On the Brittleness of Evolutionary Algorithms

Thomas Jansen

Fachbereich Informatik
Lehrstuhl für Effiziente Algorithmen und Komplexitätstheorie
Universität Dortmund
Germany

Motivation

Evolutionary algorithms are algorithms.

Motivation

Evolutionary algorithms are algorithms.

Consequence

Evolutionary Algorithms should be analyzed like algorithms are.

Motivation

Evolutionary algorithms are algorithms.

Consequence

Evolutionary Algorithms should be analyzed like algorithms are.

Analysis of (Randomized) Algorithms

1. correctness
2. (expected) run time

Motivation

Evolutionary algorithms are algorithms.

Consequence

Evolutionary Algorithms should be analyzed like algorithms are.

Analysis of (Randomized) Algorithms

1. correctness
2. (expected) run time

Evolutionary algorithms are function optimizers
(among other things). to Ken: Sorry!

Motivation

Evolutionary algorithms are algorithms.

Consequence

Evolutionary Algorithms should be analyzed like algorithms are.

Analysis of (Randomized) Algorithms

1. correctness
2. (expected) run time

Evolutionary algorithms are function optimizers
(among other things). to Ken: Sorry!

Analysis of Evolutionary Algorithms

1. convergence to global optimum √
2. expected optimization time

What Evolutionary Algorithm Should Be Analyzed?

What Evolutionary Algorithm Should Be Analyzed?

Long Term Goal all kinds of evolutionary algorithms

What Evolutionary Algorithm Should Be Analyzed?

Long Term Goal all kinds of evolutionary algorithms

Realistic not too complicated

What Evolutionary Algorithm Should Be Analyzed?

Long Term Goal all kinds of evolutionary algorithms

Realistic not too complicated

Observation in theory, analysis of extreme cases makes sense

What Evolutionary Algorithm Should Be Analyzed?

Long Term Goal all kinds of evolutionary algorithms

Realistic not too complicated

Observation in theory, analysis of extreme cases makes sense

Example infinite population models to Adam: Sorry!

What Evolutionary Algorithm Should Be Analyzed?

Long Term Goal all kinds of evolutionary algorithms

Realistic not too complicated

Observation in theory, analysis of extreme cases makes sense

Example infinite population models to Adam: Sorry!

Example population size 1

The (1+1) EA

for maximization of $f\colon \{0,1\}^n \to \mathbb{R}$

1. Initialization

   $t := 0$; Choose $x_t \in \{0,1\}^n$ uniformly at random.
2. Mutation

   $y := x_t$; Independently for each bit in y, flip this bit with probability $p_m = 1/n$.
3. Selection

   If $f(y) \geq f(x_t)$ Then $x_{t+1} := y$ Else $x_{t+1} := x_t$.
4. "Stopping Criterion"

   $t := t + 1$; Continue at line 2.

The (1+1) EA

for maximization of $f : \{0,1\}^n \to \mathbb{R}$

1. **Initialization**
   $t := 0$; Choose $x_t \in \{0,1\}^n$ uniformly at random.
2. **Mutation**
   $y := x_t$; Independently for each bit in y, flip this bit with probability $p_m = 1/n$.
3. **Selection**
   If $f(y) \geq f(x_t)$ Then $x_{t+1} := y$ Else $x_{t+1} := x_t$.
4. **"Stopping Criterion"**
   $t := t + 1$; Continue at line 2.

Is this an evolutionary algorithm?

The (1+1) EA

for maximization of $f: \{0,1\}^n \to \mathbb{R}$

1. Initialization
   $t := 0$; Choose $x_t \in \{0,1\}^n$ uniformly at random.
2. Mutation
   $y := x_t$; Independently for each bit in y, flip this bit with probability $p_m = 1/n$.
3. Selection
   If $f(y) \geq f(x_t)$ Then $x_{t+1} := y$ Else $x_{t+1} := x_t$.
4. "Stopping Criterion"
   $t := t+1$; Continue at line 2.

Is this an evolutionary algorithm?

Yes, because of

- global mutation
- fitness-based selection

The (1+1) EA for maximization of $f\colon \{0,1\}^n \to \mathbb{R}$

1. Initialization
   $t := 0$; Choose $x_t \in \{0,1\}^n$ uniformly at random.
2. Mutation
   $y := x_t$; Independently for each bit in y, flip this bit with probability $p_m = 1/n$.
3. Selection
   If $f(y) \geq f(x_t)$ Then $x_{t+1} := y$ Else $x_{t+1} := x_t$.
4. "Stopping Criterion"
   $t := t + 1$; Continue at line 2.

Is this an evolutionary algorithm?

Yes, because of

- global mutation
- fitness-based selection

Counter-Examples

- (1, 1) EA since selection independent of fitness
- "(1+1) EA" with 1-bit-mutation since no global search

What Fitness Function Should Be Analyzed?

What Fitness Function Should Be Analyzed?

Long Term Goal any classes of fitness functions

What Fitness Function Should Be Analyzed?

Long Term Goal any classes of fitness functions

Realistic not too complicated

What Fitness Function Should Be Analyzed?

Long Term Goal any classes of fitness functions

Realistic not too complicated

Observation toy problems (artificial examples) make sense

- for development of methods and analytical tools
- for illustration purposes
- as counter-examples
- as examples
- for teaching purposes

What Fitness Function Should Be Analyzed?

Long Term Goal any classes of fitness functions

Realistic not too complicated

Observation toy problems (artificial examples) make sense

- for development of methods and analytical tools
- for illustration purposes
- as counter-examples
- as examples
- for teaching purposes

Answers Darrell's open question Can we learn from toy problems?
YES!

What Fitness Function Should Be Analyzed?

Long Term Goal any classes of fitness functions

Realistic not too complicated

Observation toy problems (artificial examples) make sense

- for development of methods and analytical tools
- for illustration purposes
- as counter-examples
- as examples
- for teaching purposes

Answers Darrell's open question Can we learn from toy problems?
YES!

