Last class

- Genetic algorithms
- Evolutionary strategies
- Evolutionary programming
- Genetic programming

} Historical EA variants

- Differential evolution
- Particle swarm optimization
- Ant colony optimization
- Estimation of distribution algorithms

} Recent EA variants

Heuristic Search and Evolutionary Algorithms

Lecture 9: Theoretical Analysis of Evolutionary Algorithms

Chao Qian （钱超）

Associate Professor, Nanjing University, China

Email: qianc@nju.edu.cn
Homepage: http://www.lamda.nju.edu.cn/qianc/

Theoretical analysis

Develop solid, rigorous, and reliable knowledge

- empirical studies are limited to the experimented cases
- overcome experiment difficulties
- derive provable conclusions

Particularly for evolutionary algorithms (EAs)

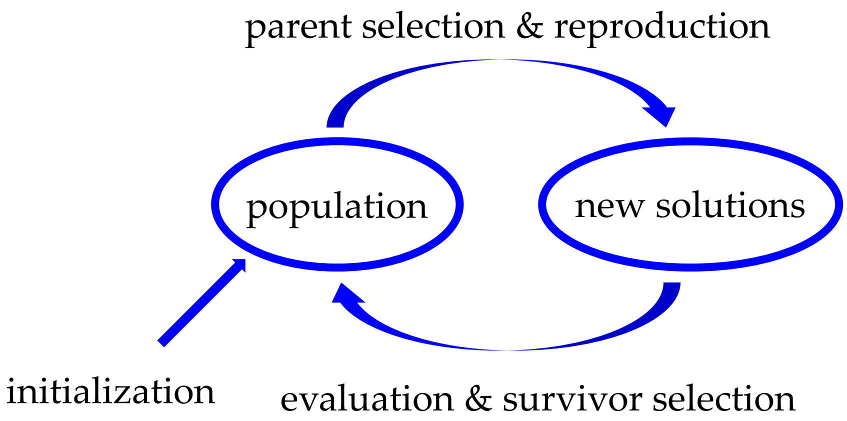
- when to use them？
- what are their merits and drawbacks?
- how different configurations affect their performance?
- design better EAs

...

Theoretical analysis of EAs

- EAs have been widely used in real applications

GA, ES, EP, GP, PSO,
ACO, DE, EDA, ……



- EAs are complex and randomized
  - The components of EAs, e.g., mutation, recombination, selection and population, can be complex
  - With the same input, the output by independent runs can be different

Theoretical analysis is very difficult

Schema theorem

Schema theorem [Holland, 1975]

- Proposed to explain how the population of EAs changes in steps

Consider a binary solution space $\{0,1\}^5 =$

00000	01000	10000	11000
00001	01001	10001	11001
00010	01010	10010	11010
00011	01011	10011	11011
00100	01100	10100	11100
00101	01101	10101	11101
00110	01110	10110	11110
00111	01111	10111	11111

A schema H is a template with "#"= "any", which defines a subspace

The order $o(H)$: the number of positions that do not have #

The defining length $d(H)$: the distance between the outermost defined positions

e.g.	$o(H)$	$d(H)$
01#1#	3	3
#1#1#	2	2
###1#	1	0

Schema theorem

Schema theorem [Holland, 1975]

- Proposed to explain how the population of EAs changes in steps

Study the change of $m(H, t)$ → the number of individuals belonging to H in the population at time t

Consider simple GA (SGA)

Representation	Binary representation
Recombination	One-point crossover
Mutation	Bit-wise mutation
Parent selection	Fitness proportional selection
Survivor selection	Generational

1. with prob. p_c, apply one-point crossover, otherwise copy them
2. for each resulting solution, apply bit-wise mutation

Schema theorem

Schema theorem [Holland, 1975]

- Proposed to explain how the population of EAs changes in steps

Study the change of $m(H,t)$ of SGA

the probability of not disrupting H by bit-wise mutation

$$E[m(H,t+1)] \geq m(H,t) \cdot \frac{\overline{f_H}}{\bar{f}} \cdot \left(1 - \left(p_c \cdot \frac{d(H)}{n-1}\right)\right) \cdot (1-p_m)^{o(H)}$$

the average fitness of individuals belonging to H in the population

the average fitness of individuals in the population

the probability of not disrupting H by one-point crossover

Schema theorem

Schema theorem [Holland, 1975]

