technische universität
dortmund

Computational Intelligence

Winter Term 2011/12

Prof. Dr. Günter Rudolph

Lehrstuhl für Algorithm Engineering (LS 11)

Fakultät für Informatik

TU Dortmund

Plans for Today

Evolutionary Algorithms

We know

- what evolutionary algorithms are and
- how we can design evolutionary algorithms.

What do we want to do now?

What do we do if we design a problem-specific algorithm?

1. prove its correctness
2. analyze its performance: (expected) run time

What does this mean for evolutionary algorithms in the context of optimization?

1. prove that max. f-value in population converges to global max. of f for $t \to \infty$
2. analyze how long this takes on average: expected optimization time

Analysis of Evolutionary Algorithms

What kind of evolutionary algorithms do we want to analyze?

clearly — all kinds of evolutionary algorithms

more realistic — very simple evolutionary algorithms
at least as starting point

For what kind of problems do we want to do analysis?

clearly — all kinds of problems

more realistic — very simple problems — "toy problems"
at least as starting point

On "Toy Problems"

better term example problems

Why should we care?

- support analysis, help to develop analytical tools
- are easy to understand, are clearly structured
- present typical situations in a paradigmatic way
- make important aspects visible
- act as counter examples
- help to discover general properties
- are important tools for further design and analyis

Upper bounds with f-based partitions

Method of f-based partitions works well with plus-selection.

Definition

Let $f\colon \{0,1\}^n \to \mathbb{R}$. A partition $L_0, L_1, \dots, L_k$ of $\{0,1\}^n$ is called f-based partition iff the following holds.

1. $\forall i, j \in \{0, \dots, k\}\colon \forall x \in L_i\colon \forall y \in L_j\colon\ (i < j \Rightarrow f(x) < f(y))$
2. $L_k = \{x \in \{0,1\}^n \mid f(x) = \max\{f(y) \mid y \in \{0,1\}^n\}\}$

Often the trivial f-based parition works well.

$k := |\{f(x) \mid x \in \{0,1\}^n\}| - 1$

$\{f(x) \mid x \in \{0,1\}^n\} = \{f_0, f_1, \dots, f_k\}$ with $f_0 < f_1 < \cdots < f_k$

for $i \in \{0, 1, \dots, k\}\colon L_i := \{x \in \{0,1\}^n \mid f(x) = f_i\}$

Example: (1+1) EA on OneMax

$\textsc{OneMax}\colon \{0,1\}^n \to \mathbb{R}$ with $\textsc{OneMax}(x) := \sum\limits_{i=1}^{n} x_i$

The (1+1) EA

1. **Initialization**
   Choose $x \in \{0,1\}^n$ uniformaly at random.
2. **Mutation**
   $y :=$ mutate(x); (standard bit mutations, $p_m = 1/n$)
3. **Selection**
   If $f(y) \geq f(x)$, Then $x := y$.
4. **"Stoppping Criterion"**
   Continue at line 2.

Method: f-based partitions

Key Observation:
(1+1) EA leaves each fitness layer at most once.

Lower bound on the probability to leave L_i:
$s_i := \min\limits_{x \in L_i} \sum\limits_{j=i+1}^{k} \sum\limits_{y \in L_j} p_m^{\mathsf{H}(x,y)} \cdot (1 - p_m)^{n - \mathsf{H}(x,y)}$

Upper bound on the expected time needed to leave L_i:
E (time to leave L_i) $\leq 1/s_i$

Upper bound on the expected optimization time:
$\mathsf{E}\left(T_{(1+1)\ \mathsf{EA}, f}\right) \leq \sum\limits_{i=0}^{k-1} 1/s_i$

Upper Bound: (1+1) EA on OneMax

Use trivial OneMax-based partition.

To leave L_i, flip exactly 1 out of $n-i$ 0-bits.

