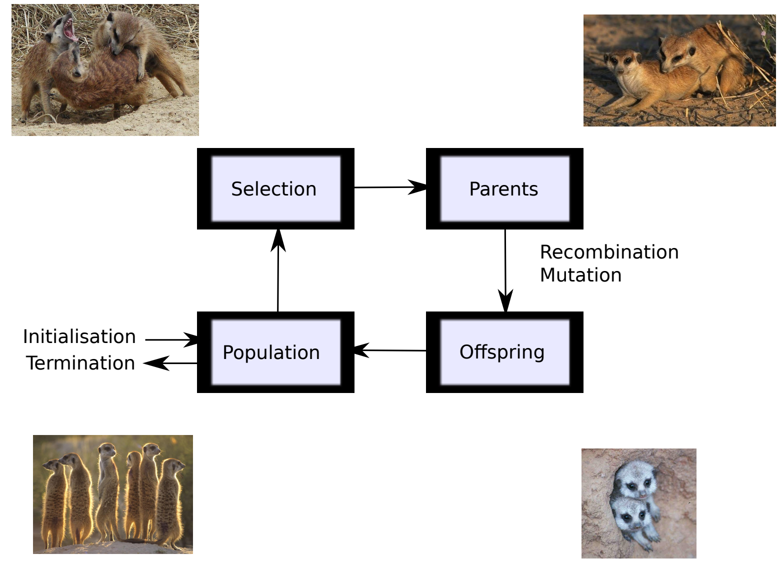
Runtime Analysis of Population-based Evolutionary Algorithms[1]

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

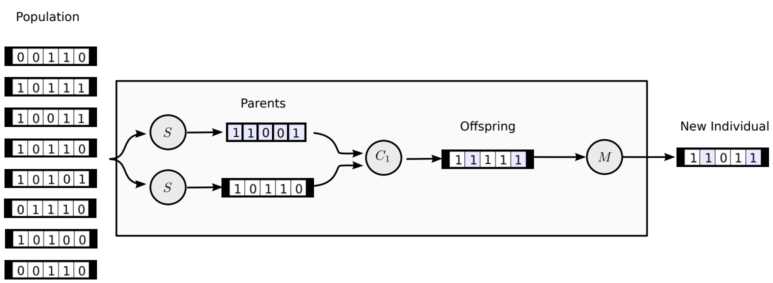
Pietro S. Oliveto
University of Sheffield
Sheffield S1 4DP, UK
`P.Oliveto@sheffield.ac.uk`

[1]The latest version of these slides are available at `http://www.cs.nott.ac.uk/~pszpl/gecco2016`

Evolutionary Algorithms



General Scheme for Evolutionary Algorithms[2]



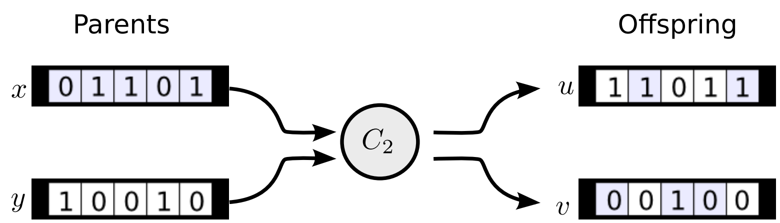
1: **initialise** a population P_0 of λ individuals uniformly at random.
2: **for** $t = 0, 1, 2, \ldots$ until termination condition **do**
3: **evaluate** the individuals in population P_t.
4: **for** $i = 1$ to λ **do**
5: **select** two parents from population P_t.
6: **recombine** the two parents.
7: **mutate** the offspring and add it to population P_{t+1}.

[2]Pseudo-code adapted from Eiben and Smith [2003].

Bitwise Mutation

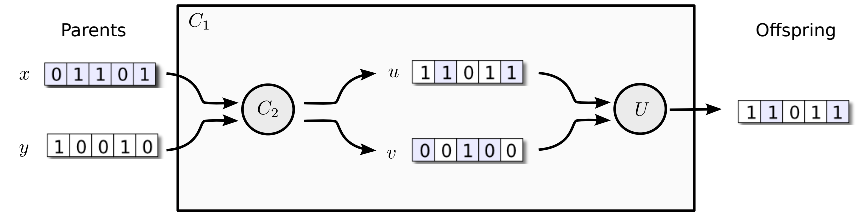
for $i = 1$ **to** n **do**
  with probability χ/n
    $x'_i := 1 - x_i$
  otherwise
    $x'_i := x_i$
return x'

Uniform Crossover - Two Offspring



for $i = 1$ **to** n **do**
  with probability $1/2$
    $u_i := x_i$ and $v_i := y_i$
  otherwise
    $u_i := y_i$ and $v_i := x_i$
return u and v.

Uniform Crossover - One Offspring



for $i = 1$ **to** n **do**
  with probability $1/2$
    $u_i := x_i$ and $v_i := y_i$
  otherwise
    $u_i := y_i$ and $v_i := x_i$
return u or v with equal probability.

Selection - Linear Ranking Goldberg and Deb [1991]

$\alpha : [0,1] \to [0,\infty)$ a *ranking function* if

$$\int_0^1 \alpha(y)dy = 1$$

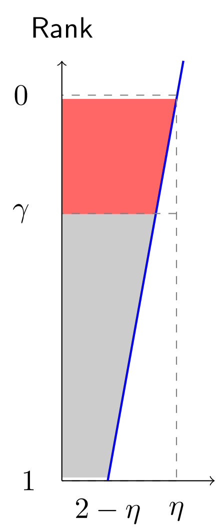
Prob. of selecting individual with rank $\leq \gamma$ is

$$\beta(\gamma) := \int_0^\gamma \alpha(y)dy$$

Linear ranking selection is obtained for

$$\alpha(\gamma) := \eta - 2(\eta - 1)\gamma,$$

where $\eta \in (1,2)$ specifies **selection pressure**.



Example - Linear Ranking Selection

for $t = 0$ to ∞ **do**
  Sort current population P_t according to fitness f, st
  $f(P_t(1)) \geq f(P_t(2)) \geq \cdots \geq f(P_t(\lambda))$.
  for $i = 1$ to λ **do**
    (Selection)
    Sample r in $\{1, ..., \lambda\}$ st. $\Pr(r \leq \gamma\lambda) = \beta(\gamma)$.
    $P_{t+1}(i) := P_t(r)$.
    (Mutation)
    Flip each bit position in $P_{t+1}(i)$ with prob. χ/n.

Example - Linear Ranking Selection

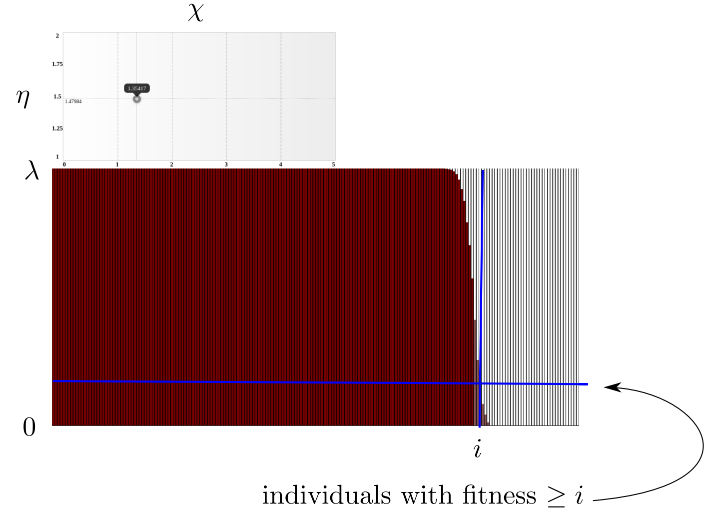
for $t = 0$ to ∞ **do**
  Sort current population P_t according to fitness f, st
  $f(P_t(1)) \geq f(P_t(2)) \geq \cdots \geq f(P_t(\lambda))$.
  for $i = 1$ to λ **do**
    (Selection)
    Sample r in $\{1, ..., \lambda\}$ st. $\Pr(r \leq \gamma\lambda) = \beta(\gamma)$.
    $P_{t+1}(i) := P_t(r)$.
    (Mutation)
    Flip each bit position in $P_{t+1}(i)$ with prob. χ/n.

Problem

- Is it possible to predict the behaviour of this and other EAs?
- Can we parameterise the EA so that it optimises f efficiently, e.g.

$$\textsc{OneMax}(x) := \sum_{i=1}^{n} x_i$$

Interactive Visualisation of Evolutionary Algorithm



http://www.cs.nott.ac.uk/~pszpl/ea/

Evolutionary Algorithms are Algorithms

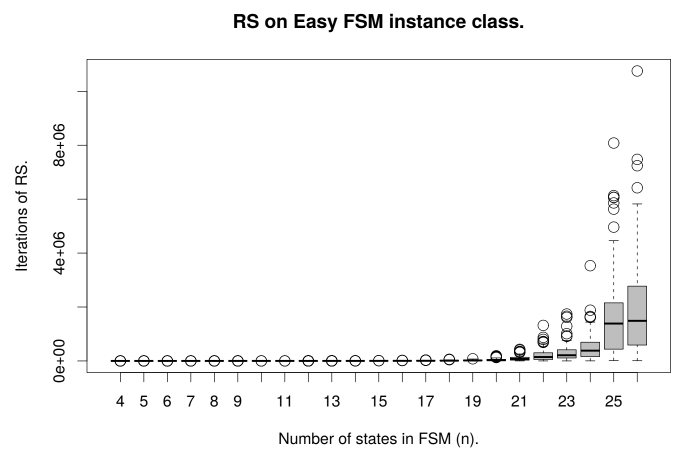
Criteria for evaluating algorithms

1. Correctness
   - Does the algorithm always give the correct output?
2. Computational Complexity
   - How much computational resources does the algorithm require to solve the problem?

Same criteria also applicable to evolutionary algorithms

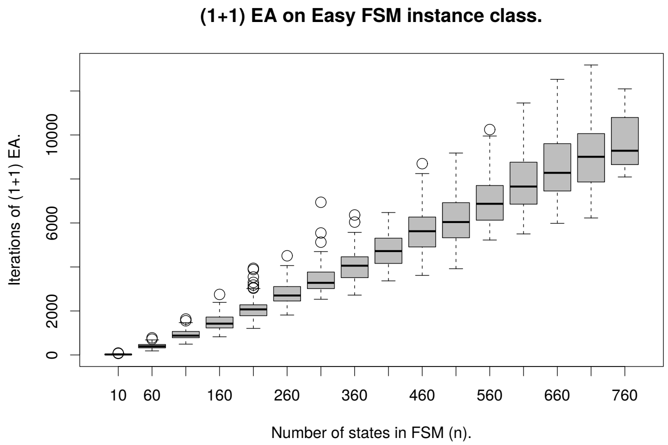
1. Correctness.
   - Discover global optimum in finite time?
2. Computational Complexity.
   - Time (number of function evaluations) most relevant computational resource.

Runtime as a function of problem size



- Exponential $\Longrightarrow$ Algorithm impractical on problem.

Runtime as a function of problem size



- Exponential $\Longrightarrow$ Algorithm impractical on problem.
- Polynomial $\Longrightarrow$ Possibly efficient algorithm.