Very well-known example $\text{ONEMAX}(x) = \sum_{i=1}^{n} x[i]$

Selected Aspects of the History of EC Theory

Selected Aspects of the History of EC Theory

Heinz Mühlenbein (1992)

How evolutionary algorithms really work:
mutation and hillclimbing. (PPSN)

Selected Aspects of the History of EC Theory

Heinz Mühlenbein (1992)

How evolutionary algorithms really work:
mutation and hillclimbing. (PPSN)

Theorem $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\textsc{OneMax}}\right) = O(n \log n)$

Selected Aspects of the History of EC Theory

Heinz Mühlenbein (1992)

How evolutionary algorithms really work:
mutation and hillclimbing. (PPSN)

Theorem $\quad \mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\textsc{OneMax}}\right) = O(n \log n)$

Observation $\quad \forall \text{linear } f\colon \mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Selected Aspects of the History of EC Theory

Heinz Mühlenbein (1992)

How evolutionary algorithms really work:
mutation and hillclimbing. (PPSN)

Theorem $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\textsc{OneMax}}\right) = O(n \log n)$

Observation $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Observation $\forall$unimodal f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Selected Aspects of the History of EC Theory

Heinz Mühlenbein (1992)

How evolutionary algorithms really work:
mutation and hillclimbing. (PPSN)

Theorem $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\textsc{OneMax}}\right) = O(n \log n)$

Observation $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Observation $\forall$unimodal f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

What are linear functions? What are unimodal functions?

Selected Aspects of the History of EC Theory

Heinz Mühlenbein (1992)

How evolutionary algorithms really work: mutation and hillclimbing. (PPSN)

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

Observation $\forall$linear $f\colon E\left(T_{(1+1)\ \text{EA},f}\right) = O(n \log n)$

Observation $\forall$unimodal $f\colon E\left(T_{(1+1)\ \text{EA},f}\right) = O(n \log n)$

What are linear functions? What are unimodal functions?

Definition $f\colon \{0,1\}^n \to \mathbb{R}$ linear

$$:\Leftrightarrow \exists w_0, w_1, \dots, w_n \in \mathbb{R}\colon \forall x \in \{0,1\}^n\colon f(x) = w_0 + \sum_{i=1}^{n} w_i \cdot x[i]$$

(w. l. o. g. $w_0 = 0,\ w_1 \geq w_2 \geq \cdots \geq w_n$)

Selected Aspects of the History of EC Theory

Heinz Mühlenbein (1992)

How evolutionary algorithms really work:
mutation and hillclimbing. (PPSN)

Theorem $\mathrm{E}\left(T_{(1+1)\ \text{EA},\textsc{OneMax}}\right) = O(n \log n)$

Observation $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \text{EA},f}\right) = O(n \log n)$

Observation $\forall$unimodal f: $\mathrm{E}\left(T_{(1+1)\ \text{EA},f}\right) = O(n \log n)$

What are linear functions? What are unimodal functions?

Definition $f\colon \{0,1\}^n \to \mathbb{R}$ linear

$:\Leftrightarrow \exists w_0, w_1, \dots, w_n \in \mathbb{R}\colon \forall x \in \{0,1\}^n\colon f(x) = w_0 + \sum_{i=1}^{n} w_i \cdot x[i]$

(w. l. o. g. $w_0 = 0,\ w_1 \geq w_2 \geq \cdots \geq w_n$)

Definition $f\colon \{0,1\}^n \to \mathbb{R}$ unimodal

$:\Leftrightarrow \forall x \in \{0,1\}^n\colon f(x)$ max. $\vee\ \exists y\colon \mathrm{H}(x,y) = 1 \wedge f(y) > f(x)$

Selected Aspects of the History of EC Theory

Heinz Mühlenbein (1992)

How evolutionary algorithms really work:
mutation and hillclimbing. (PPSN)

Theorem $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\textsc{OneMax}}\right) = O(n \log n)$

Observation $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Observation $\forall$unimodal f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Selected Aspects of the History of EC Theory

Heinz Mühlenbein (1992)

How evolutionary algorithms really work: mutation and hillclimbing. (PPSN)

Theorem $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},\textsc{OneMax}}\right) = O(n \log n)$

Observation $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Observation $\forall$unimodal f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Droste, Jansen, Wegener (1998)

A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Selected Aspects of the History of EC Theory

Heinz Mühlenbein (1992)
How evolutionary algorithms really work:
mutation and hillclimbing. (PPSN)

Theorem $\mathrm{E}\left(T_{(1+1)\text{ EA},\textsc{OneMax}}\right) = O(n \log n)$

Observation $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\text{ EA},f}\right) = O(n \log n)$

Observation $\forall$unimodal f: $\mathrm{E}\left(T_{(1+1)\text{ EA},f}\right) = O(n \log n)$

Droste, Jansen, Wegener (1998)
A rigorous complexity analysis of the (1+1) evolutionary
algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\text{ EA},f}\right) = O(n \log n)$

Droste, Jansen, Wegener (1998)
On the optimization of unimodal functions with
the (1+1) evolutionary algorithm. (PPSN)

Theorem $\exists$unimodal f: $\mathrm{E}\left(T_{(1+1)\text{ EA},f}\right) = \Theta\left(n^{3/2} \cdot 2^{\sqrt{n}}\right)$

A Bit More on (1+1) EA for Linear Functions

Droste, Jansen, Wegener (1998)

A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall$linear f : $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

A Bit More on (1+1) EA for Linear Functions

Droste, Jansen, Wegener (1998)

A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather complicated, long, and tedious.

A Bit More on (1+1) EA for Linear Functions

Droste, Jansen, Wegener (1998)

A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather complicated, long, and tedious.

Important first step

Analysis of $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA}}, \textsc{BinVal}\right)$

with $\textsc{BinVal}(x) := \sum_{i=1}^{n} 2^{n-i} \cdot x[i]$

A Bit More on (1+1) EA for Linear Functions

Droste, Jansen, Wegener (1998)

A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather complicated, long, and tedious.