$s_i \geq \binom{n-i}{1} \cdot \frac{1}{n} \cdot \left(1-\frac{1}{n}\right)^{n-1} \geq \frac{n-i}{en}$

$$\begin{aligned} \mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\mathrm{OneMax}}\right) &\leq \sum_{i=0}^{n-1} \frac{en}{n-i} = en \cdot \sum_{i=1}^{n} \frac{1}{i} \\ &< en\ln(n) + en \\ &= O(n \log n) \end{aligned}$$

Linear Functions

Observation $\mathrm{OneMax}(x) = \sum_{i=1}^{n} x[i]$
is of the form $f(x) = w_0 + \sum_{i=1}^{n} w_i \cdot x[i]$

Definition $f\colon \{0,1\}^n \to \mathbb{R}$ is called linear
if f is of the form $f(x) = w_0 + \sum_{i=1}^{n} w_i \cdot x[i]$

Are all linear functions like OneMax?

Definition different extreme example
$\mathrm{BinVal}\colon \{0,1\}^n \to \mathbb{R}$ with
$\mathrm{BinVal}(x) = \sum_{i=1}^{n} 2^{n-i} \cdot x[i]$

Upper bound for $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\mathrm{BinVal}}\right)$

Consider trivial fitness levels
$\forall i \in \{0,1,\ldots,2^n-1\} : L_i := \{x \in \{0,1\}^n \mid \mathrm{BinVal}(x) = i$

without considering s_i at best upper bound $\geq 2^n - 1$ achievable

Observation for good upper bounds number of fitness levels needs to be small

Try more clever fitness levels
$\forall i \in \{0,1,\ldots,n-1\}$:
$L_i := \left\{x \in \{0,1\}^n \setminus \left(\bigcup_{j=0}^{i-1} L_j\right) \mid \mathrm{BinVal}(x) < \sum_{j=0}^{i} 2^{n-1-j}\right\}$

Upper bound for $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\mathrm{BinVal}}\right)$ (II)

$\forall i \in \{0,1,\ldots,n-1\}$:
$L_i := \left\{x \in \{0,1\}^n \setminus \left(\bigcup_{j=0}^{i-1} L_j\right) \mid \mathrm{BinVal}(x) < \sum_{j=0}^{i} 2^{n-1-j}\right\}$

obvious $s_i \geq \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1} \geq \frac{1}{en}$

Theorem $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\mathrm{BinVal}}\right) \leq en^2$

□

Upper bounds for linear functions

Theorem f linear $\Rightarrow \mathsf{E}\left(T_{(1+1)\ \mathsf{EA},f}\right) = O(n^2)$

Proof $f(x) = \sum\limits_{i=1}^{n} w_i x[i]$ mit $w_1 \geq w_2 \geq \cdots \geq w_n$

Definition fitness levels for $i \in \{0, 1, \ldots, n-1\}$

$$L_i := \left\{ x \in \{0,1\}^n \setminus \left(\bigcup_{j=0}^{i-1} L_j \right) \mid f(x) < \sum_{j=1}^{i+1} w_j \right\}$$

$$L_n := \{1^n\}$$

Observation in order to leave L_i:
sufficient to mutate left-most 0-bit

thus $\mathsf{E}\left(T_{(1+1)\ \mathsf{EA},f}\right) \leq en^2$ □

A lower bound for the (1+1) EA on OneMax

The unique global optimum of OneMax is 1^n.

Event A: Initially, there are $\geq \lfloor n/2 \rfloor$ 0-bits.

Total Probability Theorem:
$\mathsf{E}\left(T_{(1+1)\ \mathsf{EA},\textsc{OneMax}}\right) \geq \mathsf{Prob}\,(A) \cdot \mathsf{E}\left(T_{(1+1)\ \mathsf{EA},\textsc{OneMax}} \mid A\right)$

$\mathsf{Prob}\,(A) \geq 1/2$

All these $\lfloor n/2 \rfloor$ bits need to be mutated at least once.

$\forall t \in \mathbb{N}$: $\mathsf{E}\left(T_{(1+1)\ \mathsf{EA},\textsc{OneMax}} \mid A\right) \geq t \cdot \mathsf{Prob}\left(T_{(1+1)\ \mathsf{EA},\textsc{OneMax}} > t \mid A\right)$

Clearly: $\mathsf{Prob}\left(T_{(1+1)\ \mathsf{EA},\textsc{OneMax}} > t \mid A\right)$
$\geq \mathsf{Prob}\,(\exists$ bit from these $\lfloor n/2 \rfloor$ bits without mutation$)$

On mutating bits. . .