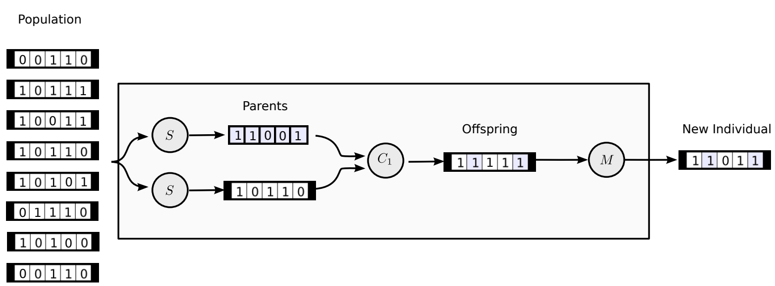
Outline

Introduction
Runtime Analysis

Upper bounds
The Level Based Theorem
Examples
Mutation and Selection
Mutation, Crossover and Selection

Lower Bounds
Negative Drift Theorem for Populations
Mutation-Selection Balance
Negative Drift with Crossover

A Model of Population-based EAs

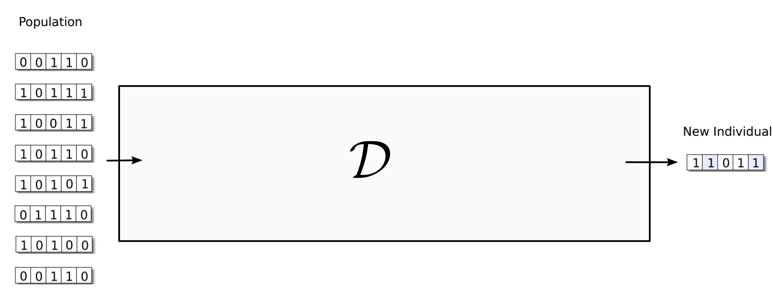


Wide range of evolutionary algorithms...

- selection mechanisms (ranking selection, (μ, λ)-selection, tournament selection, ...)
- fitness models (deterministic, stochastic, dynamic, partial, ...)
- variation operators
- search spaces (e.g. bitstrings, permutations, ...)

We will describe many of these with a general mathematical model.

A Model of Population-based EAs



Require: Search space $\mathcal{X}$ and random operator $\mathcal{D} : \mathcal{X}^\lambda \to \mathcal{X}$
1: $P_0 \sim \mathrm{Unif}(\mathcal{X}^\lambda)$
2: **for** $t = 0, 1, 2, \ldots$ until termination condition **do**
3: **for** $i = 1$ to λ **do**
4: $P_{t+1}(i) \sim \mathcal{D}(P_t)$

Runtime Analysis of Population-based EAs

Definition
Given any target subset $B(n) \subset \{0,1\}^n$ (e.g. optima), let

$$T_{B(n)} := \min_{t \in \mathbb{N}} \{t\lambda \mid P_t \cap B(n) \neq \emptyset\}$$

be the first time[3] the population contains an individual in $B(n)$.

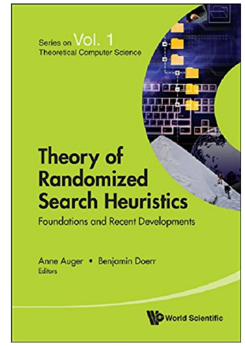
Problem
Show how

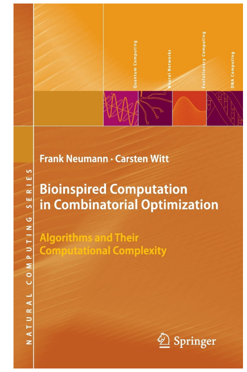
- $\mathbf{E}\left[T_{B(n)}\right]$ (the expected runtime)
- $\Pr\left(T_{B(n)} \leq t\right)$ (the "success" probability)

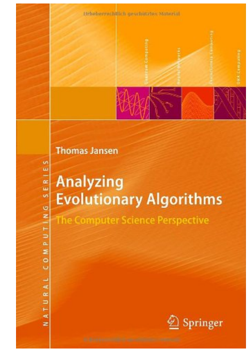
depend on the mapping $\mathcal{D}$.

[3] We here count time as the number of search points that have been sampled since the start of the algorithm. For a typical $\mathcal{D}$ that models an EA, this corresponds to the number of times the fitness function is evaluated.

Runtime Analysis of Evolutionary Algorithms



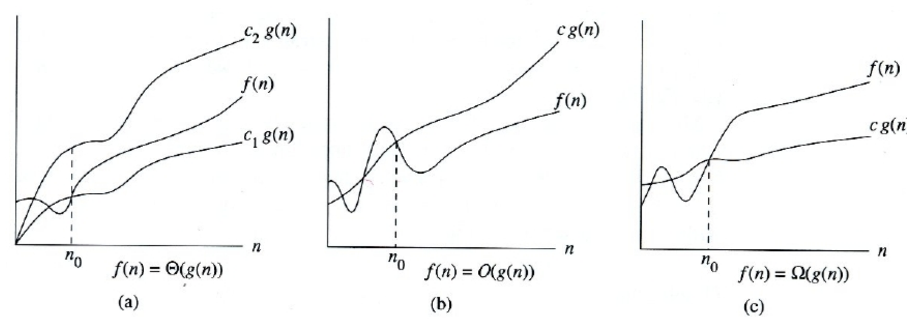




Approaches to Runtime Analysis of Populations

- Infinite population size
- Markov chain analysis He and Yao [2003]
- No parent population, or monomorphic populations
  - (1+1) EA
  - (1+λ) EA Jansen, Jong, and Wegener [2005]
  - (1,λ) EA Rowe and Sudholt [2012]
- Fitness-level techniques
  - (1+λ) EA Witt [2006]
  - (N+N) EAs Chen, He, Sun, Chen, and Yao [2009]
  - non-elitist EAs with unary variation operators Lehre [2011b], Dang and Lehre [2014]
- Classical drift analysis
  - Fitness proportionate selection Neumann, Oliveto, and Witt [2009], Oliveto and Witt [2014, 2015]
- Family trees
  - (μ+1) EA Witt [2006]
  - (μ+1) IA Zarges [2009]
- Multi-type branching processes Lehre and Yao [2012]
  - Negative drift theorem for populations Lehre [2011a]
- Level-based analysis Corus, Dang, Eremeev, and Lehre [2014]

Asymptotic notation



$$f(n) \in O(g(n)) \iff \exists \text{ constants } c, n_0 > 0 \text{ st. } 0 \leq f(n) \leq cg(n)$$
$$f(n) \in \Omega(g(n)) \iff \exists \text{ constants } c, n_0 > 0 \text{ st. } 0 \leq cg(n) \leq f(n)$$
$$f(n) \in \Theta(g(n)) \iff f(n) \in O(g(n)) \text{ and } f(n) \in \Omega(g(n))$$
$$f(n) \in o(g(n)) \iff \lim_{n \to \infty} \frac{f(n)}{g(n)} = 0$$

Level-based Theorem[4]

[4]Corus, Dang, Eremeev, and Lehre [2014]

Outline - Level-based Theorem[5]

1. Definition of levels of search space
2. Definition of "current level" of population
3. Statement of theorem and its conditions
4. Recommendations for how to apply the theorem
5. Some example applications
6. Derivation of special cases
   - Mutation-only EAs
   - Crossover
   - Mutation-only EAs with uncertain fitness (e.g. noise)

[5]It is out of scope of this tutorial to present the proof of this theorem. The proof uses drift analysis with a distance function that takes into account the current level, as well as the number of individuals above the current level.

Level partitioning of search space

Definition

$(A_1, \ldots, A_{m+1})$ a level-partitioning of search space $\mathcal{X}$ if

- $\bigcup_{j=1}^{m+1} A_j = \mathcal{X}$ (i.e., together, levels cover the search space)
- $A_i \cap A_j = \emptyset$ whenever $i \neq j$ (i.e., they are nonoverlapping)
- the last level A_{m+1} covers the optimum for the problem

To denote everything above level j, we also define

$$A_j^+ := \bigcup_{i=j+1}^{m+1} A_i$$

Current level of a population P wrt $\gamma_0 \in (0, 1)$

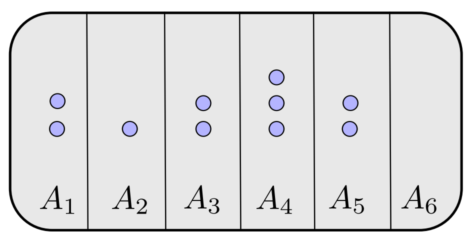
Definition

The unique integer $j \in [m]$ such that

$$|P \cap A_{j-1}^+| \geq \gamma_0 \lambda > |P \cap A_j^+|$$

Example

Current level wrt $\gamma_0 = \frac{1}{2}$ is .....

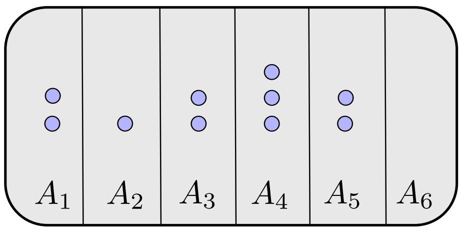


Current level of a population P wrt $\gamma_0 \in (0,1)$

Definition
The unique integer $j \in [m]$ such that

$$|P \cap A_{j-1}^{+}| \geq \gamma_0 \lambda > |P \cap A_j^{+}|$$

Example
Current level wrt $\gamma_0 = \frac{1}{2}$ is 4.



Level-based theorem (informal version)

If the following three conditions are satisfied

(G1) it is always possible to sample above the current level

(G2) the proportion of the population above the current level increases in expectation

(G3) the population size is large enough

then the expected time to reach the last level cannot be too high.

Level-based Theorem[6] (1/2) (setup)

- Given a level-partitioning $(A_1, \ldots, A_{m+1})$ of $\mathcal{X}$
- m *upgrade probabilities* $z_1, \ldots, z_m \in (0,1]$ and $z_{\min} := \min_i z_i$
- a parameter $\delta \in (0,1)$, and
- a constant $\gamma_0 \in (0,1)$,

[6]This version of the theorem simplifies some of the conditions at the cost of a slightly less precise bound on the runtime.