- Proposed to explain how the population of EAs changes in steps

Study the change of $m(H,t)$ of SGA

$$E[m(H,t+1)] \geq m(H,t) \cdot \frac{\overline{f_H}}{\bar{f}} \cdot \left(1 - \left(p_c \cdot \frac{d(H)}{n-1}\right)\right) \cdot (1-p_m)^{o(H)}$$

⇩

Low-order and short schemata of above-average fitness will increase their instances from generation to generation

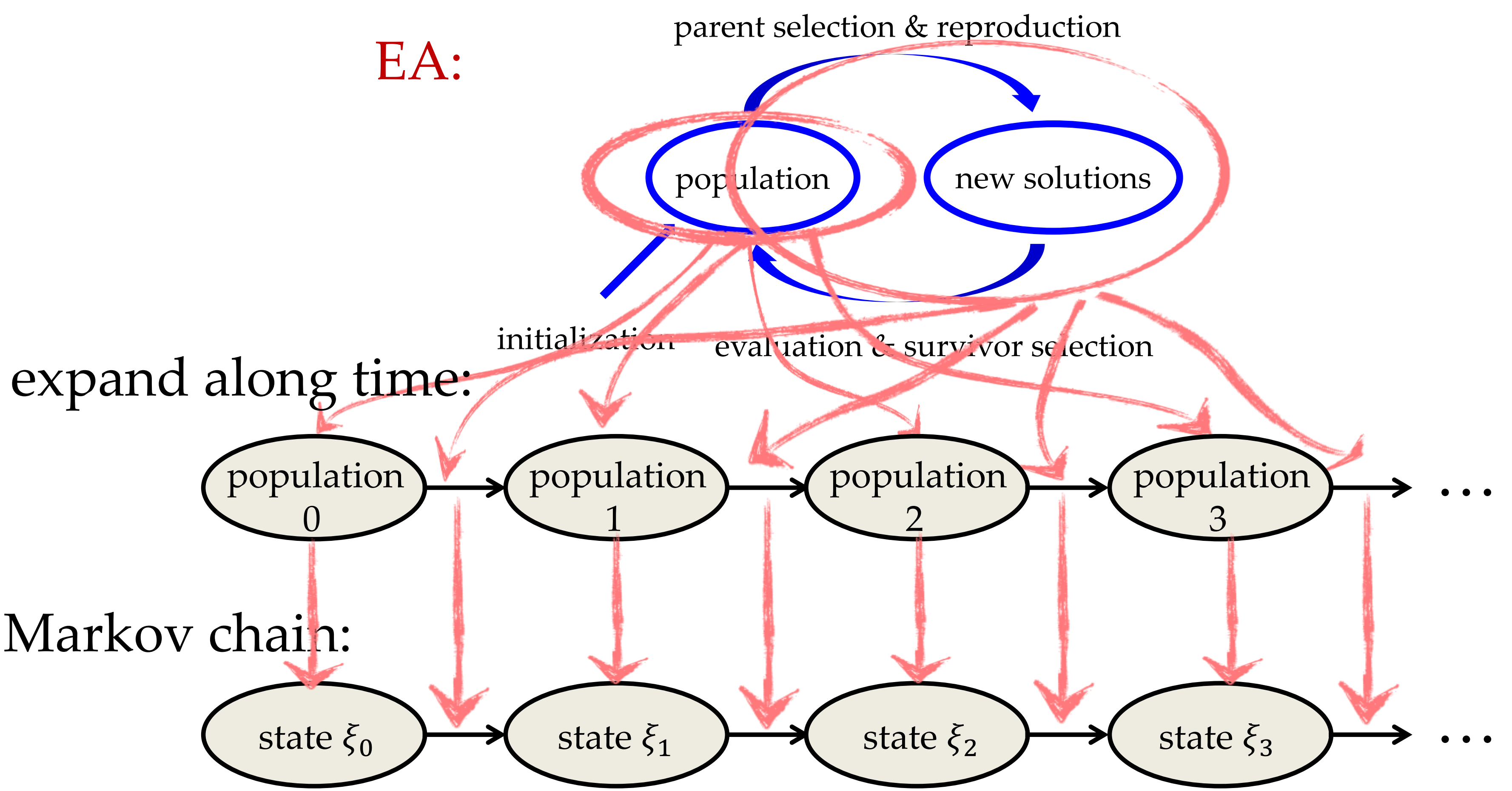
- Critiqued from several directions, and even wrong
- Cannot explain the global performance of EAs

Optimization-oriented theories

As an optimization algorithm, we concern:

- does an EA converge?
- how fast an EA converges?
- ...

