earliest point in time $t \in \mathbb{N}_0$ such that $X_t = 0$.
If there exist $\delta, c_{\min}, c_{\max} > 0$ such that

1. $E[X_t - X_{t+1} \mid X_t] \geq \delta X_t$ *and*
2. $c_{\min} \leq X_t \leq c_{\max}$,

for all $t < T$, then

$$E[T] \leq \frac{2}{\delta} \cdot \ln\left(1 + \frac{c_{\max}}{c_{\min}}\right)$$

(1+1)-EA Analysis for ONEMAX

Theorem

The expected time for the (1+1)-EA to optimise ONEMAX is $O(n \ln n)$

Proof

- Distance: let X_t be the number of zeroes in step t;
- $E[X_{t+1}|X_t] \leq X_t - 1 \cdot \frac{X_t}{en} = X_t \cdot \left(1 - \frac{1}{en}\right)$
- $E[X_t - X_{t+1}|X_t = i] \geq X_t - X_t \cdot \left(1 - \frac{1}{en}\right) = X_t/(en)$ $(\delta = 1/(en))$
- $1 = c_{\min} \leq X_t \leq c_{\max} = n$

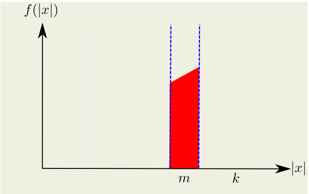
Hence,

$$E[T] \leq \frac{2}{\delta} \cdot \ln\left(1 + \frac{c_{\max}}{c_{\min}}\right) = 2en\ln(1+n) = O(n \ln n)$$

Linear Unitation Block: Upper Bound

Theorem

The expected time for the (1+1)-EA to optimise the Linear Unitation Block is $O(n \ln((m+k)/k))$



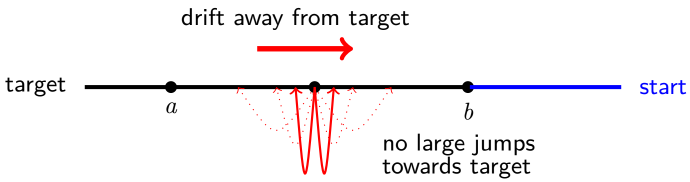
Proof

- Distance: let i be the number of zeroes;
- $E[X_{t+1}|X_t] \leq X_t - 1 \cdot \frac{X_t}{en} = X_t \left(1 - \frac{1}{en}\right)$
- $E[X_t - X_{t+1}|X_t] \geq X_t - X_t\left(1 - \frac{1}{en}\right) = \frac{1}{en}X_t$ $(\delta := \frac{1}{en})$
- $k = c_{\min} \leq X_t \leq c_{\max} = m + k$

Hence,

$$E[T] \leq \frac{2}{\delta} \cdot \ln\left(1 + \frac{c_{\max}}{c_{\min}}\right) = 2en\ln(1+(m+k)/k) = O(n\ln((m+k)/k))$$

Simplified Drift Theorem



Theorem (Simplified Negative-Drift Theorem, [Oliveto and Witt, 2011])

Suppose there exist three constants δ,ϵ,r *such that for all* $t \geq 0$:

1. $E(\Delta_t(i)) \geq \epsilon$ *for* $a < i < b$,
2. $\text{Prob}(|\Delta_t(i)| = j) \leq \frac{1}{(1+\delta)^{j-r}}$ *for* $i > a$ *and* $j \geq 1$.

Then

$$\text{Prob}(T^* \leq 2^{c^*(b-a)}) = 2^{-\Omega(b-a)}$$

.