Level-based Theorem (2/2) [Corus, Dang, Eremeev, and Lehre, 2014]

If for any $j \in [m]$, any $\gamma \in [0, \gamma_0]$ and any pop. $P \in \mathcal{X}^{\lambda}$ where

$$\gamma\lambda \leq |P \cap A_j^{+}| < \gamma_0 \lambda \leq |P \cap A_{j-1}^{+}|$$

a new individual $y \sim \mathcal{D}(P)$ is in A_j^{+} with probability

$$\Pr\left(y \in A_j^{+}\right) \geq \begin{cases} z_j & \text{if } \gamma = 0 \\ \gamma(1+\delta) & \text{if } \gamma > 0 \end{cases} \qquad \text{(G1\&G2)}$$

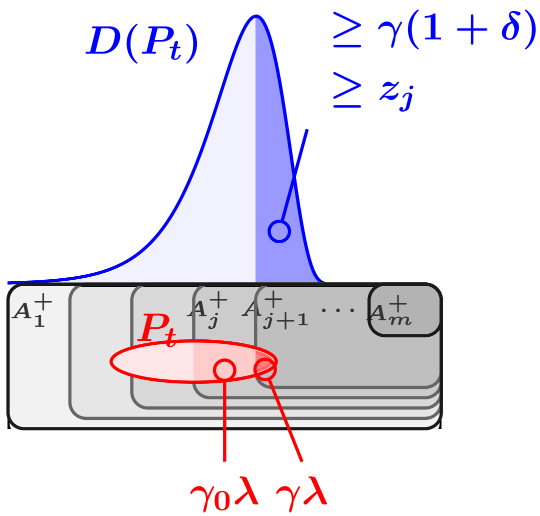
and the population size λ is bounded from below by

$$\lambda \geq \frac{8}{\gamma_0 \delta^2}\left(\ln\left(\frac{m}{\gamma_0 \delta^7 z_{\min}}\right) + 11\right) \qquad \text{(G3)}$$

then the expected time to reach the last level A_{m+1} is less than

$$\frac{1536}{\delta^5}\left(m\lambda \ln(\lambda) + \sum_{j=1}^{m} \frac{1}{z_j}\right)$$

Level-based Theorem visualised



Suggested recipe for application of level-based theorem

1. Find a partition $(A_1, \ldots, A_{m+1})$ of $\mathcal{X}$ that reflects the state of the algorithm, and where A_{m+1} is the goal state.
2. Find parameters γ_0 and δ and a configuration of the algorithm (e.g., mutation rate, selective pressure) such that whenever $|P \cap A_j^+| = \gamma\lambda > 0$, condition (G2) holds, i.e.,

$$\mathbf{Pr}\left(y \in A_j^+\right) \geq \gamma(1+\delta)$$

3. For each level $j \in [m]$, estimate a lower bound $z_j \in (0,1)$ such that whenever $|P \cap A_j^+| = 0$, condition (G1) holds, i.e.,

$$\mathbf{Pr}\left(y \in A_j^+\right) \geq z_j$$

4. Calculate the sufficient population size λ from condition (G3).
5. Read off the bound on expected runtime.

Simple Example to Illustrate Theorem

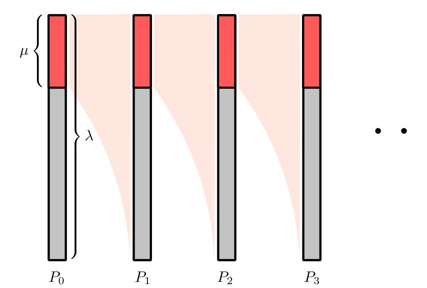
Problem

- search space $\mathcal{X} = \{1, \cdots, m+1\}$
- fitness function $f(x) = x$ (to be maximised)

Evolutionary Algorithm

for $t = 0, 1, 2, \ldots$ until termination condition **do**
  for $i = 1$ to λ **do**
    Select a parent x from P_t using (μ, λ)-selection
    Obtain y by mutating x
    Set i-th offspring $P_{t+1}(i) = y$

(μ, λ)-selection mechanism

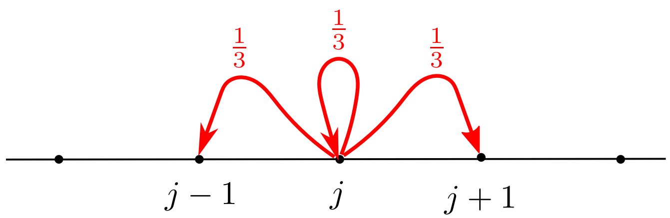


1. Sort the current population $P = (x_1, \ldots, x_\lambda)$ such that

$$f(x_1) \geq f(x_2) \geq \ldots \geq f(x_\lambda)$$

2. **return** $\mathbf{Unif}(x_1, \ldots, x_\mu)$

A simple mutation operator...



$$\mathbf{Pr}\left(V(x) = y\right) = \begin{cases} \frac{1}{3} & \text{if } y \in \{x-1, x, x+1\} \\ 0 & \text{otherwise.} \end{cases}$$

Step 1: Level-partition

Problem

- search space $\mathcal{X} = \{1, \cdots, m+1\}$
- fitness function $f(x) = x$ (to be maximised)

Step 1: Level-partition

Problem

- search space $\mathcal{X} = \{1, \cdots, m+1\}$
- fitness function $f(x) = x$ (to be maximised)

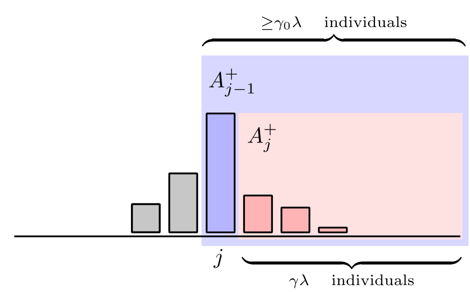
Level-partition of $\mathcal{X}$

$$A_j := \{j\}$$
$$A_j^+ = \{j+1, j+2, \ldots, m+1\}$$

Properties of a Population at Level j

- Assume A_j is the current level of the population P, i.e.,

$$\gamma\lambda = |P \cap A_j^+| < \gamma_0\lambda \leq |P \cap A_{j-1}^+| \qquad (1)$$

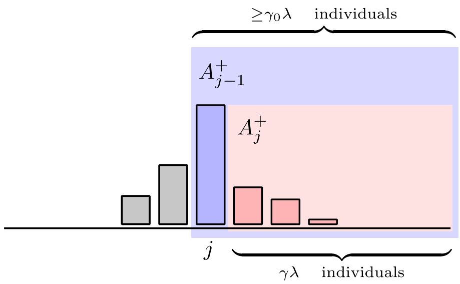


- (μ, λ) selects parent u.a.r. among best μ individuals
- by choosing parameter $\gamma_0 := \mu/\lambda$, assumption (1) implies
  - $\mathbf{Pr}\left(\text{select parent in } A_{j-1}^+\right) =$
  - $\mathbf{Pr}\left(\text{select parent in } A_j^+\right) =$

Properties of a Population at Level j

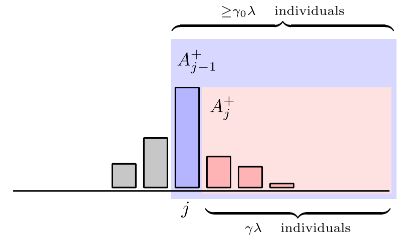
- Assume A_j is the current level of the population P, i.e.,

$$\gamma\lambda = |P \cap A_j^+| < \gamma_0\lambda \leq |P \cap A_{j-1}^+| \qquad (1)$$



- (μ, λ) selects parent u.a.r. among best μ individuals
- by choosing parameter $\gamma_0 := \mu/\lambda$, assumption (1) implies
  - $\mathbf{Pr}\left(\text{select parent in } A_{j-1}^+\right) = 1$
  - $\mathbf{Pr}\left(\text{select parent in } A_j^+\right) = \frac{\gamma\lambda}{\mu}$

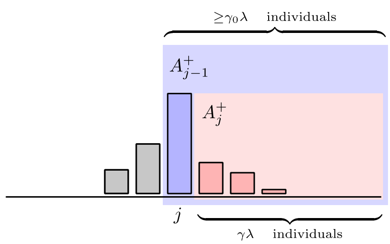
Condition (G2)



$$\mathbf{Pr}\left(y \in A_j^+\right)$$

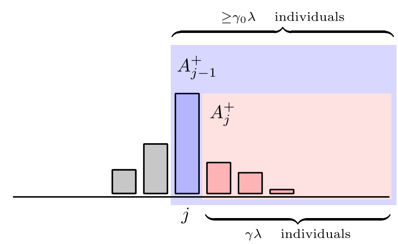
$$\geq \gamma(1+\delta)$$

Condition (G2)



$$\begin{aligned}\mathbf{Pr}\left(y \in A_j^+\right) &\geq \mathbf{Pr}\left(\text{select parent in } A_j^+\right) \cdot \mathbf{Pr}\left(\text{do not downgrade}\right)\\ &\geq \gamma \cdot \frac{\lambda}{\mu} \cdot \left(1 - \frac{1}{3}\right)\\ \\ &\geq \gamma(1+\delta)\end{aligned}$$

Condition (G2)

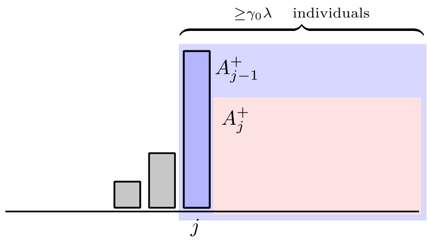


Assuming that $\frac{\lambda}{\mu} = \frac{9}{4} = \frac{1+\frac{1}{2}}{1-\frac{1}{3}}$

$$\begin{aligned}\mathbf{Pr}\left(y \in A_j^+\right) &\geq \mathbf{Pr}\left(\text{select parent in } A_j^+\right) \cdot \mathbf{Pr}\left(\text{do not downgrade}\right)\\ &\geq \gamma \cdot \frac{\lambda}{\mu} \cdot \left(1 - \frac{1}{3}\right)\\ &= \gamma\left(1 + \frac{1}{2}\right).\\ &\geq \gamma(1+\delta)\end{aligned}$$

$\implies$ Condition (G2) satisfied for $\delta = 1/2$.

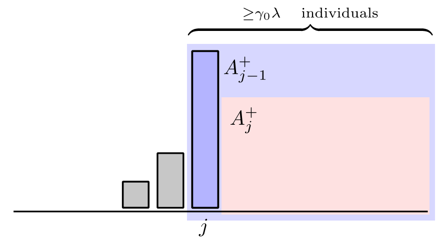
Condition (G1)



$$\mathbf{Pr}\left(y \in A^+_j\right) \geq$$

$$= z_j > 0$$

Condition (G1)



$$\begin{aligned}\mathbf{Pr}\left(y \in A^+_j\right) &\geq \mathbf{Pr}\left(\text{select parent in } A_j\right) \cdot \mathbf{Pr}\left(\text{upgrade offspring to } A^+_j\right) \\ &\geq 1 \cdot \frac{1}{3} \\ &= z_j > 0\end{aligned}$$

$\Longrightarrow$ Condition (G1) satisfied by choosing $z_j := \frac{1}{3}$ for all $j \in [m]$.

Condition (G3) - Sufficiently Large Population

Recall that $\gamma_0 = \mu/\lambda = 4/9$ and $\delta = 1/2$ and $z_* = 1/3$

$$\frac{8}{\gamma_0 \delta^2}\left(\ln\left(\frac{m}{z_* \gamma_0 \delta^7}\right) + 11\right)$$

$$\leq \lambda$$

Condition (G3) - Sufficiently Large Population

Recall that $\gamma_0 = \mu/\lambda = 4/9$ and $\delta = 1/2$ and $z_* = 1/3$

$$\begin{aligned}&\frac{8}{\gamma_0 \delta^2}\left(\ln\left(\frac{m}{z_* \gamma_0 \delta^7}\right) + 11\right) \\ &= 72\left(\ln\left(864m\right) + 11\right) \\ &< 72(\ln(m) + 18) \leq \lambda\end{aligned}$$

Hence, choosing $\lambda \geq 72(\ln(m) + 18)$ sufficient to satisfy (G3).