Prob (1 specific bit flips) $= \frac{1}{n}$

Prob (1 specfic bit does not flip) $= 1 - \frac{1}{n}$

Prob (1 specific bit does not flip in t steps) $= \left(1 - \frac{1}{n}\right)^t$

Prob (1 specific bit flips at least once in t steps) $= 1 - \left(1 - \frac{1}{n}\right)^t$

Prob ($\lfloor n/2 \rfloor$ specific bits all flip at least once in t steps)
$= \left(1 - \left(1 - \frac{1}{n}\right)^t\right)^{\lfloor n/2 \rfloor}$

Prob ($\exists$ bit out of $\lfloor n/2 \rfloor$ specific bits that never flips in t steps)
$= 1 - \left(1 - \left(1 - \frac{1}{n}\right)^t\right)^{\lfloor n/2 \rfloor}$

Choosing t appropriately. . .

Prob ($\exists$ bit out of $\lfloor n/2 \rfloor$ specific bits that never flips in t steps)
$= 1 - \left(1 - \left(1 - \frac{1}{n}\right)^t\right)^{\lfloor n/2 \rfloor}$

$t := (n-1) \ln n$

$$\begin{aligned}
\mathsf{Prob}\,(\cdots) &= 1 - \left(1 - \left(1 - \frac{1}{n}\right)^{(n-1)\ln n}\right)^{\lfloor n/2 \rfloor} \\
&\geq 1 - \left(1 - \left(\frac{1}{e}\right)^{\ln n}\right)^{\lfloor n/2 \rfloor} \\
&\geq 1 - \left(1 - \frac{1}{n}\right)^{\lfloor n/2 \rfloor} \\
&\geq 1 - \left(\frac{1}{e}\right)^{1/2} = 1 - e^{-1/2} > 0.39
\end{aligned}$$

Putting things together. . .

Prob $\left(T_{\text{(1+1) EA},\textsc{OneMax}} > (n-1)\ln n \mid A\right) \geq 1 - e^{-1/2}$

E $\left(T_{\text{(1+1) EA},\textsc{OneMax}} \mid A\right)$
$\geq (n-1)\ln n \cdot$ Prob $\left(T_{\text{(1+1) EA},\textsc{OneMax}} > t \mid A\right)$
$\geq \left(1 - e^{-1/2}\right)(n-1)\ln n$

E $\left(T_{\text{(1+1) EA},\textsc{OneMax}}\right)$
$\geq$ Prob $(A) \cdot$ E $\left(T_{\text{(1+1) EA},\textsc{OneMax}} \mid A\right)$
$\geq \frac{1}{2} \cdot \left(1 - e^{-1/2}\right)(n-1)\ln n$
$> 0.19(n-1)\ln n = \Omega(n \log n)$

Conclusion: E $\left(T_{\text{(1+1) EA},\textsc{OneMax}}\right) = \Theta(n \log n)$

The Coupon Collector Theorem

Scenario Collect coupons of n different types
unless you have at least one of each type.
Obtain single coupons, each time independently
each type with equal probability.
Let T be the number of coupons obtained at the end.

Theorem

1. E $(T) = n \ln n + O(n)$
2. $\forall \beta \geq 1$: Prob $(T > \beta n \ln n) \leq n^{-(\beta-1)}$
3. $\forall c \in \mathbb{R}$: Prob $(T > n \ln n + cn) \leq 1 - e^{-e^{-c}}$

A More Flexibel Proof Method

Observations

- f-based partitions restricted to "well behaving" functions
- direct lower bound often too difficult

How can we find a more flexibel method?

Observation f-based partition measure progress by $f(x_{t+1}) - f(x_t)$

Idea consider a more general measure of progress

Define distance $d\colon Z \to \mathbb{R}_0^+$, ($Z$ set of all populations)
with $d(P) = 0 \Leftrightarrow P$ contains optimal solution

Caution "Distance" need not be a metric!