Important first step

Analysis of $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA}}, \textsc{BinVal}\right)$

with $\textsc{BinVal}(x) := \sum_{i=1}^{n} 2^{n-i} \cdot x[i]$

Property of BinVal

left-most mutated bit decides about acceptance

A Bit More on (1+1) EA for Linear Functions

Droste, Jansen, Wegener (1998)

A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall \text{linear } f \colon \mathrm{E}\left(T_{(1+1)\ \mathrm{EA}, f}\right) = O(n \log n)$

Proof rather complicated, long, and tedious.

Important first step

Analysis of $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA}}, \textsc{BinVal}\right)$

with $\textsc{BinVal}(x) := \sum_{i=1}^{n} 2^{n-i} \cdot x[i]$

Property of BinVal

left-most mutated bit decides about acceptance

Extreme Case

$0111 \cdots 11 \rightsquigarrow 1000 \cdots 00$

Even More on (1+1) EA for Linear Functions

Droste, Jansen, Wegener (1998)

A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather complicated, long, and tedious.

Jun He, Xin Yao (2001)

Drift analysis and average time complexity of evolutionary algorithms. (AI)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Even More on (1+1) EA for Linear Functions

Droste, Jansen, Wegener (1998)

A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather complicated, long, and tedious.

Jun He, Xin Yao (2001)

Drift analysis and average time complexity of evolutionary algorithms. (AI)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather simple and very elegant

due to powerful general method drift analysis

Still More on (1+1) EA for Linear Functions

Droste, Jansen, Wegener (1998)

A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather complicated, long, and tedious.

Jun He, Xin Yao (2001)

Drift analysis and average time complexity of evolutionary algorithms. (AI)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather simple and very elegant

Still More on (1+1) EA for Linear Functions

Droste, Jansen, Wegener (1998)

A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather complicated, long, and tedious.

Jun He, Xin Yao (2001)

Drift analysis and average time complexity of evolutionary algorithms. (AI)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather simple and very elegant

Jun He, Xin Yao (2002)

Erratum to: Drift analysis and average time complexity of evolutionary algorithms. (AI)

Still More on (1+1) EA for Linear Functions

Droste, Jansen, Wegener (1998)

A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with Boolean inputs. (CEC)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather complicated, long, and tedious.

Jun He, Xin Yao (2001)

Drift analysis and average time complexity of evolutionary algorithms. (AI)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof rather simple and very elegant

Jun He, Xin Yao (2002)

Erratum to: Drift analysis and average time complexity of evolutionary algorithms. (AI)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

using mutation probability $p_m = 1/(2n)$

Final Word on (1+1) EA for Linear Functions

Final Word on (1+1) EA for Linear Functions

Jun He, Xin Yao (2004)

A study of drift analysis for estimating computation time of evolutionary algorithms. (Natural Computing)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Final Word on (1+1) EA for Linear Functions

Jun He, Xin Yao (2004)

A study of drift analysis for estimating computation time of evolutionary algorithms. (Natural Computing)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof a bit more complicated but still elegant
combining central idea from Droste/Jansen/Wegener (1998) with drift analysis

Final Word on (1+1) EA for Linear Functions

Jun He, Xin Yao (2004)

A study of drift analysis for estimating computation time of evolutionary algorithms. (Natural Computing)

Theorem $\forall$linear f: $E\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$

Proof a bit more complicated but still elegant
combining central idea from Droste/Jansen/Wegener (1998) with drift analysis

What happened in the papers He/Yao (2001/2002)?

Final Word on (1+1) EA for Linear Functions

Jun He, Xin Yao (2004)

A study of drift analysis for estimating computation time of evolutionary algorithms. (Natural Computing)

Theorem $\forall$linear f: $E\left(T_{(1+1)\text{ EA},f}\right) = O(n \log n)$

Proof a bit more complicated but still elegant
combining central idea from Droste/Jansen/Wegener (1998)
with drift analysis

What happened in the papers He/Yao (2001/2002)?

Remember BinVal

Mutation with 1 bit flipping $0 \rightsquigarrow 1$, all other bits flipping $1 \rightsquigarrow 0$ may be accepted.

Mutation with 1 bit flipping $1 \rightsquigarrow 0$, all other bits flipping $0 \rightsquigarrow 1$ may be rejected.

Final Word on (1+1) EA for Linear Functions

Jun He, Xin Yao (2004)

A study of drift analysis for estimating computation time of evolutionary algorithms. (Natural Computing)

Theorem $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA}, f}\right) = O(n \log n)$

Proof a bit more complicated but still elegant
combining central idea from Droste/Jansen/Wegener (1998)
with drift analysis

What happened in the papers He/Yao (2001/2002)?

Remember BinVal

Mutation with 1 bit flipping $0 \rightsquigarrow 1$, all other bits flipping $1 \rightsquigarrow 0$ may be accepted.

Mutation with 1 bit flipping $1 \rightsquigarrow 0$, all other bits flipping $0 \rightsquigarrow 1$ may be rejected.

Worst Case Assumption This is always the case.

The Model — A bit more formal...

Definition partial order on $\{0,1\}^n$

$$x \leq y :\Leftrightarrow \forall i \in \{1,2,\ldots,n\}\colon x[i] \leq y[i]$$

The Model — A bit more formal...

Definition partial order on $\{0,1\}^n$

$x \leq y :\Leftrightarrow \forall i \in \{1,2,\ldots,n\} : x[i] \leq y[i]$

PO-EA

1. **Initialization**
   $t := 0$; Choose $x_t \in \{0,1\}^n$ uniformly at random.
2. **Mutation**
   $y := x_t$; Independently for each bit in y, flip this bit with probability $p_m = 1/n$.
3. **Selection**
   If $(x_t \leq y) \vee (\neg(y \leq x_t) \wedge (\text{ONEMAX}(y) \leq \text{ONEMAX}(x_t)))$
   Then $x_{t+1} := y$ Else $x_{t+1} := x_t$.
4. **"Stopping Criterion"**
   $t := t+1$; Continue at line 2.

PO-EA Some Observations

Is identical with (1+1) EA on linear functions for "pure" mutations.

Is identical with (1+1) EA on linear functions for 1-bit mutation.

PO-EA Some Observations

Is identical with (1+1) EA on linear functions for “pure” mutations.

Is identical with (1+1) EA on linear functions for 1-bit mutation.