Example: Summary

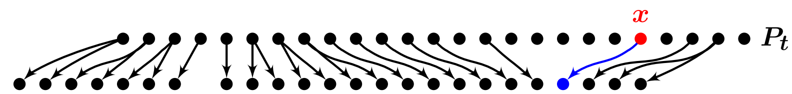
We have shown that if $\lambda \geq 72(\ln(m) + 18)$ and $\mu = 4\lambda/9$

- (G1) is satisfied for $z_j = 1/3$ for all $j \in [m]$
- (G2) is satisfied for $\delta = 1/2$, and
- (G3) is satisfied

hence, by the level-based theorem, the expected running time of the EA is no more than

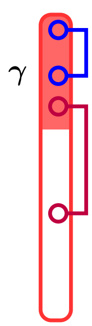
$$\frac{1536}{\delta^5}\left(m\lambda\ln(\lambda) + \sum_{j=1}^{m}\frac{1}{z_j}\right) = O(m\lambda\ln\lambda)$$

Population-Selection Variation Algorithm (PSVA)



for $t = 0$ to ∞ **do**
  for $i = 1$ to λ **do**
    Sample i-th parent x according to $p_{\text{sel}}(P_t, \cdot)$
    Sample i-th offspring $P_{t+1}(i)$ according to $p_{\text{var}}(x, \cdot)$

Selective Pressure



Cumulative selection probability
p_{sel} has cumulative selection probability $\beta(\gamma)$ if

$$\forall P \in \mathcal{X}^\lambda \quad \forall\gamma \in (0, \gamma_0)$$
$$\mathbf{Pr}\ (\ f(p_{\text{sel}}(P)) \geq f(\gamma\text{-ranked})\ )\ \geq\ \beta(\gamma)$$

- If $f(P) = (8, 7, 6, 5, 5, 4, 3, 2)$, then $\beta(3/8) \approx \mathbf{Pr}$ (select an individual ≥ 6).
- 2-tournament selection, $\beta(\gamma) \geq \gamma^2 + 2\gamma(1-\gamma)$
- linear ranking-selection $\beta(\gamma) = \gamma(\eta(1-\gamma)+\gamma)$
- (μ, λ)-selection $\beta(\gamma) \geq \gamma\lambda/\mu$

Corollary for PSVA

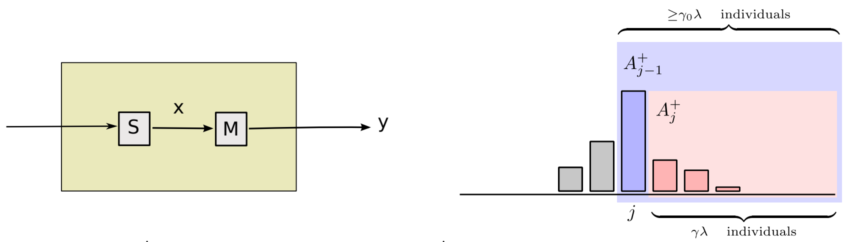
If for any $j \in [m]$

(C1) $p_{\text{var}}(y \in A_j^+ \mid x \in A_{j-1}^+) \geq s_j \geq s_{\min}$
(C2) $p_{\text{var}}(y \in A_j^+ \mid x \in A_j^+) \geq p_0$
(C3) $\beta(\gamma) \geq \frac{\gamma(1+\delta)}{p_0}$
(C4) $\lambda \geq \frac{8}{\gamma_0\delta^2}\left(\ln\left(\frac{m}{\gamma_0^2\delta^7 s_{\min}}\right) + 11\right)$

then the expected time to reach the last level A_{m+1} is less than

$$\frac{1536}{\delta^5}\left(m\lambda\ln(\lambda) + \frac{p_0}{\gamma_0}\sum_{j=1}^{m}\frac{1}{s_j}\right)$$

Proof of Corollary: (C2) & (C3) $\Longrightarrow$ (G2)

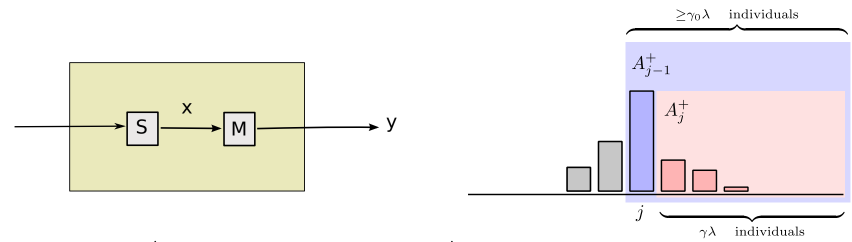


If $|P \cap A_{j-1}^+| \geq \gamma_0\lambda > |P \cap A_j^+| =: \gamma\lambda$ and $y \sim \mathcal{D}(P)$

then $\mathbf{Pr}\left(y \in A_j^+\right) \geq$

$\geq \gamma(1+\delta).$

Proof of Corollary: (C2) & (C3) $\Longrightarrow$ (G2)



If $|P \cap A_{j-1}^+| \geq \gamma_0\lambda > |P \cap A_j^+| =: \gamma\lambda$ and $y \sim \mathcal{D}(P)$

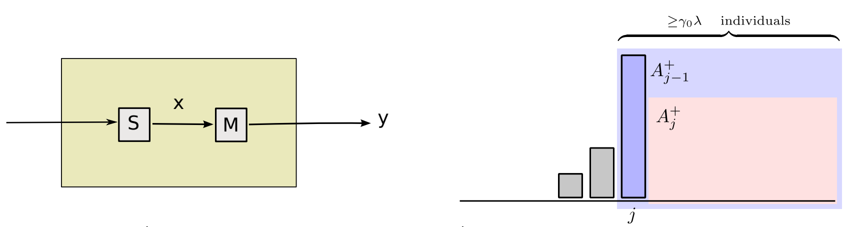
then $\mathbf{Pr}\left(y \in A_j^+\right) \geq \mathbf{Pr}\left(x \in A_j^+\right) \mathbf{Pr}\left(y \in A_j^+ \mid x \in A_j^+\right)$

(i.e., select x from level $j+1$
and do not downgrade it)

$\geq \beta(\gamma)p_0$

$\geq \gamma(1+\delta).$

Proof of Corollary: (C1) & (C3) $\Longrightarrow$ (G1)

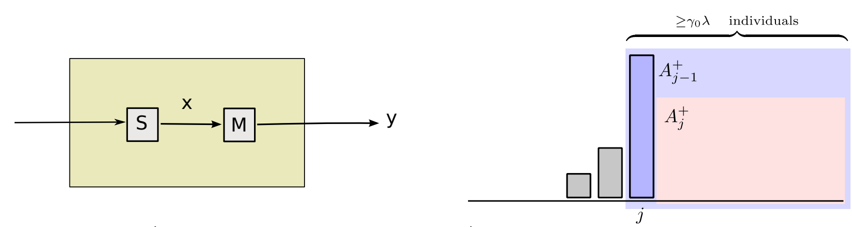


If $|P \cap A_{j-1}^+| \geq \gamma_0\lambda$ and $|P \cap A_j^+| = 0$ and $y \sim \mathcal{D}(P)$

$\mathbf{Pr}\left(y \in A_j^+\right) \geq$

$= z_j > 0$

Proof of Corollary: (C1) & (C3) $\Longrightarrow$ (G1)



If $|P \cap A_{j-1}^+| \geq \gamma_0\lambda$ and $|P \cap A_j^+| = 0$ and $y \sim \mathcal{D}(P)$

$\mathbf{Pr}\left(y \in A_j^+\right) \geq \mathbf{Pr}\left(x \in A_{j-1}^+\right) \mathbf{Pr}\left(y \in A_j^+ \mid x \in A_{j-1}^+\right)$

(i.e., select x from level j and upgrade it)

$\geq \beta(\gamma_0)s_j$

$\geq \gamma_0(1+\delta)s_j/p_0$

$= z_j > 0$

Example Application

$$\text{LEADINGONES}(x) = \sum_{i=1}^{n} \prod_{j=1}^{i} x_j$$

Partition into $n+1$ levels

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

Example Application

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

If $\lambda/\mu > e^{\chi}(1+\delta)$ and $\lambda > c'' \ln(n)$ then

(C1) $p_{\text{var}}\left(y \in A_j^+ \mid x \in A_j\right) \geq \frac{\chi(1-\delta)}{ne^{\chi}}$

(C2) $p_{\text{var}}\left(y \in A_j \cup A_j^+ \mid x \in A_j\right) \geq \frac{1-\delta}{e^{\chi}}$

(C3) $\beta(\gamma) \geq \gamma\lambda/\mu > \gamma(1+\delta)e^{\chi}$

(C4) $\lambda > c'' \ln(n)$

Example Application

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

If $\lambda/\mu > e^{\chi}(1+\delta)$ and $\lambda > c'' \ln(n)$ then

(C1) $p_{\text{var}}\left(y \in A_j^+ \mid x \in A_j\right) \geq \frac{\chi(1-\delta)}{ne^{\chi}} \quad =: \quad s_j =: s_*$

(C2) $p_{\text{var}}\left(y \in A_j \cup A_j^+ \mid x \in A_j\right) \geq \frac{1-\delta}{e^{\chi}} \quad =: \quad p_0$

(C3) $\beta(\gamma) \geq \gamma\lambda/\mu > \gamma(1+\delta)e^{\chi} \quad = \quad \gamma(1+\delta)/p_0$

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

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

Exercise: Our first example, linear ranking selection

How to set the following parameters

- population size λ
- selective pressure η
- mutation rate χ/n

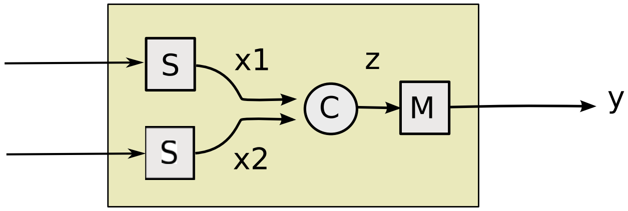
so that the EA optimises LEADINGONES efficiently?

Hints

- (C1), (C2), and (C4) already satisfied as for (μ, λ)-selection
- Remains to show (C3), i.e., $\beta(\gamma) \geq (1+\delta)\gamma/p_0$
- Linear ranking has cumulative selection probability

$$\beta(\gamma) = \gamma(\eta(1-\gamma)+\gamma)$$

Genetic Algorithms with Crossover



Definition (Genetic Algorithm)

for $t = 0, 1, 2, \ldots$ until termination condition **do**
  for $i = 1$ to λ **do**
    Select parents x_1 and x_2 from population P_t acc. to p_{sel}
    Create z by applying a crossover operator to x_1 and x_2.
    Create y by applying a mutation operator to y.