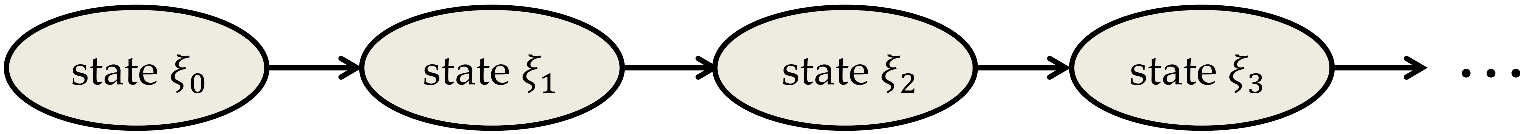
Markov chain modeling

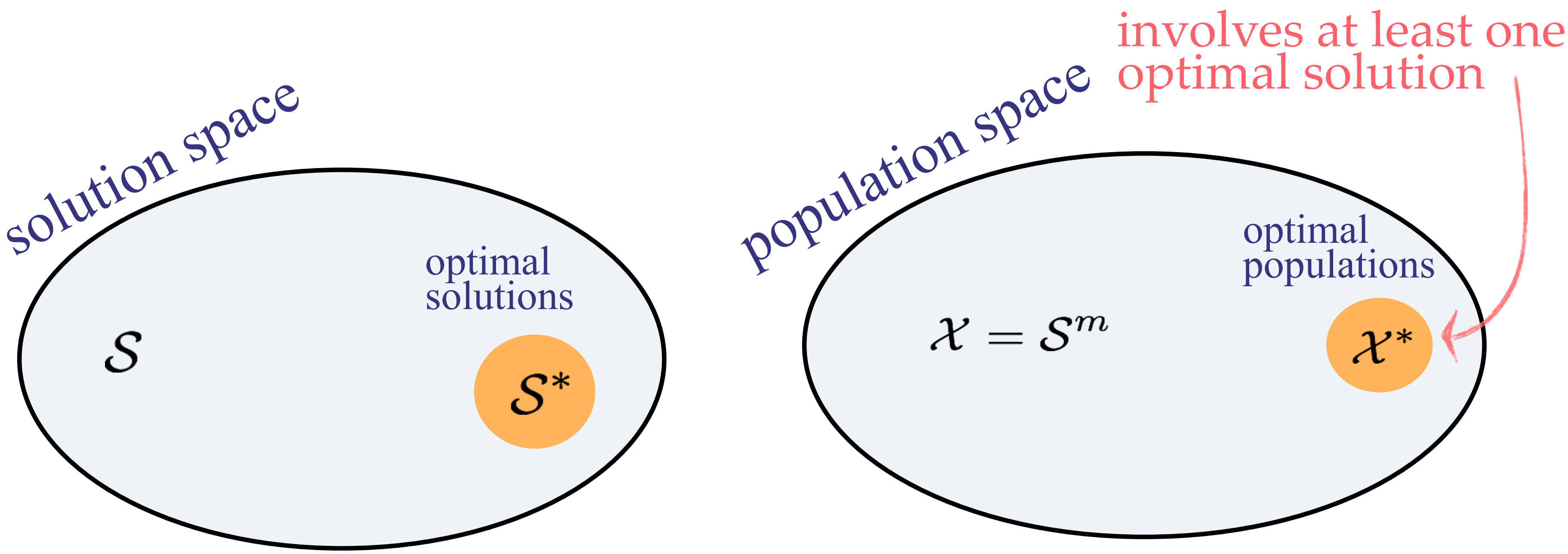


$$P(\xi_t|\xi_{t-1},\dots,\xi_1,\xi_0) = P(\xi_t|\xi_{t-1})$$

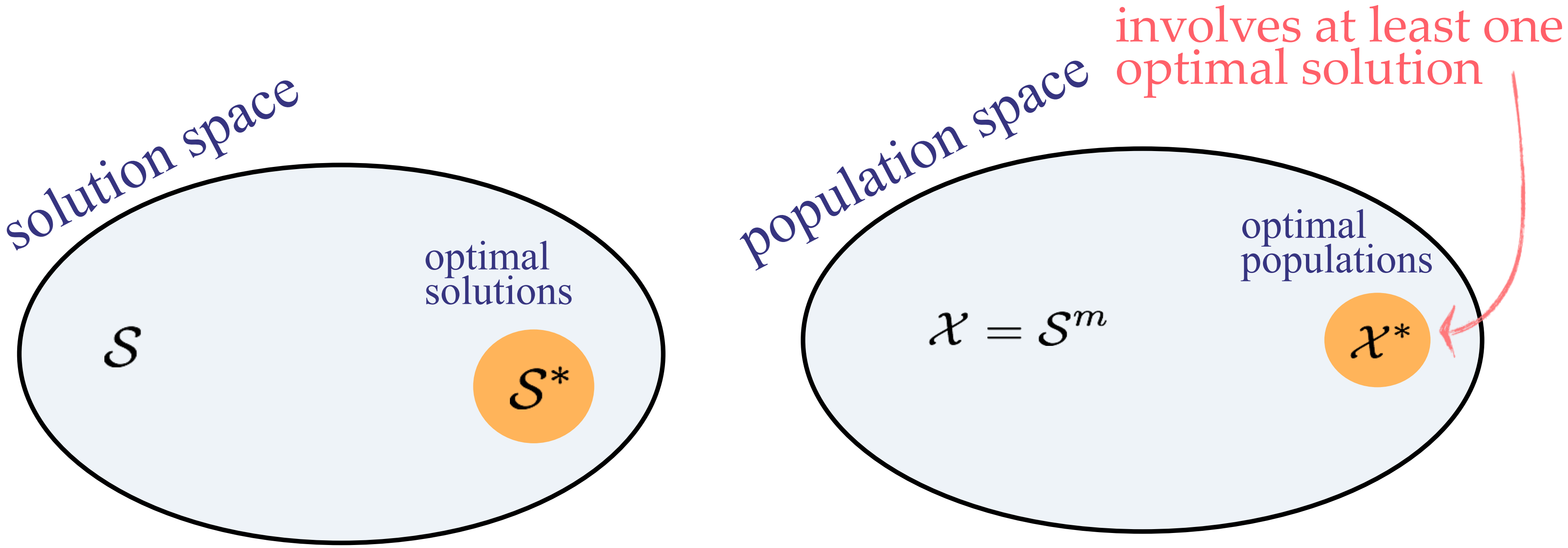
Markov chain modeling

Markov chain: $P(\xi_t|\xi_{t-1},\ldots,\xi_1,\xi_0) = P(\xi_t|\xi_{t-1})$





Size of population space



What is the size of population space?

$$\binom{|\mathcal{S}| + m - 1}{m}$$

Convergence

Does an EA converge to the optimal solutions?

$$\lim_{t \to +\infty} P(\xi_t \in \mathcal{X}^*) = 1$$

An EA that

1. uses global operators
2. preserves the best solution

converges to the optimal solutions

[Rudolph, 1998]

$$\Rightarrow \quad \forall x: P(\xi_{t+1} \in \mathcal{X}^* \mid \xi_t = x) > 0$$

$$\Downarrow$$

$$P(\exists t: \xi_t \in \mathcal{X}^*) = 1 - \prod_{t=0}^{+\infty} P(\xi_t \notin \mathcal{X}^*) = 1 \quad \Leftarrow \quad \prod_{t=0}^{+\infty} P(\xi_t \notin \mathcal{X}^*) = 0$$