PO-EA is not identical with (1+1) EA on any function:

$$0101 \underbrace{\rightsquigarrow}_{(\neg\leq)\wedge(\textsc{OneMax})} 0010 \underbrace{\rightsquigarrow}_{\text{pure}} 0110 \underbrace{\rightsquigarrow}_{(\neg\leq)\wedge(\textsc{OneMax})} 0001 \underbrace{\rightsquigarrow}_{\text{pure}} 0101$$

PO-EA Some Observations

Is identical with (1+1) EA on linear functions for "pure" mutations.

Is identical with (1+1) EA on linear functions for 1-bit mutation.

PO-EA is not identical with (1+1) EA on any function:

$$0101 \underbrace{\rightsquigarrow}_{(\neg\leq)\wedge(\textsc{OneMax})} 0010 \underbrace{\rightsquigarrow}_{\text{pure}} 0110 \underbrace{\rightsquigarrow}_{(\neg\leq)\wedge(\textsc{OneMax})} 0001 \underbrace{\rightsquigarrow}_{\text{pure}} 0101$$

$111\cdots111$ is only stable state

Definition optimization time = first hitting time of 1^n

PO-EA Some Observations

Is identical with (1+1) EA on linear functions for "pure" mutations.

Is identical with (1+1) EA on linear functions for 1-bit mutation.

PO-EA is not identical with (1+1) EA on any function:

$$0101 \underbrace{\rightsquigarrow}_{(\neg\leq)\wedge(\textsc{OneMax})} 0010 \underbrace{\rightsquigarrow}_{\text{pure}} 0110 \underbrace{\rightsquigarrow}_{(\neg\leq)\wedge(\textsc{OneMax})} 0001 \underbrace{\rightsquigarrow}_{\text{pure}} 0101$$

$111\cdots111$ is only stable state

Definition optimization time = first hitting time of 1^n

Optimization time of PO-EA is upper bound
for $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right)$ for linear f
as observed by Jun He and Xin Yao

Drift Analysis

Central Notion distance

for Z = set of all populations ($Z = \{0,1\}^n$ for (1+1) EA)

define distance $d\colon Z \to \mathbb{R}_0^+$
with $d(P) = 0 \Leftrightarrow P$ contains optimum

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

Drift Analysis

Central Notion distance

for Z = set of all populations ($Z = \{0,1\}^n$ for (1+1) EA)

define distance $d \colon Z \to \mathbb{R}_0^+$
with $d(P) = 0 \Leftrightarrow P$ contains optimum

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

Consider maximal distance $M := \max\{d(P) \mid P \in Z\}$
and decreasement in distance $D_t := d(P_{t-1}) - d(P_t)$

Drift Analysis

Central Notion distance

for Z = set of all populations ($Z = \{0,1\}^n$ for (1+1) EA)

define distance $d\colon Z \to \mathbb{R}_0^+$
with $d(P) = 0 \Leftrightarrow P$ contains optimum

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

Consider maximal distance $M := \max\{d(P) \mid P \in Z\}$
and decreasement in distance $D_t := d(P_{t-1}) - d(P_t)$

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

Drift Analysis

Central Notion distance

for Z = set of all populations ($Z = \{0,1\}^n$ for (1+1) EA)
define distance $d\colon Z \to \mathbb{R}_0^+$
with $d(P) = 0 \Leftrightarrow P$ contains optimum

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

Consider maximal distance $M := \max\{d(P) \mid P \in Z\}$
and decreasement in distance $D_t := d(P_{t-1}) - d(P_t)$

Definition $\mathrm{E}\left(D_t \mid T \geq t\right)$ is called drift.

Worst Case Perspective $\Delta := \min\left\{\mathrm{E}\left(D_t \mid T \geq t\right) \mid t \in \mathbb{N}_0\right\}$

Drift Analysis

Central Notion distance

for Z = set of all populations ($Z = \{0,1\}^n$ for (1+1) EA)

define distance $d\colon Z \to \mathbb{R}_0^+$
with $d(P) = 0 \Leftrightarrow P$ contains optimum

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

Consider maximal distance $M := \max\{d(P) \mid P \in Z\}$
and decreasement in distance $D_t := d(P_{t-1}) - d(P_t)$

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

Worst Case Perspective $\Delta := \min\{\mathrm{E}(D_t \mid T \geq t) \mid t \in \mathbb{N}_0\}$

Drift Theorem (He/Yao (2001))
$\Delta > 0 \Rightarrow \mathrm{E}(T) \leq M/\Delta$

Proof of the Drift Theorem

Drift Theorem $\Delta > 0 \Rightarrow \mathrm{E}\left(T\right) \leq \frac{M}{\Delta}$

Proof

Proof of the Drift Theorem

Drift Theorem $\Delta > 0 \Rightarrow \mathrm{E}\left(T\right) \leq \frac{M}{\Delta}$

Proof

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

Proof of the Drift Theorem

Drift Theorem $\Delta > 0 \Rightarrow \mathrm{E}\left(T\right) \leq \frac{M}{\Delta}$

Proof

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

Proof of the Drift Theorem

Drift Theorem $\Delta > 0 \Rightarrow \mathrm{E}(T) \leq \frac{M}{\Delta}$

Proof

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

$$= \sum_{t=1}^{\infty} \mathrm{Prob}(T = t) \cdot \sum_{i=1}^{t} \mathrm{E}(D_i \mid T = t)$$

Proof of the Drift Theorem

Drift Theorem $\Delta > 0 \Rightarrow \mathrm{E}(T) \leq \frac{M}{\Delta}$

Proof

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

Proof of the Drift Theorem

Drift Theorem $\Delta > 0 \Rightarrow \mathrm{E}\left(T\right) \leq \frac{M}{\Delta}$

Proof

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

Proof of the Drift Theorem (cont.)

$$M \geq \sum_{i=1}^{\infty} \sum_{t=i}^{\infty} \mathrm{Prob}\,(T = t) \cdot \mathrm{E}\,(D_i \mid T = t)$$

Proof of the Drift Theorem (cont.)

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

Proof of the Drift Theorem (cont.)

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

Proof of the Drift Theorem (cont.)