Corollary for Genetic Algorithms

If for any $j \in [m]$

(C1) $p_{\mathrm{var}}(y \in A_j^+ \mid x \in A_{j-1}^+) \geq s_j \geq s_*$

(C2) $p_{\mathrm{var}}(y \in A_j^+ \mid x \in A_j^+) \geq p_0$

(C3) $p_{\mathrm{xor}}(x \in A_j^+ \mid u \in A_{j-1}^+, v \in A_j^+) \geq \varepsilon_1$

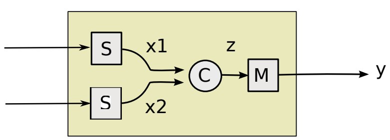
(C4) $\beta(\gamma) \geq \gamma\sqrt{\frac{1+\delta}{p_0\varepsilon_1\gamma_0}}$

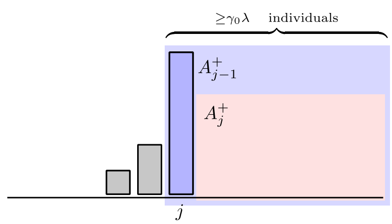
(C5) $\lambda \geq \frac{8}{\delta^2\gamma_0}\left(\ln\left(\frac{m}{\gamma_0^2\delta^7 s_*}\right) + 8\right)$

then the expected time to reach the last level A_{m+1} is less than

$$\frac{1536}{\delta^5}\left(m\lambda\ln(\lambda) + \frac{p_0}{\gamma_0}\sum_{j=1}^{m}\frac{1}{s_j}\right)$$

Proof of Corollary: (C1) and (C4) $\Longrightarrow$ (G1)

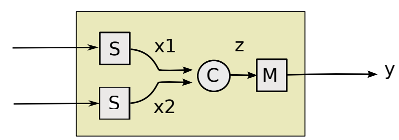


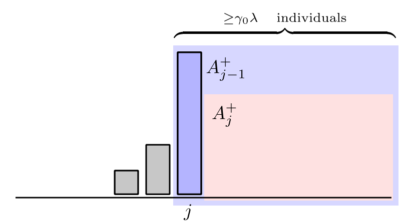


$$\mathbf{Pr}\left(y \in A_j^+\right)$$

$$=: z_j.$$

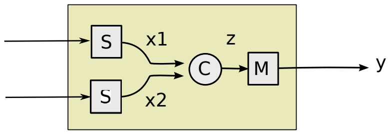
Proof of Corollary: (C1) and (C4) $\Longrightarrow$ (G1)

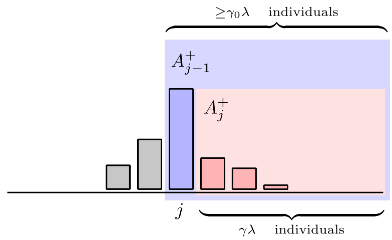




$$\begin{aligned}
\mathbf{Pr}\left(y \in A_j^+\right) &\geq \mathbf{Pr}\left(z \in A_{j-1}^+\right) s_j \\
&\geq \mathbf{Pr}\left(x_1, x_2 \in A_{j-1}^+\right) \varepsilon_1 s_j \\
&\geq \beta(\gamma_0)^2 \varepsilon_1 s_j \\
&\geq \gamma_0^2\left(\frac{1+\delta}{p_0\varepsilon_1\gamma_0}\right)\varepsilon_1 s_j \\
&\geq \gamma_0(1+\delta)s_j/p_0 \\
&=: z_j.
\end{aligned}$$

Proof of Corollary: (C2), (C3), and (C4) $\Longrightarrow$ (G2)

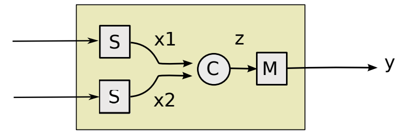


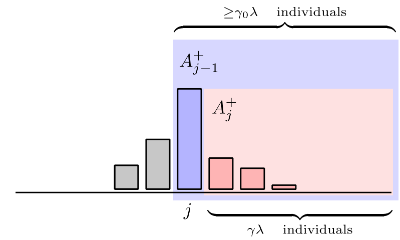


$$\mathbf{Pr}\left(y \in A_j^+\right)$$

$$\geq \gamma(1+\delta).$$

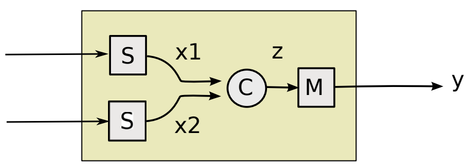
Proof of Corollary: (C2), (C3), and (C4) $\Longrightarrow$ (G2)





$$\begin{aligned}
\mathbf{Pr}\left(y \in A_j^+\right) &\geq \mathbf{Pr}\left(z \in A_j^+\right) p_0 \\
&\geq \mathbf{Pr}\left(x_1 \in A_{j-1}^+\right) \mathbf{Pr}\left(x_2 \in A_j^+\right) \varepsilon_1 p_0 \\
&\geq \beta(\gamma_0)\beta(\gamma)\varepsilon_1 p_0 \\
&\geq \gamma_0 \left(\frac{1+\delta}{p_0\varepsilon_1\gamma_0}\right)\gamma\varepsilon_1 p_0 \\
&\geq \gamma(1+\delta).
\end{aligned}$$

Example application 1 – (μ,λ) GA on LeadingOnes



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

(μ,λ) Genetic Algorithm (GA)

for $t = 0, 1, 2, \ldots$ until termination condition **do**
  for $i = 1$ to λ **do**
    Select a parent x from population P_t acc. to (μ,λ)-selection
    Select a parent y from population P_t acc. to (μ,λ)-selection
    Apply uniform crossover to x and y, i.e. $z := \text{crossover}(x, y)$
    Create $P_{t+1}(i)$ by flipping each bit in z with probability χ/n.

Theorem

If $\lambda > c\log(n)$ for a sufficiently large constant $c > 0$, and $\frac{\lambda}{\mu} > 2e^{\chi}(1+\delta)$ for any constant $\delta > 0$, then the expected runtime of (μ,λ) GA on LeadingOnes *is $O(n\lambda\log(\lambda) + n^2)$.*

Example application 1 – (μ,λ) GA on LeadingOnes

If $\lambda/\mu > \ldots$ and $\lambda > c''\ln(n)$ then

(C1) $p_{\mathsf{var}}(y \in A_j^+ \mid x \in A_{j-1}^+) \geq \dfrac{\chi(1-\delta)}{ne^{\chi}} =: s_j =: s_*$

(C2) $p_{\mathsf{var}}(y \in A_j^+ \mid x \in A_j^+) \geq \dfrac{1-\delta}{e^{\chi}} =: p_0$

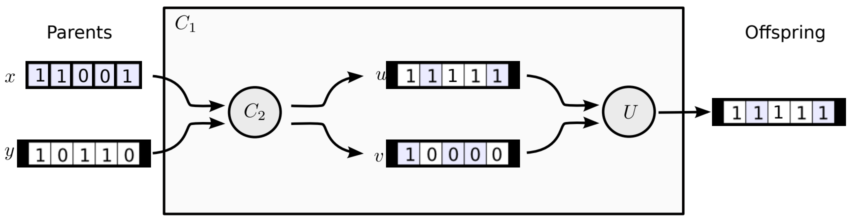
(C3) $p_{\mathsf{xor}}(x \in A_j^+ \mid u \in A_{j-1}^+, v \in A_j^+) \overset{?}{\geq} \varepsilon_1 > 0$

(C4) $\beta(\gamma) \geq \frac{\gamma\lambda}{\mu} \overset{?}{\geq} \gamma\sqrt{\frac{1+\delta}{p_0\varepsilon_1\gamma_0}}$

(C5) $\lambda > c''\ln(n) \geq \dfrac{8}{\delta^2\gamma_0}\left(\ln\left(\dfrac{m}{\gamma_0^2\delta^7 s_*}\right) + 8\right)$

- (C1) and (C2) hold as for mutation-only EAs.
- (C5) holds if the constant $c'' > 0$ is large enough $(m = n)$
- Remains to show that (C3) and (C4) can be satisfied
  - Need to determine the parameter ε_1.
  - Need to determine a lower bound for the ratio λ/μ.

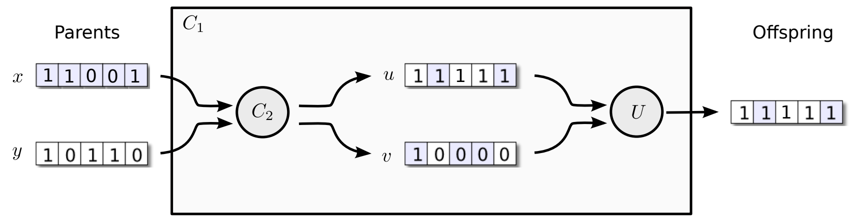
Condition (C3) – (μ,λ) GA on LeadingOnes



Assume that $x \in A_{\geq j+1}$ and $y \in A_{\geq j}$, then

$$\begin{matrix} x_1 = \cdots = x_j = 1 \\ y_1 = \cdots = y_j = 1 \end{matrix} \implies \begin{matrix} u_1 = \cdots = u_j = 1 \\ v_1 = \cdots = u_j = 1 \end{matrix}$$

Condition (C3) – (μ,λ) GA on LeadingOnes



Assume that $x \in A_{\geq j+1}$ and $y \in A_{\geq j}$, then

$$\begin{matrix} x_1 = \cdots = x_j = 1 \\ y_1 = \cdots = y_j = 1 \end{matrix} \implies \begin{matrix} u_1 = \cdots = u_j = 1 \\ v_1 = \cdots = u_j = 1 \end{matrix}$$

Without loss of generality, $u_{j+1} = x_{j+1} = 1$, hence

$$\mathbf{Pr}\left(\text{crossover}(x, y) \in A_{\geq j+1} \mid x \in A_{\geq j+1} \text{ and } y \in A_{\geq j}\right) \geq \frac{1}{2} =: \varepsilon$$

Condition (C4)

We need for all $\gamma \in (0, \gamma_0)$

$$\beta(\gamma) \geq \gamma \sqrt{\frac{1+\delta}{p_0 \varepsilon_1 \gamma_0}}$$

Condition (C4)

We have chosen the parameters

$$\gamma_0 := \mu/\lambda$$
$$p_0 := (1-\delta)e^{-\chi}$$
$$\varepsilon_1 := \frac{1}{2}$$

We need for all $\gamma \in (0, \gamma_0)$

$$\beta(\gamma) \geq \gamma \sqrt{\frac{1+\delta}{p_0 \varepsilon_1 \gamma_0}} = \gamma \sqrt{2e^{\chi} \left(\frac{1+\delta}{1-\delta}\right) \left(\frac{\lambda}{\mu}\right)}$$

Condition (C4)

We have chosen the parameters

$$\gamma_0 := \mu/\lambda$$
$$p_0 := (1-\delta)e^{-\chi}$$
$$\varepsilon_1 := \frac{1}{2}$$

We need to choose μ and λ such that for all $\gamma \in (0, \gamma_0)$

$$\beta(\gamma) \geq \frac{\gamma\lambda}{\mu} \geq \gamma\sqrt{\frac{1+\delta}{p_0\varepsilon_1\gamma_0}} = \gamma\sqrt{2e^{\chi}\left(\frac{1+\delta}{1-\delta}\right)\left(\frac{\lambda}{\mu}\right)}$$

Condition (C4)

We have chosen the parameters

$$\gamma_0 := \mu/\lambda$$
$$p_0 := (1-\delta)e^{-\chi}$$
$$\varepsilon_1 := \frac{1}{2}$$

We need to choose μ and λ such that for all $\gamma \in (0, \gamma_0)$

$$\beta(\gamma) \geq \frac{\gamma\lambda}{\mu} \geq \gamma\sqrt{\frac{1+\delta}{p_0\varepsilon_1\gamma_0}} = \gamma\sqrt{2e^{\chi}\left(\frac{1+\delta}{1-\delta}\right)\left(\frac{\lambda}{\mu}\right)}$$

i.e., it is sufficient to choose μ and λ such that

$$\frac{\lambda}{\mu} \geq 2e^{\chi}\left(\frac{1+\delta}{1-\delta}\right).$$

Example application 1 – (μ,λ) GA on LeadingOnes

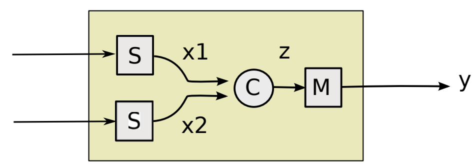
If $\lambda/\mu > 2e^{\chi}\left(\frac{1+\delta}{1-\delta}\right)$ for any const. $\delta > 0$, and $\lambda > c'' \ln(n)$ then

- **(C1)** $p_{\text{var}}(y \in A_j^+ \mid x \in A_{j-1}^+) \geq \frac{\chi(1-\delta)}{ne^{\chi}} =: s_j =: s_*$
- **(C2)** $p_{\text{var}}(y \in A_j^+ \mid x \in A_j^+) \geq \frac{1-\delta}{e^{\chi}} =: p_0$
- **(C3)** $p_{\text{xor}}(x \in A_j^+ \mid u \in A_{j-1}^+, v \in A_j^+) \geq \frac{1}{2} =: \varepsilon_1 > 0$
- **(C4)** $\beta(\gamma) \geq \frac{\gamma\lambda}{\mu} \geq \gamma\sqrt{\frac{1+\delta}{p_0\varepsilon_1\gamma_0}}$
- **(C5)** $\lambda > c'' \ln(n) \geq \frac{8}{\delta^2\gamma_0}\left(\ln\left(\frac{m}{\gamma_0^2\delta^7 s_*}\right) + 8\right)$

Hence, the expected runtime of (μ,λ) GA on LeadingOnes is

$$\mathcal{O}(n\lambda\log(\lambda) + n^2).$$

Example application 2 – (μ,λ) GA on Onemax



$$\textsc{OneMax}(x) := \sum_{i=1}^{n} x_i.$$

(μ,λ) Genetic Algorithm (GA)

for $t = 0, 1, 2, \ldots$ until termination condition **do**
  for $i = 1$ to λ **do**
    Select a parent x from population P_t acc. to (μ, λ)-selection
    Select a parent y from population P_t acc. to (μ, λ)-selection
    Apply uniform crossover to x and y, i.e. $z :=$ crossover(x, y)
    Create $P_{t+1}(i)$ by flipping each bit in z with probability χ/n.