But life is limited! How fast does it converge?

Running time complexity

Convergence analysis

$lim_{t \to +\infty} P(\xi_t \in \mathcal{X}^*) = 1$?

Running time analysis

$\tau = \min \{t \geq 0 \mid \xi_t \in \mathcal{X}^*\}$

The leading theoretical aspect

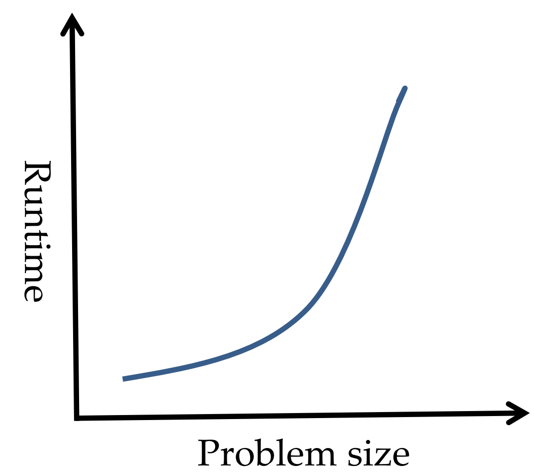
[Auger & Doerr, 2011; Neumann & Witt, 2012]

The number of iterations until finding an optimal or approximate solution for the first time

Running time complexity

- The number of iterations × the number of fitness evaluations in each iteration
- Usually grows with the problem size and expressed in asymptotic notations

  e.g., (1+1)-EA solving LeadingOnes: $O(n^2)$



Running time complexity

Convergence analysis

$lim_{t\to+\infty} P(\xi_t \in \mathcal{X}^*) = 1$?