Needle in a Haystack

Theorem (Oliveto,Witt, Algorithmica 2011)

Let $\eta > 0$ *be constant. Then there is a constant* $c > 0$ *such that with probability* $1 - 2^{-\Omega(n)}$ *the (1+1)-EA on* Needle *creates only search points with at most* $n/2 + \eta n$ *ones in* 2^{cn} *steps.*

Proof Idea

- By Chernoff bounds the probability that the initial bit string has less than $n/2 - \gamma n$ zeroes is $e^{-\Omega(n)}$.
- we set $b := n/2 - \gamma n$ and $a := n/2 - 2\gamma n$ where $\gamma := \eta/2$;

Proof of Condition 1

$$E(\Delta(i)) = \frac{n-i}{n} - \frac{i}{n} = \frac{n-2i}{n} \geq 2\gamma = \epsilon$$

Proof of Condition 2

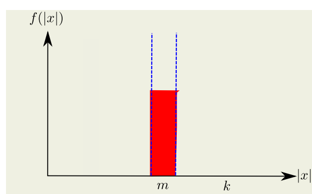
$$\Pr(|\Delta(i)| \geq j) \leq \binom{n}{j}\left(\frac{1}{n}\right)^j \leq \left(\frac{n^j}{j!}\right)\left(\frac{1}{n}\right)^j \leq \frac{1}{j!} \leq \left(\frac{1}{2}\right)^{j-1}$$

This proves Condition 2 by setting $\delta = r = 1$.

Plateau Block Function: Lower Bound

Let $k + m < (1/2 - \epsilon)n$.

$$PlateauBlock_r(|x|) = \begin{cases} a & \text{if } k \leq n - |x| \leq k + m \\ 0 & \text{otherwise.} \end{cases}$$



Theorem

The time for the (1+1)-EA to optimise $PlateauBlock_r$ *is at least* $2^{\Omega(m)}$ *with probability at least* $1 - 2^{-\Omega(m)}$.

Proof

Let X_t be the number of 0-bits at time t.

$$E(\Delta(t)) = \frac{n - X_t}{n} - \frac{X_t}{n} = 1 - \frac{2X_t}{n} \geq \frac{n}{n} - \frac{2(k+m)}{n} = \frac{n - 2(k+m)}{n}$$

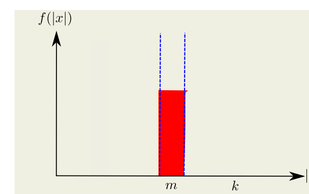
If $2(k+m) < n(1-\epsilon)$ by the simplified drift theorem

$$P(T < 2^{cm}) = 2^{-\Omega(m)}$$

Plateau Block Function: Upper Bound

Theorem

The expected time for the (1+1)-EA to optimise $PlateauBlock_r$ *is at most* $e^{O(m)}$.



Proof

We calculate the probability p of m consecutive steps across the plateau

$$\prod_{i=m+1}^{k+m} p_i \geq \prod_{i=1}^{m} \frac{k+i}{en} \geq \left(\frac{1}{en}\right)^m \frac{(k+m)!}{k!} \geq \left(\frac{1}{en}\right)^m \left(\frac{k+m}{e}\right)^m = \left(\frac{k+m}{e^2 n}\right)^m$$

where

$$\frac{(k+m)!}{k!} = m! \cdot \frac{(k+m)!}{m!k!} = m!\binom{k+m}{m} \geq \left(\frac{m}{e}\right)^m \left(\frac{k+m}{m}\right)^m = \left(\frac{k+m}{e}\right)^m$$

Hence,

$$\mathbb{E}[T] \leq m \cdot 1/p = m\left(\frac{e^2 n}{k+m}\right)^m$$

Some history[6]

Origins

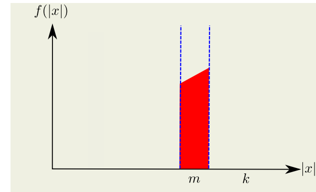
- Stability of equilibria in ODEs (Lyapunov, 1892)
- Stability of Markov Chains (see eg [Meyn and Tweedie, 1993])
- 1982 paper by Hajek [Hajek, 1982]
  - Simulated annealing (1988) [Sasaki and Hajek, 1988]

Drift Analysis of Evolutionary Algorithms

- Introduced to EC in 2001 by He and Yao [He and Yao, 2001, He and Yao, 2004] (additive drift)
  - (1+1) EA on linear functions: $O(n \ln n)$ [He and Yao, 2001]
  - (1+1) EA on maximum matching by Giel and Wegener [Giel and Wegener, 2003]
- Simplified drift in 2008 by Oliveto and Witt [Oliveto and Witt, 2011]
- Multiplicative drift by Doerr et al [Doerr et al., 2010b]
  - (1+1) EA on linear functions: $en \ln(n) + O(n)$ [Witt, 2012]
- Variable drift by Johannsen [Johannsen, 2010] and Mitavskiy et al. [Mitavskiy et al., 2009]
- Population drift by Lehre [Lehre, 2011c]