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

Proof of the Drift Theorem (cont.)

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

Proof of the Drift Theorem (cont.)

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

Proof of the Drift Theorem (cont.)

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

Proof of the Drift Theorem (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 $\mathrm{E}\left(T\right) \leq \frac{M}{\Delta}$ □

Drift Theorem for Lower Bounds

We made use of

- $M = \max\{d(P) \mid P \in Z\}$
- $\Delta_l = \min\{\mathrm{E}\left(d(P_{t-1}) - d(P_t) \mid T \geq t\right)\}$

and had

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

Changes for a Lower Bound Method

Observation just two inequalities "in wrong direction"

1. $M \geq \sum \cdots$ mit $M = \max d(P) \mid P \in Z$
2. $\sum \cdots \geq \Delta_l \cdot \sum \cdots$ mit
   $\Delta_l = \min\{\mathrm{E}\left(d(P_{t-1}) - d(P_t) \mid T \geq t\right)\}$

Changes for a Lower Bound Method

Observation just two inequalities “in wrong direction”

1. $M \geq \sum \cdots$ mit $M = \max d(P) \mid P \in Z$
2. $\sum \cdots \geq \Delta_l \cdot \sum \cdots$ mit
   $\Delta_l = \min\{\mathrm{E}\left(d(P_{t-1}) - d(P_t) \mid T \geq t\right)\}$

obviously for lower bound

- $\Delta_u = \max\{\mathrm{E}\left(d(P_{t-1}) - d(P_t) \mid T \geq t\right)\}$
  sufficient for “$\leq$”
- instead of M we need lower bound M_u on $d(P_0)$

Changes for a Lower Bound Method

Observation just two inequalities "in wrong direction"

1. $M \geq \sum \cdots$ mit $M = \max d(P) \mid P \in Z$
2. $\sum \cdots \geq \Delta_l \cdot \sum \cdots$ mit
   $\Delta_l = \min\{\mathrm{E}\left(d(P_{t-1}) - d(P_t) \mid T \geq t\right)\}$

obviously for lower bound

- $\Delta_u = \max\{\mathrm{E}\left(d(P_{t-1}) - d(P_t) \mid T \geq t\right)\}$
  sufficient for "$\leq$"
- instead of M we need lower bound M_u on $d(P_0)$

this yields

- $\mathrm{E}\left(T \mid d(P_0) \geq M_u\right) \geq M_u/\Delta_u$
- $\mathrm{E}\left(T\right) \geq \mathrm{Prob}\left(d(P_0) \geq M_u\right) \cdot \mathrm{E}\left(T \mid d(P_0) \geq M_u\right) \geq$
  $\mathrm{Prob}\left(d(P_0) \geq M_u\right) \cdot M_u/\Delta_u$
- $\mathrm{E}\left(T\right) \geq \sum \mathrm{Prob}\left(d(P_0) \geq d\right) \cdot d/\Delta_u \geq \mathrm{E}\left(d(P_0)\right)/\Delta_u$

Results (of different kinds)

when using drift analysis for the optimization time of PO-EA

What distance function do we use?

Results (of different kinds)

when using drift analysis for the optimization time of PO-EA

What distance function do we use?

simplest $d(x) := n - \textsc{OneMax}(x)$

Results (of different kinds)

when using drift analysis for the optimization time of PO-EA

What distance function do we use?

simplest $d(x) := n - \text{OneMax}(x)$

Theorem

$$\mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \text{OneMax}(x_t) = n - z\right) = \Theta\left(\left(\frac{z}{n}\right)^2\right)$$

Results (of different kinds)

when using drift analysis for the optimization time of PO-EA

What distance function do we use?

simplest $d(x) := n - \textsc{OneMax}(x)$

Theorem

$\mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \textsc{OneMax}(x_t) = n - z\right) = \Theta\left(\left(\frac{z}{n}\right)^2\right)$

Sketch of Proof

$$\mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \textsc{OneMax}(x_t) = n - z\right)$$
$$= \left(\sum_{b_0=1}^{z} b_0 \binom{z}{b_0} \left(\frac{1}{n}\right)^{b_0} \left(1 - \frac{1}{n}\right)^{n-b_0}\right)$$
$$+ \sum_{b_0=1}^{z} \sum_{b_1=b_0+1}^{n-z} (b_0 - b_1) \binom{z}{b_0} \cdot \binom{n-z}{b_1} \cdot \left(\frac{1}{n}\right)^{b_0+b_1} \cdot \left(1 - \frac{1}{n}\right)^{n-b_0-b_1}$$

Sketch of Upper Bound

$$\begin{aligned}
& \mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \textsc{OneMax}(x_t) = n - z\right) \\
\leq \quad & \left(\sum_{b_0=1}^{z} b_0 \binom{z}{b_0} \left(\frac{1}{n}\right)^{b_0} \left(1 - \frac{1}{n}\right)^{n-b_0} \right) \\
& + \sum_{b_0=1}^{z} \sum_{b_1=0}^{n-z} (b_0 - b_1) \binom{z}{b_0} \binom{n-z}{b_1} \left(\frac{1}{n}\right)^{b_0+b_1} \left(1 - \frac{1}{n}\right)^{n-b_0-b_1}
\end{aligned}$$

Sketch of Upper Bound

$$
\begin{aligned}
& \mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \textsc{OneMax}(x_t) = n - z\right) \\
\leq\ & \left(\sum_{b_0=1}^{z} b_0 \binom{z}{b_0} \left(\frac{1}{n}\right)^{b_0} \left(1 - \frac{1}{n}\right)^{n-b_0}\right) \\
& \quad + \sum_{b_0=1}^{z} \sum_{b_1=0}^{n-z} (b_0 - b_1) \binom{z}{b_0} \binom{n-z}{b_1} \left(\frac{1}{n}\right)^{b_0+b_1} \left(1 - \frac{1}{n}\right)^{n-b_0-b_1} \\
=\ & \ldots \\
=\ & \frac{z}{n} - \frac{n-z}{n} + \left(1 - \frac{1}{n}\right)^{z} \frac{n-z}{n} \\
=\ & \frac{1}{n} \cdot \left(2z - n + \left(1 - \frac{1}{n}\right)^{z} (n - z)\right) \\
=\ & O\left(\left(\frac{z}{n}\right)^2\right) \quad \checkmark
\end{aligned}
$$