Drift

Define distance $d\colon Z \to \mathbb{R}_0^+$, ($Z$ set of all populations)
with $d(P) = 0 \Leftrightarrow P$ contains optimal solution

Observation $T = \min\{t \mid d(P_t) = 0\}$

Consider maximum distance $M := \max\{d(P) \mid P \in Z\}$,
decrease in distance $D_t := d(P_{t-1}) - d(P_t)$

Definition E $(D_t \mid T \geq t)$ is called drift.

Pessimistic point of view $\Delta := \min\{\text{E}(D_t \mid T \geq t) \mid t \in \mathbb{N}_0\}$

Drift Theorem (Upper Bound) $\Delta > 0 \Rightarrow$ E $(T) \leq M/\Delta$

Upper Bound Drift Theorem

Theorem (Drift Theorem (Upper Bound))

Let A be some evolutionary algorithm, P_t its t-th population, f some function, Z the set of all possible populations, $d\colon Z \to \mathbb{R}_0^+$ some distance measure with
$d(P) = 0 \Leftrightarrow P$ *contains an optimum of* f,
$M = \max\{d(P) \mid P \in Z\}$, $D_t := d(P_{t-1}) - d(P_t)$,
$\Delta := \min\{E(D_t \mid T \geq t) \mid t \in \mathbb{N}_0\}$.
$\Delta > 0 \Rightarrow E(T_{A,f}) \leq M/\Delta$

Proof

Observe $M \geq \mathrm{E}\left(\sum_{t=1}^{T} D_t\right)$

Proof of the Drift Theorem (Upper Bound)

$$\begin{aligned}
M &\geq \mathrm{E}\left(\sum_{t=1}^{T} D_t\right) = \sum_{t=1}^{\infty} \mathrm{Prob}\left(T = t\right) \cdot \mathrm{E}\left(\sum_{i=1}^{T} D_i \mid T = t\right) \\
&= \sum_{t=1}^{\infty} \mathrm{Prob}\left(T = t\right) \cdot \sum_{i=1}^{t} \mathrm{E}\left(D_i \mid T = t\right) \\
&= \sum_{t=1}^{\infty} \sum_{i=1}^{t} \mathrm{Prob}\left(T = t\right) \cdot \mathrm{E}\left(D_i \mid T = t\right) \\
&= \sum_{i=1}^{\infty} \sum_{t=i}^{\infty} \mathrm{Prob}\left(T = t\right) \cdot \mathrm{E}\left(D_i \mid T = t\right)
\end{aligned}$$

Proof of the Drift Theorem (Upper Bound) (cont.)

$$\begin{aligned}
M &\geq \sum_{i=1}^{\infty} \sum_{t=i}^{\infty} \mathrm{Prob}\left(T = t\right) \cdot \mathrm{E}\left(D_i \mid T = t\right) \\
&= \sum_{i=1}^{\infty} \sum_{t=i}^{\infty} \mathrm{Prob}\left(T \geq i\right) \cdot \mathrm{Prob}\left(T = t \mid T \geq i\right) \cdot \mathrm{E}\left(D_i \mid T = t\right) \\
&= \sum_{i=1}^{\infty} \mathrm{Prob}\left(T \geq i\right) \sum_{t=i}^{\infty} \mathrm{Prob}\left(T = t \mid T \geq i\right) \cdot \mathrm{E}\left(D_i \mid T = t \wedge T \geq i\right) \\
&= \sum_{i=1}^{\infty} \mathrm{Prob}\left(T \geq i\right) \sum_{t=1}^{\infty} \mathrm{Prob}\left(T = t \mid T \geq i\right) \cdot \mathrm{E}\left(D_i \mid T = t \wedge T \geq i\right) \\
&= \sum_{i=1}^{\infty} \mathrm{Prob}\left(T \geq i\right) \mathrm{E}\left(D_i \mid T \geq i\right) \geq \Delta \cdot \sum_{i=1}^{\infty} \mathrm{Prob}\left(T \geq i\right) = \Delta \cdot \mathrm{E}\left(T\right)
\end{aligned}$$

thus $\mathrm{E}\left(T\right) \leq \frac{M}{\Delta}$ □

LeadingOnes Using the Drift Theorem

Remember $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\textsc{LeadingOnes}}\right) = O(n^2)$
using f-based partitions

Definition $d(x) := n - \textsc{LeadingOnes}(x)$

Observe $M = \max\left\{d(x) \mid x \in \{0,1\}^n\right\} = n$

Observe $\mathrm{E}\left(d(x_{t-1}) - d(x_t) \mid T > t\right) \geq 1 \cdot \frac{1}{n} \cdot \left(1 - \frac{1}{n}\right)^{n-1} \geq \frac{1}{en}$

Thus $\mathrm{E}\left(T\right) \leq \frac{n}{1/en} = en^2$

same result Is there no advantage?