Running time analysis

$\tau = \min \{t \geq 0 \mid \xi_t \in \mathcal{X}^*\}$

The leading theoretical aspect

[Auger & Doerr, 2011; Neumann & Witt, 2012]

The number of iterations until finding an optimal or approximate solution for the first time

A quick guide to asymptotic notations:

Let g and f be two functions defined on the real numbers.

- $g \in O(f)$: $\exists M > 0$ such that $g(x) \leq M \cdot f(x)$ for all sufficiently large x
- $g \in \Omega(f)$: $f \in O(g)$
- $g \in \Theta(f)$: $g \in O(f)$ and $g \in \Omega(f)$

$g \in O(f) \to g \leq f$

$g \in \Omega(f) \to g \geq f$

$g \in \Theta(f) \to g = f$

Running time complexity

Convergence analysis

$lim_{t\to+\infty} P(\xi_t \in \mathcal{X}^*) = 1$?

Running time analysis

$\tau = \min\{t \geq 0 \mid \xi_t \in \mathcal{X}^*\}$

The leading theoretical aspect

[Auger & Doerr, 2011; Neumann & Witt, 2012]

The number of iterations until finding an optimal or approximate solution for the first time

EAs are randomized algorithms

- They do not perform the same operations even if the input is the same
- They do not output the same result if run twice!

τ is a random variable.
We are interested in:

- $\mathrm{E}[\tau]$
- $P(\tau \leq T)$

Expectation

[Expectation] The expectation of a discrete random variable X is

$$E[X] = \sum_i i \cdot P(X = i)$$

where the sum is over all values in the range of X.

[Binomial Random Variable] A binomial random variable $X \sim B(n, p)$ with parameters n and p represents the number of successes in n independent experiments each of which succeeds with probability p.

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad E[X] = np$$

Expectation

[Expectation] The expectation of a discrete random variable X is

$$E[X] = \sum_i i \cdot P(X = i)$$

where the sum is over all values in the range of X.

[Geometric Random Variable] A geometric random variable X with parameter p represents the number of trials until the first success, where each trial succeeds with probability p.

$$P(X = i) = (1 - p)^{i-1}p \qquad\qquad E[X] = 1/p$$

Properties of expectation

[Law of Total Probability] For disjoint $B_1, B_2, \ldots, B_n$ that $\cup_{i=1}^{n} B_i = \Omega$,

$$P(A) = \sum_i P(A \wedge B_i) = \sum_i P(A \mid B_i)P(B_i)$$

[Law of Total Expectation] For disjoint $B_1, B_2, \ldots, B_n$ that $\cup_{i=1}^{n} B_i = \Omega$,

$$E[X] = \sum_i E[X \mid B_i]P(B_i)$$

[Linearity of Expectation] For any collection of discrete random variables $X_1, X_2, \ldots, X_n$ with finite expectations,

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$$

How to calculate the expectation

Two common ways of calculating $E[X]$:

- Let $X = X_1 + X_2 + \cdots + X_n$, then $E[X] = \sum_{i=1}^{n} E[X_i]$
- $E[X] = E[E[X \mid Y]]$

Example: *[Binomial Random Variable]* A binomial random variable $X \sim B(n, p)$ with parameters n and p represents the number of successes in n independent experiments each of which succeeds with probability p.

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad E[X] = np$$

Tail inequalities

[Markov's inequality] Let X be a random variable taking only non-negative values, and $E[X]$ its expectation. For any $t > 0$,

$$P(X \geq t) \leq E[X]/t$$

[Chernoff bounds] Let $X_1, X_2, \dots, X_n$ be independent Poisson trials, and $X = \sum_{i=1}^{n} X_i$. For any $\delta > 0$,

$$P(X \geq (1+\delta)E[X]) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{E[X]}$$

$$P(X \leq (1-\delta)E[X]) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{E[X]}$$

For a uniformly randomly sampled Boolean vector $\boldsymbol{x} \in \{0,1\}^n$, what is the probability of having no more than $2n/3$ 1-bits?