Sketch of Lower Bound

from proof of upper bound

$$\mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \textsc{OneMax}(x_t) = n - z\right)$$

$$= \frac{1}{n} \cdot \left(2z - n + \left(1 - \frac{1}{n}\right)^z (n - z)\right)$$

$$- \sum_{b_0=1}^{z} \sum_{b_1=0}^{b_0} (b_0 - b_1) \binom{z}{b_0} \binom{n-z}{b_1} \left(\frac{1}{n}\right)^{b_0+b_1} \left(1 - \frac{1}{n}\right)^{n-b_0-b_1}$$

Sketch of Lower Bound

from proof of upper bound

$$\mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \textsc{OneMax}(x_t) = n - z\right)$$

$$= \frac{1}{n} \cdot \left(2z - n + \left(1 - \frac{1}{n}\right)^z (n - z)\right)$$

$$- \sum_{b_0=1}^{z} \sum_{b_1=0}^{b_0} (b_0 - b_1) \binom{z}{b_0} \binom{n-z}{b_1} \left(\frac{1}{n}\right)^{b_0+b_1} \left(1 - \frac{1}{n}\right)^{n-b_0-b_1}$$

$$= \cdots = \Omega\left(\left(\frac{z}{n}\right)^2\right) \qquad \square$$

Sketch of Lower Bound

from proof of upper bound

$$\begin{aligned}
&\mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \textsc{OneMax}(x_t) = n - z\right) \\
&= \frac{1}{n} \cdot \left(2z - n + \left(1 - \frac{1}{n}\right)^z (n - z)\right) \\
&- \sum_{b_0=1}^{z} \sum_{b_1=0}^{b_0} (b_0 - b_1) \binom{z}{b_0} \binom{n-z}{b_1} \left(\frac{1}{n}\right)^{b_0+b_1} \left(1 - \frac{1}{n}\right)^{n-b_0-b_1} \\
&= \cdots = \Omega\left(\left(\frac{z}{n}\right)^2\right) \qquad \square
\end{aligned}$$

Drift Analysis leads to
- lower bound $\Omega(n)$:-(
- upper bound $O(n^3)$:-(

correct but not tight

Results for Simple d Reconsidered

With $d(x) := n - \text{ONEMAX}(x)$ we have

$$\mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \text{ONEMAX}(x_t) = n - z\right) = \Theta\left(\left(\frac{z}{n}\right)^2\right)$$

Results for Simple d Reconsidered

With $d(x) := n - \textsc{OneMax}(x)$ we have
$\mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \textsc{OneMax}(x_t) = n - z\right) = \Theta\left(\left(\frac{z}{n}\right)^2\right)$

more involved analysis following the lines of the $O(n \log n)$-proof in Droste/Jansen/Wegener (1998) leads to upper bound $O(n^2)$

Results for Simple d Reconsidered

With $d(x) := n - \text{OneMax}(x)$ we have
$\mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \text{OneMax}(x_t) = n - z\right) = \Theta\left(\left(\frac{z}{n}\right)^2\right)$

more involved analysis following the lines of the $O(n \log n)$-proof in Droste/Jansen/Wegener (1998)
leads to upper bound $O(n^2)$

a slightly more involved conditional drift analysis
using drift for $\text{OneMax}(x_t) \geq n - \sqrt{n}$ and
proving lower bound on probability for not crossing this boundary
leads to lower bound $\Omega\left(\sqrt{n} \cdot \left(\frac{n}{\sqrt{n}}\right)^2\right) = \Omega\left(n^{3/2}\right)$

Results for Simple d Reconsidered

With $d(x) := n - \text{ONEMAX}(x)$ we have
$\mathrm{E}\left(d(x_{t+1}) - d(x_t) \mid \text{ONEMAX}(x_t) = n - z\right) = \Theta\left(\left(\frac{z}{n}\right)^2\right)$

more involved analysis following the lines of the
$O(n \log n)$-proof in Droste/Jansen/Wegener (1998)
leads to upper bound $O(n^2)$

a slightly more involved conditional drift analysis
using drift for $\text{ONEMAX}(x_t) \geq n - \sqrt{n}$ and
proving lower bound on probability for not crossing this boundary
leads to lower bound $\Omega\left(\sqrt{n} \cdot \left(\frac{n}{\sqrt{n}}\right)^2\right) = \Omega\left(n^{3/2}\right)$

Observation He and Yao could not succeed in 2001
since this slight modification of the model
actually significantly increased the expected optimization time.

What is the Truth?

What is the Truth?

Consider different distance measure

$$d(x) := \sqrt{n - \textsc{OneMax}(x)}$$

What is the Truth?

Consider different distance measure

$$d(x) := \sqrt{n - \textsc{OneMax}(x)}$$

$$\mathrm{E}\left(d(x_t) - d(x_{t+1}) \mid \textsc{OneMax}(x_t) = n - z\right)$$

$$= \sum_{b_0=0}^{z} \sum_{b_1=0}^{n-z} \left(\binom{z}{b_0} \cdot \binom{n-z}{b_1} \cdot \left(\frac{1}{n}\right)^{b_0+b_1} \cdot \left(1 - \frac{1}{n}\right)^{n-b_0-b_1} \right.$$

$$\left. \cdot \mathbf{1}_{(b_1=0<b_0)\vee(b_1>b_0>0)} \cdot \left(\tilde{d}(z) - \tilde{d}(z + b_1 - b_0)\right) \right)$$

What is the Truth?