Theorem

If $\lambda > c\log(n)$ for a sufficiently large constant $c > 0$, and $\frac{\lambda}{\mu} > 2e^{\chi}(1+\delta)$ for any constant $\delta > 0$, then the expected runtime of (μ,λ) GA on OneMax is $O(n\lambda\log(\lambda))$.

Example application 2 – (μ,λ) GA on Onemax

If $\lambda/\mu > \ldots$ and $\lambda > c'' \ln(n)$ then

(C1) $p_{\text{var}}(y \in A_j^+ \mid x \in A_{j-1}^+) \geq \frac{\chi(n-j)(1-\delta)}{ne^\chi} =: s_j$

(C2) $p_{\text{var}}(y \in A_j^+ \mid x \in A_j^+) \geq \frac{1-\delta}{e^\chi} =: p_0$

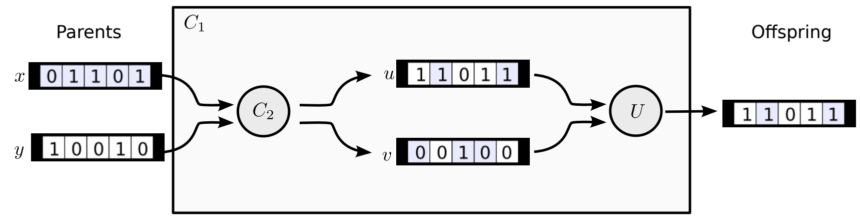
(C3) $p_{\text{xor}}(x \in A_j^+ \mid u \in A_{j-1}^+, v \in A_j^+) \overset{?}{\geq} \varepsilon_1 > 0$

(C4) $\beta(\gamma) \geq \frac{\gamma\lambda}{\mu} \overset{?}{\geq} \gamma\sqrt{\frac{1+\delta}{p_0\varepsilon_1\gamma_0}}$

(C5) $\lambda > c'' \ln(n) \geq \frac{8}{\delta^2\gamma_0}\left(\ln\left(\frac{m}{\gamma_0^2\delta^7 s_*}\right) + 8\right)$

- (C1) and (C2) hold as for mutation-only EAs.
- (C5) holds if the constant $c'' > 0$ is large enough $(m = n + 1)$
- Remains to show that (C3) and (C4) can be satisfied
  - Need to determine the parameter ε_1.
  - Need to determine a lower bound for the ratio λ/μ.

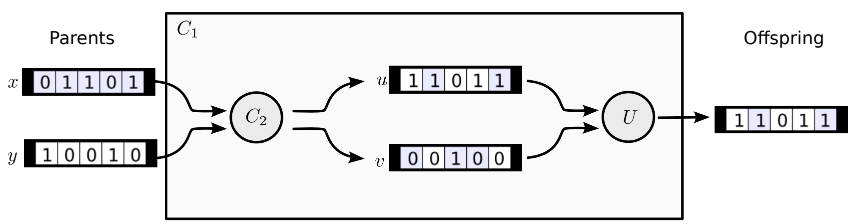
Condition (C3) – (μ,λ) GA on OneMax



Proof.
Assume that $x \in A_{\geq j+1}$ and $y \in A_{\geq j}$,

$$2j + 1 \leq |x| + |y|$$

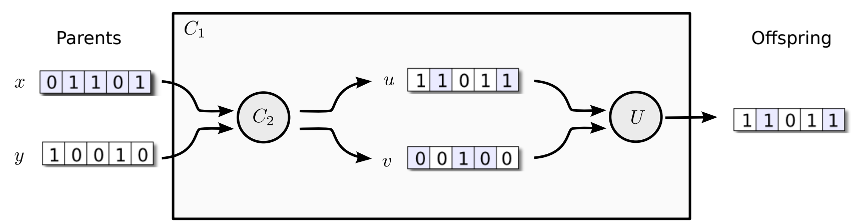
Condition (C3) – (μ,λ) GA on OneMax



Proof.
Assume that $x \in A_{\geq j+1}$ and $y \in A_{\geq j}$,

$$\begin{aligned} 2j + 1 &\leq |x| + |y| \\ &= |u| + |v| \end{aligned}$$

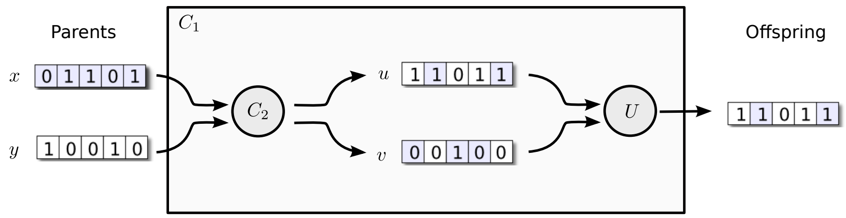
Condition (C3) – (μ,λ) GA on OneMax



Proof.
Assume that $x \in A_{\geq j+1}$ and $y \in A_{\geq j}$, and w.l.o.g. that $|u| \geq |v|$

$$\begin{aligned} 2j + 1 &\leq |x| + |y| \\ &= |u| + |v| \\ &\leq 2|u|. \end{aligned}$$

Condition (C3) – (μ,λ) GA on OneMax



Proof.
Assume that $x \in A_{\geq j+1}$ and $y \in A_{\geq j}$, and w.l.o.g. that $|u| \geq |v|$

$$\begin{aligned} 2j+1 &\leq |x|+|y| \\ &= |u|+|v| \\ &\leq 2|u|. \end{aligned}$$

Therefore $\mathbf{Pr}\left(u \in A_{\geq j+1}\right) = 1$ and

$$\mathbf{Pr}\left(\text{crossover}(x,y) \in A_{\geq j+1} \mid x \in A_{\geq j+1} \text{ and } y \in A_{\geq j}\right) \geq \frac{1}{2} =: \varepsilon.$$

Lower Bounds

Example application 2 – (μ,λ) GA on Onemax

If $\lambda/\mu > 2e^{\chi}\left(\frac{1+\delta}{1-\delta}\right)$ for any const. $\delta > 0$, and $\lambda > c'' \ln(n)$ then

(C1) $p_{\text{var}}(y \in A_j^+ \mid x \in A_{j-1}^+) \geq \dfrac{\chi(n-j)(1-\delta)}{ne^{\chi}} =: s_j$

(C2) $p_{\text{var}}(y \in A_j^+ \mid x \in A_j^+) \geq \dfrac{1-\delta}{e^{\chi}} =: p_0$

(C3) $p_{\text{xor}}(x \in A_j^+ \mid u \in A_{j-1}^+, v \in A_j^+) \geq \dfrac{1}{2} =: \varepsilon_1 > 0$

(C4) $\beta(\gamma) \geq \frac{\gamma\lambda}{\mu} \geq \gamma\sqrt{\frac{1+\delta}{p_0\varepsilon_1\gamma_0}}$

(C5) $\lambda > c'' \ln(n) \geq \dfrac{8}{\delta^2\gamma_0}\left(\ln\left(\dfrac{m}{\gamma_0^2\delta^7 s_*}\right)+8\right)$

Hence, the expected runtime of (μ,λ) GA on OneMax is

$$\mathcal{O}(n\lambda\log(\lambda)+n\log(n)).$$

Lower Bounds

Problem
Consider a sequence of populations $P_1, \ldots$ over a search space $\mathcal{X}$, and a target region $A \subset \mathcal{X}$ (e.g., the set of optimal solutions), let

$$T_A := \min\{\, \lambda t \mid P_t \cap A \neq \emptyset \,\}$$

We would like to prove statements on the form

$$\mathbf{Pr}\left(T_A \leq t(n)\right) \leq e^{-\Omega(n)}. \qquad (2)$$

- i.e., with overwhelmingly high probability, the target region A has not been found in $t(n)$ evaluations
- lower bounds often harder to prove than upper bounds
- will present an easy to use method that is applicable in many situations

Algorithms considered for lower bounds

Definition (Non-elitist EA with selection and mutation)

for $t = 0, 1, 2, \ldots$ until termination condition **do**
  for $i = 1$ to λ **do**
    Select parent x from population P_t according to p_{sel}
    Flip each position in x independently with probability χ/n.
    Let the i-th offspring be $P_{t+1}(i) := x$.
    (i.e., create offspring by mutating the parent)

Assumptions

- population size $\lambda \in \mathbf{poly}(n)$, i.e. not exponentially large
- bitwise mutation with probability χ/n, but no crossover.
- results hold for any non-elitist selection scheme p_{sel} that satisfy some mild conditions to be described later.

Reproductive rate[7]

Definition
For any population $P = (x_1, \ldots, x_\lambda)$ let $p_{\text{sel}}(x_i)$ be the probability that individual x_i is selected from the population P

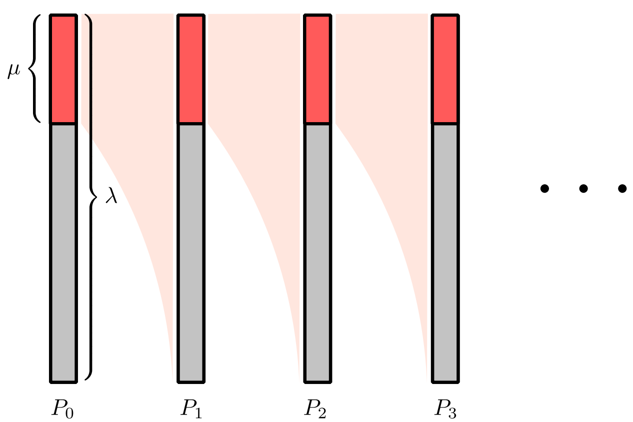
- The reproductive rate of individual x_i is $\lambda \cdot p_{\text{sel}}(x_i)$.
- The reproductive rate of a selection mechanism is bounded from above by α_0 if

$$\forall P \in \mathcal{X}^\lambda, \ \forall x \in P \quad \lambda \cdot p_{\text{sel}}(x) \ \leq \ \alpha_0$$

(i.e., no individual gets more than α_0 offspring in expectation)

[7] The reproductive rate of an individual as defined here corresponds to the notion of "fitness" as used in the field of population genetics, i.e., the expected number of offspring.