Tail inequalities

[Markov's inequality] Let X be a random variable taking only non-negative values, and $E[X]$ its expectation. For any $t > 0$,

$$P(X \geq t) \leq E[X]/t$$

⇓

[Chernoff bounds] Let $X_1, X_2, \ldots, X_n$ be independent Poisson trials, and $X = \sum_{i=1}^{n} X_i$. For any $\delta > 0$,

$$P(X \geq (1+\delta)E[X]) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{E[X]}$$

$$P(X \leq (1-\delta)E[X]) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{E[X]}$$

Union bound

[Union bound] For any finite or countably finite sequence of events $E_1, E_2, \ldots$, it holds that

$$P\left(\bigcup_{i\geq 1} E_i\right) \leq \sum_{i\geq 1} P(E_i)$$

Bit-wise mutation

For a Boolean vector $\boldsymbol{x} \in \{0,1\}^n$ with i 0-bits, after flipping each bit with prob. $1/n$ independently, what is the upper bound on the probability of decreasing the number of 0-bits by j?

E_i: j specific 0-bits of $\boldsymbol{x}$ are flipped

$$\leq P\left(\bigcup_{i\geq 1} E_i\right) \leq \binom{i}{j}\left(\frac{1}{n}\right)^j$$

$\left(\frac{1}{n}\right)^j \rightarrow P(E_i)$

Example of running time analysis

An extremely simplified EA
missing some features of *real* EAs

(1+1)-EA

$s \leftarrow$ a randomly drawn solution from $\mathcal{X}$
for t=1,2,... **do**
  $s' \leftarrow mutate(s)$
  if $f(s') \geq f(s)$ **then**
    $s \leftarrow s'$
  end if
  terminate if meets a stopping criterion
end for

no population

one-bit mutation
randomly choose one bit and change its value

bit-wise mutation
flip each bit with prob. $1/n$ independently

no crossover

for maximization, allow neutral changes

find an optimal solution

Example of running time analysis

Probing problem OneMax: $\arg\max_{\boldsymbol{x}\in\{0,1\}^n} \sum_{i=1}^{n} x_i$

count the number of 1 bits

fitness: $f(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$

EAs do not have the knowledge of the problems

only able to call $f(\boldsymbol{x})$

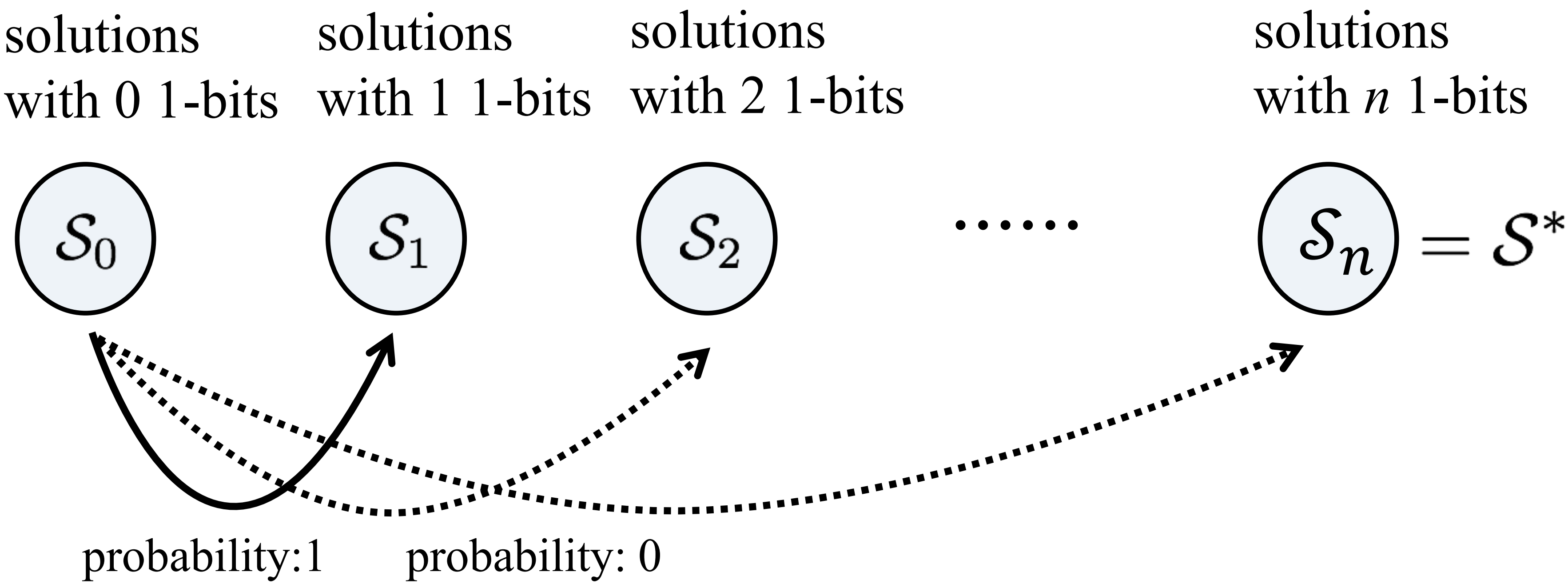
no difference with any other function $f : \{0,1\}^n \to \mathbb{R}$