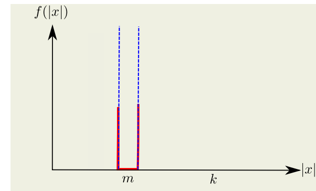
[6] More on drift in GECCO 2012 tutorial by Lehre http://www.cs.nott.ac.uk/~pkl/drift

Summary - (1+1) EA on Functions of Unitation



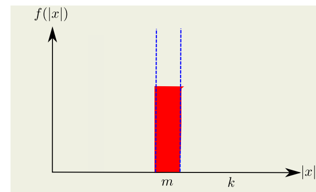
Linear blocks

- $\Theta\left(n \ln\left(\frac{m+k}{k}\right)\right)$



Gap blocks

- $O\left(\left(\frac{nm}{m+k}\right)^m\right)$
- $\Omega\left(\left(\frac{nm}{e(m+k)}\right)^m\right)$



Plateau blocks

- $e^{O(m)}$ if $k < n(1/2 - \varepsilon)$
- $\Theta(m)$ if $k > n(1/2 + \varepsilon)$

Final Overview

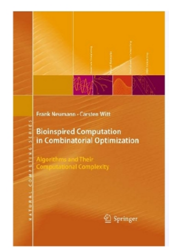
Overview

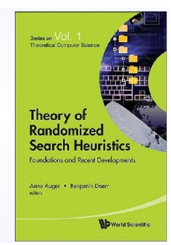
- Tail Inequalities
- Artificial Fitness Levels
- Drift Analysis

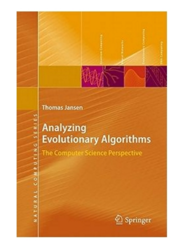
Other Techniques (Not covered)

- Family Trees [Witt, 2006]
- Gambler's Ruin & Martingales [Jansen and Wegener, 2001]
- Probability Generating Functions [Doerr et al., 2011]
- Branching Processes [Lehre and Yao, 2012]
- ...

Further Reading







Thank you!

Acknowledgements

http://www.project-sage.eu

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 618091 (SAGE).

This work was supported in part by the Engineering and Physical Sciences Research Council under grant no EP/M004252/1.

EPSRC
Engineering and Physical Sciences Research Council

References I

Bäck, T. (1993).
Optimal mutation rates in genetic search.
In *In Proceedings of the Fifth International Conference on Genetic Algorithms (ICGA)*, pages 2–8.

Dang, D.-C. and Lehre, P. K. (2014).
Upper bounds on the expected runtime of non-elitist populations from fitness-levels.
To appear in Proceedings of The Genetic and Evolutionary Computation Conference (GECCO) 2014, Vancouver, Canada.

Doerr, B., Fouz, M., and Witt, C. (2011).
Sharp bounds by probability-generating functions and variable drift.
In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, GECCO '11, pages 2083–2090, New York, NY, USA. ACM.

Doerr, B., Johannsen, D., and Winzen, C. (2010a).
Multiplicative drift analysis.
In *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, GECCO '10, pages 1449–1456. ACM.

Doerr, B., Johannsen, D., and Winzen, C. (2010b).
Multiplicative drift analysis.
In *GECCO '10: Proceedings of the 12th annual conference on Genetic and evolutionary computation*, pages 1449–1456, New York, NY, USA. ACM.

Droste, S., Jansen, T., and Wegener, I. (1998).
A rigorous complexity analysis of the (1 + 1) evolutionary algorithm for separable functions with boolean inputs.
Evolutionary Computation, 6(2):185–196.

References II

Giel, O. and Wegener, I. (2003).
Evolutionary algorithms and the maximum matching problem.
In *Proceedings of the 20th Annual Symposium on Theoretical Aspects of Computer Science (STACS 2003)*, pages 415–426.