Advantage being more general and applicable

Example f-based partitions not applicable
for comma selection

$(1, n)$ EA and LeadingOnes

Theorem $\mathsf{E}\left(T_{(1,\,n)\text{ EA},\text{LeadingOnes}}\right) = O(n^2)$

Proof with drift analysis
$d(x) := n - \text{LeadingOnes}(x)$ thus $M = n$

$$\begin{aligned} & \mathsf{E}\left(d(x_{t-1}) - d(x_t) \mid T > t\right) \\ \geq\ & 1 \cdot \left(1 - \left(1 - \frac{1}{en}\right)^n\right) - n \cdot \left(1 - \left(1 - \frac{1}{n}\right)^n\right)^n \\ =\ & \Omega(1) \end{aligned}$$

thus $\mathsf{E}(T) = O(n)$
thus $\mathsf{E}\left(T_{(1,\,n)\text{ EA},\text{LeadingOnes}}\right) = n \cdot \mathsf{E}(T) = O(n^2)$ □

Another Drift Theorem

Remember distance $d\colon Z \to \mathbb{R}_0^+$ with $d(P) = 0 \Leftrightarrow P$ optimal
$M := \max\{d(P) \mid P \in Z\}$, $D_t := d(P_{t-1}) - d(P_t)$
$\Delta := \min\{\mathsf{E}(D_t \mid T \geq t) \mid t \in \mathbb{N}_0\}$
$\Delta > 0 \Rightarrow \mathsf{E}(T) \leq \frac{M}{\Delta}$

Observe M can be replaced by $\mathsf{E}(d(P_0))$ ✔

In addition

Theorem Let $d\colon Z \to \mathbb{N}_0$ be distance, rest as before.
$\exists c \in \mathbb{R}^+\colon \forall P_{t-1}\colon \mathsf{E}(d(P_{t-1}) - d(P_t) \mid P_t) \geq \frac{d(P_{t-1})}{c}$
$\Rightarrow \mathsf{E}(T) \leq c \cdot \mathsf{E}\left(H_{d(P_0)}\right)$

Proof idea Apply drift theorem to $d' := H_d$.

Proving the Logarithmic Drift Theorem

Theorem Let $d\colon Z \to \mathbb{N}_0$ be distance, rest as before.
$\exists c \in \mathbb{R}^+\colon \forall P_{t-1}\colon \mathsf{E}(d(P_{t-1}) - d(P_t) \mid P_t) \geq \frac{d(P_{t-1})}{c}$
$\Rightarrow \mathsf{E}(T) \leq c \cdot \mathsf{E}\left(H_{d(P_0)}\right)$

Proof Observe $H_{d(P)} = 0 \Leftrightarrow d(P) = 0$ ✔

Compute $H_k - H_l = \sum\limits_{i=1}^{k} \frac{1}{i} - \sum\limits_{i=1}^{l} \frac{1}{i}$
$= \sum\limits_{i=l+1}^{k} \frac{1}{i} \geq \frac{k-l}{k}$

thus $\mathsf{E}\left(H_{d(P_{t-1})} - H_{d(P_t)} \mid P_{t-1}\right) \geq \mathsf{E}\left(\frac{d(P_{t-1}) - d(P_t)}{d(P_{t-1})} \mid P_{t-1}\right)$
$= \frac{\mathsf{E}(d(P_{t-1}) - d(P_t) \mid P_{t-1})}{d(P_{t-1})} \geq \frac{1}{c}$

thus $\mathsf{E}(T) \leq c \cdot \mathsf{E}\left(H_{d(P_0)}\right)$ □