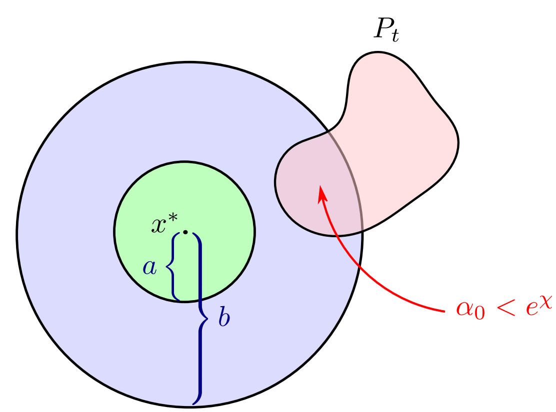
(μ, λ)-selection mechanism



Probability of selecting i-th individual is $p_i \in \{0, \frac{1}{\mu}\}$.

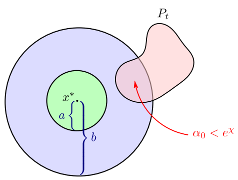
- reproductive rate bounded by $\alpha_0 = \lambda/\mu$

Negative Drift Theorem for Populations (informal)



If individuals closer than b of target has reproductive rate $\alpha_0 < e^\chi$,
then it takes exponential time $e^{c(b-a)}$ to reach within a of target.

Negative Drift Thm. for Populations [Lehre, 2011a]



Consider the non-elitist EA with

- population size $\lambda = \text{poly}(n)$
- bitwise mutation rate χ/n for $0 < \chi < n$

let $T := \min\{t \mid H(P_t, x^*) \leq a\}$ for any $x^* \in \{0,1\}^n$.

If there are constants $\alpha_0 \geq 1$, $\delta > 0$ and integers $a(n)$ and $b(n) < \frac{n}{\chi}$ where $b(n) - a(n) = \omega(\ln n)$, st.

(C1) If $a(n) < H(x, x^*) < b(n)$ then $\lambda \cdot p_{\text{sel}}(x) \leq \alpha_0$.

(C2) $\psi := \ln(\alpha_0)/\chi + \delta < 1$

(C3) $b(n) < \min\left\{\frac{n}{5}, \frac{n}{2}\left(1 - \sqrt{\psi(2-\psi)}\right)\right\}$

then there exist constants $c, c' > 0$ such that

$$\Pr\left(T \leq e^{c(b(n)-a(n))}\right) \leq e^{-c'(b(n)-a(n))}.$$

The worst individuals have low reproductive rate

Lemma

Consider any selection mechanism which for $x, y \in P$ satisfies

(a) *If $f(x) > f(y)$, then $p_{sel}(x) > p_{sel}(y)$.*
(selection probabilities are monotone wrt fitness)

(b) *If $f(x) = f(y)$, then $p_{sel}(x) = p_{sel}(y)$.*
(ties are drawn randomly)

If $f(x) = \min_{y \in P} f(y)$, then $p_{sel}(x) \leq 1/\lambda$.
(individuals with lowest fitness have reproductive rate ≤ 1)

Proof.

- By (a) and (b), $p_{\text{sel}}(x) = \min_{y \in P} p_{\text{sel}}(y)$.
- $1 = \sum_{x \in P} p_{\text{sel}}(x) \geq \lambda \cdot \min_{y \in P} p_{\text{sel}}(y) = \lambda \cdot p_{\text{sel}}(x)$.

□

Example 1: Needle in the haystack

Definition

$$\textsc{Needle}(x) = \begin{cases} 1 & \text{if } x = 1^n \\ 0 & \text{otherwise.} \end{cases}$$

Theorem

The optimisation time of the non-elitist EA with any selection mechanism satisfying the properties above[8] on $\textsc{Needle}$ is at least e^{cn} with probability $1 - e^{-\Omega(n)}$ for some constant $c > 0$.

[8] From black-box complexity theory, it is known that $\textsc{Needle}$ is hard for all search heuristics (Droste et al 2006).

Example 1: Needle in the haystack (proof[9])

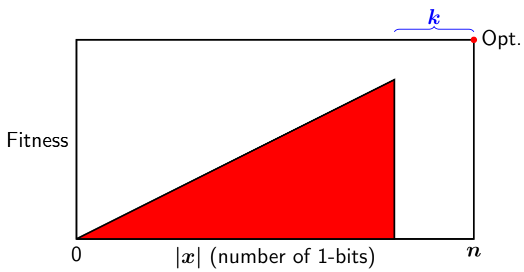
- Apply negative drift theorem with $a(n) := 1$.
- By previous lemma, can choose $\alpha_0 = 1$ for any $b(n)$, hence $\psi = \ln(\alpha)/\chi + \delta = \delta < 1$ for all χ and $\delta < 1$.
- Choosing the parameters $\delta := 1/10$ and $b(n) := n/6$ give

$$\min\left\{\frac{n}{5}, \frac{n}{2}\left(1 - \sqrt{\psi(2-\psi)}\right)\right\} = \frac{n}{5} < b(n).$$

- It follows that $\Pr\left(T \leq e^{c(b(n)-a(n))}\right) \leq e^{-\Omega(n)}$.

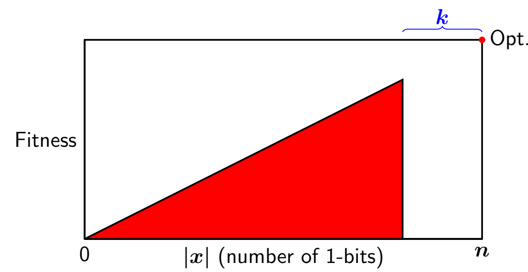
[9] For simplicity, we assume that $b(n) \leq n/\chi$.

Exercise: Optimisation time on Jump_k



$$\text{JUMP}_k(x) := \begin{cases} 0 & \text{if } n-k \le |x| < n, \\ |x| & \text{otherwise.} \end{cases}$$

Exercise: Optimisation time on Jump_k



Recipe

- $a(n) = 1$
- $b(n) = k$
- $\alpha_0 = 1$ as before
- small δ

$$\text{JUMP}_k(x) := \begin{cases} 0 & \text{if } n-k \le |x| < n, \\ |x| & \text{otherwise.} \end{cases}$$

When the best individuals have low reproductive rate

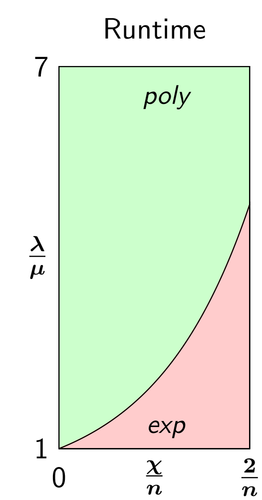
Remark

- The negative drift conditions hold trivially if $\alpha_0 < e^{\chi}$ holds for all individuals.

Example (Insufficient selective pressure)

Selection mechanism	Parameter settings
Linear ranking selection	$\eta < e^{\chi}$
k-tournament selection	$k < e^{\chi}$
(μ,λ)-selection	$\lambda < \mu e^{\chi}$
Any in cellular EAs	$\Delta(G) < e^{\chi}$

Mutation-selection balance



Example

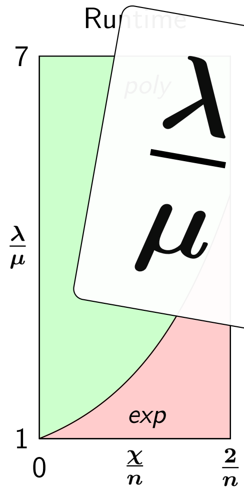
The runtime T of a non-elitist EA with

- (μ, λ)-selection
- bit-wise mutation rate χ/n
- population size $\lambda > c\log(n)$

on LeadingOnes has expected value

$$\mathbf{E}[T] = \begin{cases} e^{\Omega(n)} & \text{if } \lambda < \mu e^{\chi} \\ O(n\lambda\ln(\lambda) + n^2) & \text{if } \lambda > \mu e^{\chi} \end{cases}$$

Mutation-selection balance



$$\frac{\lambda}{\mu} > e^{\chi}$$

Example

The runtime T of a non-elitist EA with

- (μ, λ)-selection
- bit-wise mutation rate χ/n
- population size $\lambda > c \log(n)$

on LeadingOnes has expected value

$$\mathbf{E}[T] = \begin{cases} e^{\Omega(n)} & \text{if } \lambda < \mu e^{\chi} \\ O(n\lambda \ln(\lambda) + n^2) & \text{if } \lambda > \mu e^{\chi} \end{cases}$$

Other Example Applications

Expected runtime of EA with bit-wise mutation rate χ/n

Selection Mechanism	High Selective Pressure
Fitness Proportionate	$\nu > f_{\max} \ln(2e^{\chi})$
Linear Ranking	$\eta > e^{\chi}$
k-Tournament	$k > e^{\chi}$
(μ, λ)	$\lambda > \mu e^{\chi}$
Cellular EAs	
OneMax	$O(n\lambda^2)$
LeadingOnes	$O(n\lambda^2 + n^2)$
Linear Functions	$O(n\lambda^2 + n^2)$
r-Unimodal	$O(r\lambda^2 + nr)$
Jump$_r$	$O(n\lambda^2 + (n/\chi)^r)$

Other Example Applications

Expected runtime of EA with bit-wise mutation rate χ/n

Selection Mechanism	High Selective Pressure	Low Selective Pressure
Fitness Proportionate	$\nu > f_{\max} \ln(2e^{\chi})$	$\nu < \chi/\ln 2$ and $\lambda \geq n^3$
Linear Ranking	$\eta > e^{\chi}$	$\eta < e^{\chi}$
k-Tournament	$k > e^{\chi}$	$k < e^{\chi}$
(μ, λ)	$\lambda > \mu e^{\chi}$	$\lambda < \mu e^{\chi}$
Cellular EAs		$\Delta(G) < e^{\chi}$
OneMax	$O(n\lambda^2)$	$e^{\Omega(n)}$
LeadingOnes	$O(n\lambda^2 + n^2)$	$e^{\Omega(n)}$
Linear Functions	$O(n\lambda^2 + n^2)$	$e^{\Omega(n)}$
r-Unimodal	$O(r\lambda^2 + nr)$	$e^{\Omega(n)}$
Jump$_r$	$O(n\lambda^2 + (n/\chi)^r)$	$e^{\Omega(n)}$

Fitness proportional selection + crossover Oliveto and Witt [2014, 2015]

Definition (Simple Genetic Algorithm (SGA) (Goldberg 1989))

for $t = 0, 1, 2, \ldots$ until termination condition **do**
  for $i = 1$ to λ **do**
    Select two parents x and y from P_t proportionally to fitness
    Obtain z by applying uniform crossover to x and y with $p = 1/2$
    Flip each position in z independently with $p = 1/n$.
    Let the i-th offspring be $P_{t+1}(i) := x$.
    (i.e., create offspring by crossover followed by mutation)

Application to OneMax

Expected Behaviour

- Backward drift due to mutation close to the optimum
- no positive drift due to crossover
- selection too weak to keep positive fluctuations

Difficulties When Introducing Crossover:

- Variance of offspring distribution
- # flipping bits due to mutation Poisson-distributed → variance $O(1)$
- # of one-bits created by crossover binomially distributed according to Hamming distance of parents and $1/2$ → deviation $\Omega(\sqrt{n})$ possible

Negative Drift Theorem With Scaling

Let X_t, $t \geq 0$, random variable describing a stochastic process over finite state space $S \subseteq \mathbb{R}$;

If there ∃ interval $[a, b]$ and, possibly depending on $\ell := b - a$, bound $\epsilon(\ell) > 0$ and scaling factor $r(\ell)$ st.