Consider different distance measure

$$d(x) := \sqrt{n - \textsc{OneMax}(x)}$$

$$\begin{aligned}
&\mathrm{E}\left(d(x_t) - d(x_{t+1}) \mid \textsc{OneMax}(x_t) = n - z\right) \\
&= \sum_{b_0=0}^{z} \sum_{b_1=0}^{n-z} \left(\binom{z}{b_0} \cdot \binom{n-z}{b_1} \cdot \left(\frac{1}{n}\right)^{b_0+b_1} \cdot \left(1 - \frac{1}{n}\right)^{n-b_0-b_1} \right. \\
&\left. \cdot \mathbf{1}_{(b_1=0<b_0)\vee(b_1>b_0>0)} \cdot \left(\tilde{d}(z) - \tilde{d}(z + b_1 - b_0)\right) \right) \\
&= \cdots = \Omega\left(\frac{\sqrt{z}}{n}\right)
\end{aligned}$$

What is the Truth?

Consider different distance measure

$$d(x) := \sqrt{n - \textsc{OneMax}(x)}$$

$$\mathrm{E}\left(d(x_t) - d(x_{t+1}) \mid \textsc{OneMax}(x_t) = n - z\right)$$
$$= \sum_{b_0=0}^{z} \sum_{b_1=0}^{n-z} \left(\binom{z}{b_0} \cdot \binom{n-z}{b_1} \cdot \left(\frac{1}{n}\right)^{b_0+b_1} \cdot \left(1 - \frac{1}{n}\right)^{n-b_0-b_1} \right.$$
$$\left. \cdot \mathbf{1}_{(b_1=0<b_0)\vee(b_1>b_0>0)} \cdot \left(\tilde{d}(z) - \tilde{d}(z + b_1 - b_0)\right) \right)$$
$$= \cdots = \Omega\left(\frac{\sqrt{z}}{n}\right)$$

Drift Analysis leads to

upper bound $O\left(n^{3/2}\right)$

What is the Truth?

Consider different distance measure

$$d(x) := \sqrt{n - \text{ONEMAX}(x)}$$

$$\begin{aligned}
&\mathrm{E}\left(d(x_t) - d(x_{t+1}) \mid \text{ONEMAX}(x_t) = n - z\right) \\
&= \sum_{b_0=0}^{z} \sum_{b_1=0}^{n-z} \left(\binom{z}{b_0} \cdot \binom{n-z}{b_1} \cdot \left(\frac{1}{n}\right)^{b_0+b_1} \cdot \left(1 - \frac{1}{n}\right)^{n-b_0-b_1} \right. \\
&\left. \cdot \mathbf{1}_{(b_1=0<b_0)\vee(b_1>b_0>0)} \cdot \left(\tilde{d}(z) - \tilde{d}(z + b_1 - b_0)\right) \right) \\
&= \cdots = \Omega\left(\frac{\sqrt{z}}{n}\right)
\end{aligned}$$

Drift Analysis leads to

upper bound $O\left(n^{3/2}\right)$

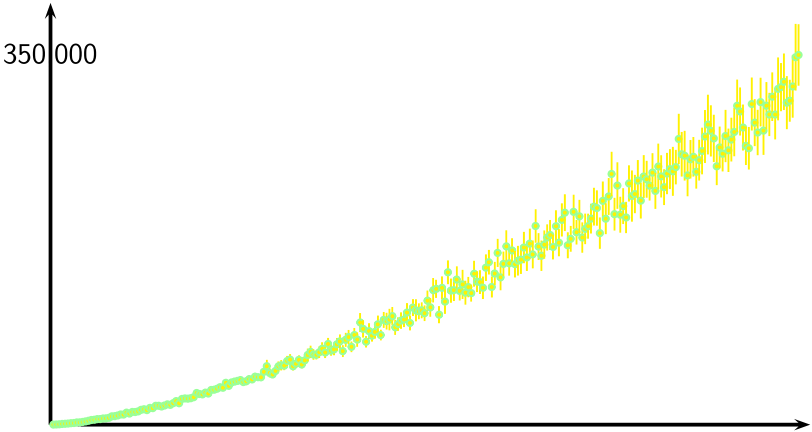
Theorem $\quad \mathrm{E}\left(T_{\text{PO-EA}}\right) = \Theta\left(n^{3/2}\right)$

Empirical Results

simple, direct implementation, 100 runs, $n \in \{10, 20, \ldots, 2560\}$

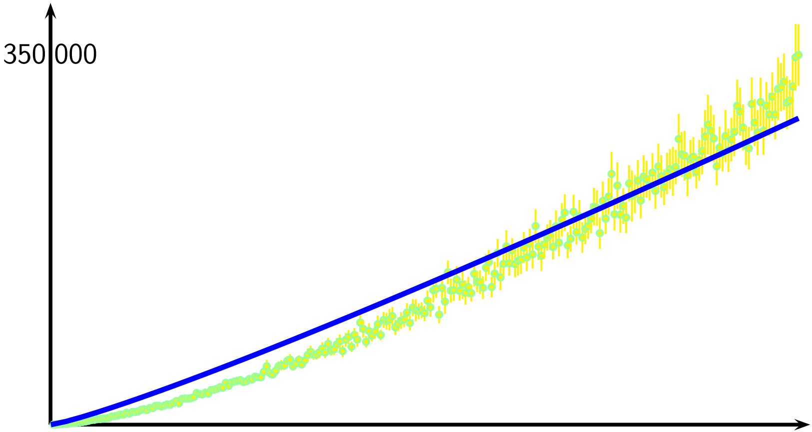
Empirical Results

simple, direct implementation, 100 runs, $n \in \{10, 20, \ldots, 2560\}$



Empirical Results

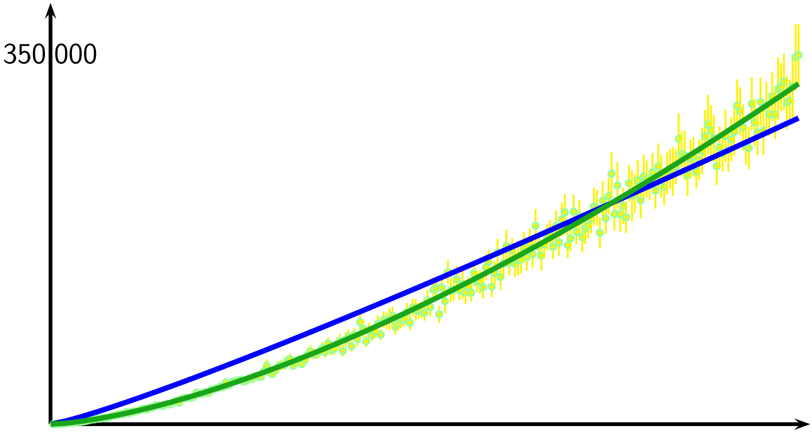
simple, direct implementation, 100 runs, $n \in \{10, 20, \ldots, 2560\}$