Goldberg, D. E. (1989).
Genetic Algorithms for Search, Optimization, and Machine Learning.
Addison-Wesley.

Hajek, B. (1982).
Hitting-time and occupation-time bounds implied by drift analysis with applications.
Advances in Applied Probability, 13(3):502–525.

He, J. and Yao, X. (2001).
Drift analysis and average time complexity of evolutionary algorithms.
Artificial Intelligence, 127(1):57–85.

He, J. and Yao, X. (2004).
A study of drift analysis for estimating computation time of evolutionary algorithms.
Natural Computing: an international journal, 3(1):21–35.

Holland, J. H. (1992).
Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence.
The MIT Press.

References III

Jägersküpper, J. (2007).
Algorithmic analysis of a basic evolutionary algorithm for continuous optimization.
Theoretical Computer Science, 379(3):329–347.

Jägersküpper, J. (2008).
A blend of markov-chain and drift analysis.
In *PPSN*, pages 41–51.

Jansen, T. and Wegener, I. (2001).
Evolutionary algorithms - how to cope with plateaus of constant fitness and when to reject strings of the same fitness.
IEEE Trans. Evolutionary Computation, 5(6):589–599.

Johannsen, D. (2010).
Random combinatorial structures and randomized search heuristics.
PhD thesis, Universität des Saarlandes.

Lehre, P. K. (2011a).
Fitness-levels for non-elitist populations.
In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, GECCO '11, pages 2075–2082. ACM.

Lehre, P. K. (2011b).
Fitness-levels for non-elitist populations.
In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, (GECCO 2011), pages 2075–2082, New York, NY, USA. ACM.

References IV

Lehre, P. K. (2011c).
Negative drift in populations.
In *Proceedings of Parallel Problem Solving from Nature - (PPSN XI)*, volume 6238 of *LNCS*, pages 244–253. Springer Berlin / Heidelberg.

Lehre, P. K. and Yao, X. (2012).
On the impact of mutation-selection balance on the runtime of evolutionary algorithms.
IEEE Transactions on Evolutionary Computation, 16(2):225–241.

Meyn, S. P. and Tweedie, R. L. (1993).
Markov Chains and Stochastic Stability.
Springer-Verlag.

Mitavskiy, B., Rowe, J. E., and Cannings, C. (2009).
Theoretical analysis of local search strategies to optimize network communication subject to preserving the total number of links.
International Journal of Intelligent Computing and Cybernetics, 2(2):243–284.

Motwani, R. and Raghavan, P. (1995).
Randomized Algorithms.
Cambridge University Press.

Oliveto, P. S. and Witt, C. (2011).
Simplified drift analysis for proving lower bounds inevolutionary computation.
Algorithmica, 59(3):369–386.

References V

Reeves, C. R. and Rowe, J. E. (2002).
Genetic Algorithms: Principles and Perspectives: A Guide to GA Theory.
Kluwer Academic Publishers, Norwell, MA, USA.

Rudolph, G. (1998).
Finite Markov chain results in evolutionary computation: A tour d'horizon.
Fundamenta Informaticae, 35(1–4):67–89.

Sasaki, G. H. and Hajek, B. (1988).
The time complexity of maximum matching by simulated annealing.
Journal of the Association for Computing Machinery, 35(2):387–403.

Sudholt, D. (2010).
General lower bounds for the running time of evolutionary algorithms.
In *PPSN (1)*, pages 124–133.

Witt, C. (2006).
Runtime analysis of the $(\mu$+1) ea on simple pseudo-boolean functions evolutionary computation.
In *GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pages 651–658, New York, NY, USA. ACM Press.

Witt, C. (2012).
Optimizing linear functions with randomized search heuristics - the robustness of mutation.
In Dürr, C. and Wilke, T., editors, *29th International Symposium on Theoretical Aspects of Computer Science (STACS 2012)*, volume 14 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 420–431, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

References VI

Wolpert, D. and Macready, W. G. (1997).
No free lunch theorems for optimization.
IEEE Trans. Evolutionary Computation, 1(1):67–82.