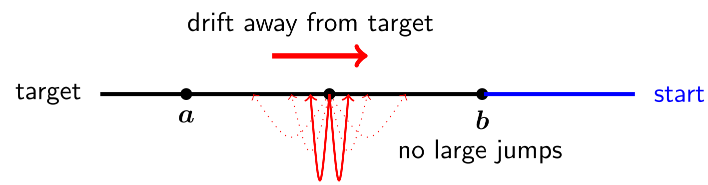
(C1) $E(X_{t+1} - X_t \mid X_0, \dots, X_t \ \wedge\ a < X_t < b) \ \geq\ \epsilon$,

(C2) $\mathrm{Prob}(|X_{t+1} - X_t| \geq jr \mid X_0, \dots, X_t \ \wedge\ a < X_t) \ \leq\ e^{-j}$ for $j \in \mathbb{N}_0$,

(C3) $1 \leq r \leq \min\{\epsilon^2 \ell, \sqrt{\epsilon\ell/(132\log(\epsilon\ell))}\}$.

then

$$\mathrm{Pr}\left(T \leq e^{\epsilon\ell/(132r^2)}\right) = O(e^{-\epsilon\ell/(132r^2)}).$$



Negative Drift Theorem With Scaling

Let X_t, $t \geq 0$, random variable describing a stochastic process over finite state space $S \subseteq \mathbb{R}$;

If there ∃ interval $[a, b]$ and, possibly depending on $\ell := b - a$, bound $\epsilon(\ell) > 0$ and scaling factor $r(\ell)$ st.

(C1) $E(X_{t+1} - X_t \mid X_0, \dots, X_t \ \wedge\ a < X_t < b) \ \geq\ \epsilon$,

(C2) $\mathrm{Prob}(|X_{t+1} - X_t| \geq jr \mid X_0, \dots, X_t \ \wedge\ a < X_t) \ \leq\ e^{-j}$ for $j \in \mathbb{N}_0$,

(C3) $1 \leq r \leq \min\{\epsilon^2 \ell, \sqrt{\epsilon\ell/(132\log(\epsilon\ell))}\}$.

then

$$\mathrm{Pr}\left(T \leq e^{\epsilon\ell/(132r^2)}\right) = O(e^{-\epsilon\ell/(132r^2)}).$$

Potential Function

For drift theorem, capture whole population in one value: For $X = \{x_1, \dots, x_\mu\}$ let $g(X) := \sum_{i=1}^{\mu} e^{\kappa \textsc{OneMax}(x_i)}$.

Negative Drift Theorem With Scaling

Let X_t, $t \geq 0$, random variable describing a stochastic process over finite state space $S \subseteq \mathbb{R}$;

If there ∃ interval $[a, b]$ and, possibly depending on $\ell := b - a$, bound $\epsilon(\ell) > 0$ and scaling factor $r(\ell)$ st.

(C1) $E(X_{t+1} - X_t \mid X_0, \dots, X_t \ \wedge\ a < X_t < b) \ \geq\ \epsilon$,

(C2) $\mathrm{Prob}(|X_{t+1} - X_t| \geq jr \mid X_0, \dots, X_t \ \wedge\ a < X_t) \ \leq\ e^{-j}$ for $j \in \mathbb{N}_0$,

(C3) $1 \leq r \leq \min\{\epsilon^2 \ell, \sqrt{\epsilon\ell/(132\log(\epsilon\ell))}\}$.

then

$$\mathrm{Pr}\left(T \leq e^{\epsilon\ell/(132r^2)}\right) = O(e^{-\epsilon\ell/(132r^2)}).$$

Problem: maybe $r(\ell) = \Omega(\sqrt{\ell})$

Solution

Find bits that are "converged" within population, i.e., either ones or zeros only. Crossover is irrelevant for these.

Diversity

X_t: # individuals with 1 in some fixed position at time t

Assume uniform selection. Then:

- The probability crossover produces an individual with 1 in the fixed position is:
- $\frac{k}{\mu} \cdot \frac{k}{\mu} + 2 \cdot \frac{1}{2} \cdot \frac{k(\mu-k)}{\mu^2} = \frac{k}{\mu}$
- $\{X_t\} \approx B(\mu, k/\mu) \rightsquigarrow E(X_t \mid X_{t-1} = k) = k$ (martingale)
- But random fluctuations $\rightsquigarrow$ absorbing state 0 or μ (due to the variance).

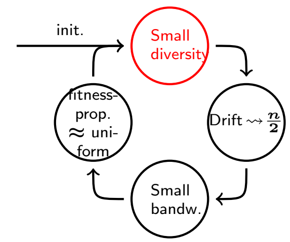
Compare fitness-prop. and uniform selection:

- Basically no difference for small population *bandwidth* (difference of best and worst OneMax-value in pop.)
- $E(X_t \mid X_{t-1} = k) = k \pm 1/(7\mu)$

Result

Let $\mu \leq n^{1/8-\epsilon}$ for an arbitrarily small constant $\epsilon > 0$. Then with probability $1 - 2^{-\Omega(n^{\epsilon/9})}$, the SGA on OneMax does not create individuals with more than $(1+c)\frac{n}{2}$ or less than $(1-c)\frac{n}{2}$ one-bits, for arbitrarily small constant $c > 0$, within the first $2^{n^{\epsilon/10}}$ generations. In particular, it does not reach the optimum then.

Overall Proof Structure



Not a loop, but in each step only exponentially small failure prob.

Summary

- Runtime analysis of evolutionary algorithms
  - mathematically rigorous statements about EA performance
  - most previous results on simple EAs, such as (1+1) EA
  - special techniques developed for population-based EAs
- Level-based method Corus et al. [2014]
  - EAs analysed from the perspective of EDAs
  - Upper bounds on expected optimisation time
  - Example applications include crossover and noise
- Negative drift theorem Lehre [2011a]
  - reproductive rate vs selective pressure
  - exponential lower bounds
  - mutation-selection balance
- Diversity + Bandwidth analysis for fitness proportional selection
  - analysis of crossover
  - low selection pressure
  - exponential lower bounds

Acknowledgements

- Dogan Corus, University of Sheffield, UK
- Duc-Cuong Dang, University of Nottingham, UK
- Anton Eremeev, Omsk Branch of Sobolev Institute of Mathematics, Russia
- Carsten Witt, DTU, Lyngby, Denmark

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 618091 (SAGE) and by the EPSRC under grant no EP/M004252/1.

References I

Tianshi Chen, Jun He, Guangzhong Sun, Guoliang Chen, and Xin Yao. A new approach for analyzing average time complexity of population-based evolutionary algorithms on unimodal problems. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(5):1092–1106, Oct. 2009. ISSN 1083-4419. doi: 10.1109/TSMCB.2008.2012167.

Dogan Corus, Duc-Cuong Dang, Anton V. Eremeev, and Per Kristian Lehre. Level-based analysis of genetic algorithms and other search processes. In *Parallel Problem Solving from Nature - PPSN XIII - 13th International Conference, Ljubljana, Slovenia, September 13-17, 2014. Proceedings*, pages 912–921, 2014. doi: 10.1007/978-3-319-10762-2_90. URL http://dx.doi.org/10.1007/978-3-319-10762-2_90.

Duc-Cuong Dang and Per Kristian Lehre. Refined Upper Bounds on the Expected Runtime of Non-elitist Populations from Fitness-Levels. In *Proceedings of the 16th Annual Conference on Genetic and Evolutionary Computation Conference (GECCO 2014)*, pages 1367–1374, 2014. ISBN 9781450326629. doi: 10.1145/2576768.2598374.

Agoston E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*. SpringerVerlag, 2003. ISBN 3540401849.

David E. Goldberg and Kalyanmoy Deb. A comparative analysis of selection schemes used in genetic algorithms. In *Foundations of Genetic Algorithms*, pages 69–93. Morgan Kaufmann, 1991.

References II

Jun He and Xin Yao. Towards an analytic framework for analysing the computation time of evolutionary algorithms. *Artificial Intelligence*, 145(1-2):59–97, 2003.

Thomas Jansen, Kenneth A. De Jong, and Ingo Wegener. On the choice of the offspring population size in evolutionary algorithms. *Evolutionary Computation*, 13 (4):413–440, 2005. doi: 10.1162/106365605774666921.

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

Per Kristian Lehre. 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, 2011b. ACM. ISBN 978-1-4503-0557-0.

Per Kristian Lehre and Xin Yao. On the impact of mutation-selection balance on the runtime of evolutionary algorithms. *Evolutionary Computation, IEEE Transactions on*, 16(2):225 –241, April 2012. ISSN 1089-778X. doi: 10.1109/TEVC.2011.2112665.

Frank Neumann, Pietro Simone Oliveto, and Carsten Witt. Theoretical analysis of fitness-proportional selection: landscapes and efficiency. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation (GECCO 2009)*, pages 835–842, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-325-9. doi: http://doi.acm.org/10.1145/1569901.1570016.

Pietro S. Oliveto and Carsten Witt. On the runtime analysis of the simple genetic algorithm. *Theoretical Computer Science*, 545:2–19, 2014.

References III

Pietro S. Oliveto and Carsten Witt. Improved time complexity analysis of the simple genetic algorithm. *Theoretical Computer Science*, 605:21–41, 2015.

Jonathan E. Rowe and Dirk Sudholt. The choice of the offspring population size in the (1,λ) ea. In *Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference*, GECCO '12, pages 1349–1356, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1177-9.

Carsten Witt. Runtime Analysis of the $(\mu + 1)$ EA on Simple Pseudo-Boolean Functions. *Evolutionary Computation*, 14(1):65–86, 2006.

Christine Zarges. On the utility of the population size for inversely fitness proportional mutation rates. In *FOGA 09: Proceedings of the tenth ACM SIGEVO workshop on Foundations of genetic algorithms*, pages 39–46, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-414-0. doi: http://doi.acm.org/10.1145/1527125.1527132.

Level based Theorem

To simplify condition (G3), note first that for $\delta \in (0, 1)$

$$\begin{aligned}
&\frac{4(1+\delta)}{\gamma_0\delta^2}\ln\left(\frac{24576(1+\delta)m}{\gamma_0\delta^7 z}\right)\\
< \quad &\frac{4\cdot 2}{\gamma_0\delta^2}\left(\ln\left(\frac{m}{z\gamma_0\delta^7}\right)+\ln(24576\cdot 2)\right)\\
< \quad &\frac{8}{\gamma_0\delta^2}\left(\ln\left(\frac{m}{z\gamma_0\delta^7}\right)+11\right)
\end{aligned}$$

Result

Assuming that $\delta \in (0,1)$ we have $\varepsilon = \delta/2$ and $c = \delta^4/384$.
If $c\lambda > 1$, then

$$\ln(1 + c\lambda) + 1 < \ln(e2c\lambda) = \ln(\lambda) + \ln\left(\frac{e\delta^4}{192}\right) < \ln(\lambda)$$

Consider the case where $c\lambda \leq 1$. By (G3), we must have $\lambda \geq 88 > 2e$. So we get

$$\ln(1 + c\lambda) + 1 \leq \ln(2) + 1 = \ln(2e) \leq \ln(\lambda)$$

We therefore have

$$\begin{aligned}
&\frac{2}{c\varepsilon}\left(m\lambda(1 + \ln(1 + c\lambda)) + \sum_{j=1}^{m} \frac{1}{z_j}\right) \\
&\leq \frac{1536}{\delta^5}\left(m\lambda \ln(\lambda) + \sum_{j=1}^{m} \frac{1}{z_j}\right)
\end{aligned}$$