with $14.346 \cdot n \cdot \ln(n)$

Empirical Results

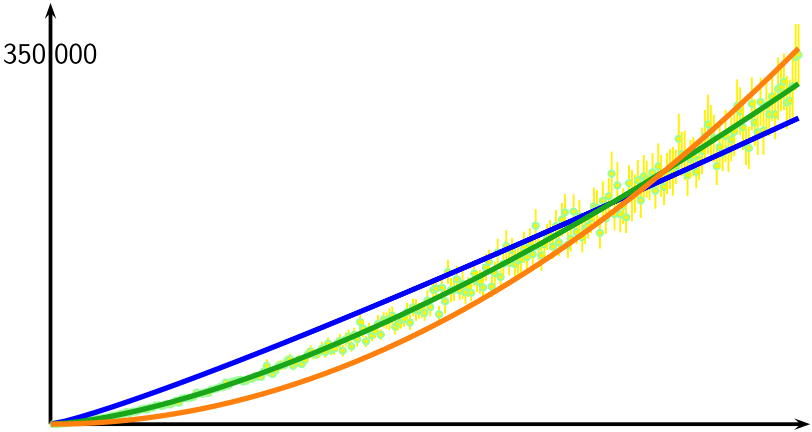
simple, direct implementation, 100 runs, $n \in \{10, 20, \ldots, 2560\}$



with $14.346 \cdot n \cdot \ln(n)$, $2.474 \cdot n^{3/2}$

Empirical Results

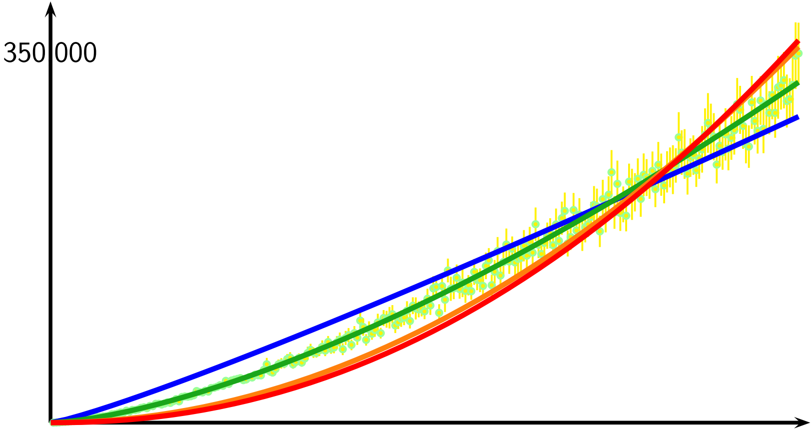
simple, direct implementation, 100 runs, $n \in \{10, 20, \ldots, 2560\}$



with $14.346 \cdot n \cdot \ln(n)$, $2.474 \cdot n^{3/2}$, $.054 \cdot n^2$

Empirical Results

simple, direct implementation, 100 runs, $n \in \{10, 20, \ldots, 2560\}$



with $14.346 \cdot n \cdot \ln(n)$, $2.474 \cdot n^{3/2}$, $.054 \cdot n^2$, $.007 \cdot n^2 \cdot \ln(n)$

Summary

- still $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$
- PO-EA abstracts from linear functions in a worst case way
- PO-EA is exact for $\textsc{BinVal}$ on $0111\cdots1$, mutations of single bits, pure mutations
- yet, there is no f such that PO-EA behaves exactly like (1+1) EA on f
- $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = \Theta(n^{3/2})$
- drift analysis is a powerful general tool

On the Choice of a Title

What has all this got to do with brittleness of evolutionary algorithms?

On the Choice of a Title

What has all this got to do with brittleness of evolutionary algorithms?

The step from the (1+1) EA on linear functions to PO-EA is quite small.

And yet the (1+1) EA's average optimization time is changed considerably.

On the Choice of a Title

What has all this got to do with brittleness of evolutionary algorithms?

The step from the (1+1) EA on linear functions to PO-EA is quite small.

And yet the (1+1) EA's average optimization time is changed considerably.

Who says that it's small?

On the Choice of a Title

What has all this got to do with brittleness of evolutionary algorithms?

The step from the (1+1) EA on linear functions to PO-EA is quite small.

And yet the (1+1) EA's average optimization time is changed considerably.

Who says that it's small? at least Jun He, Xin Yao, and me

On the Choice of a Title

What has all this got to do with brittleness of evolutionary algorithms?

The step from the (1+1) EA on linear functions to PO-EA is quite small.

And yet the (1+1) EA's average optimization time is changed considerably.

Who says that it's small? at least Jun He, Xin Yao, and me

But we now know it's not so small!

On the Choice of a Title

What has all this got to do with brittleness of evolutionary algorithms?

The step from the (1+1) EA on linear functions to PO-EA is quite small.

And yet the (1+1) EA's average optimization time is changed considerably.

Who says that it's small? at least Jun He, Xin Yao, and me

But we now know it's not so small!

PO-EA is an example where a quite small change of the mutation probability from $\frac{1}{n}$ to $\frac{1}{2n}$ changes the average optimization time considerably.

On the Brittleness of Evolutionary Algorithms

Thomas Jansen

Summary

- still $\forall$linear f: $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = O(n \log n)$
- PO-EA abstracts from linear functions in a worst case way
- PO-EA is exact for BinVal on $0111\cdots1$, mutations of single bits, pure mutations
- yet, there is no f such that PO-EA behaves exactly like (1+1) EA on f
- $\mathrm{E}\left(T_{(1+1)\ \mathrm{EA},f}\right) = \Theta(n^{3/2})$
- drift analysis is a powerful general tool
- (1+1) EA extremely sensitive to changes of p_m