Upper bound analysis

(1+1)-EA with one-bit mutation

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

the solutions with the same number of 1-bits share the same *f* value

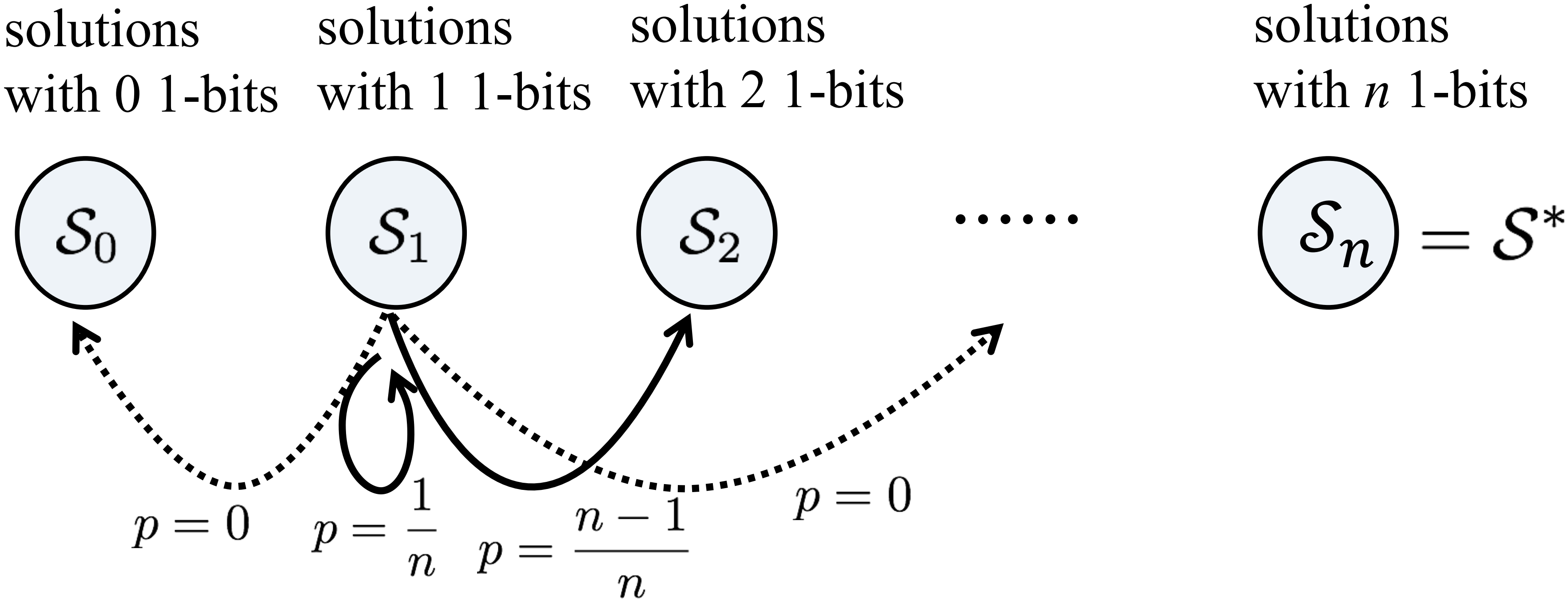


Upper bound analysis

(1+1)-EA with one-bit mutation

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

the solutions with the same number of 1-bits share the same *f* value



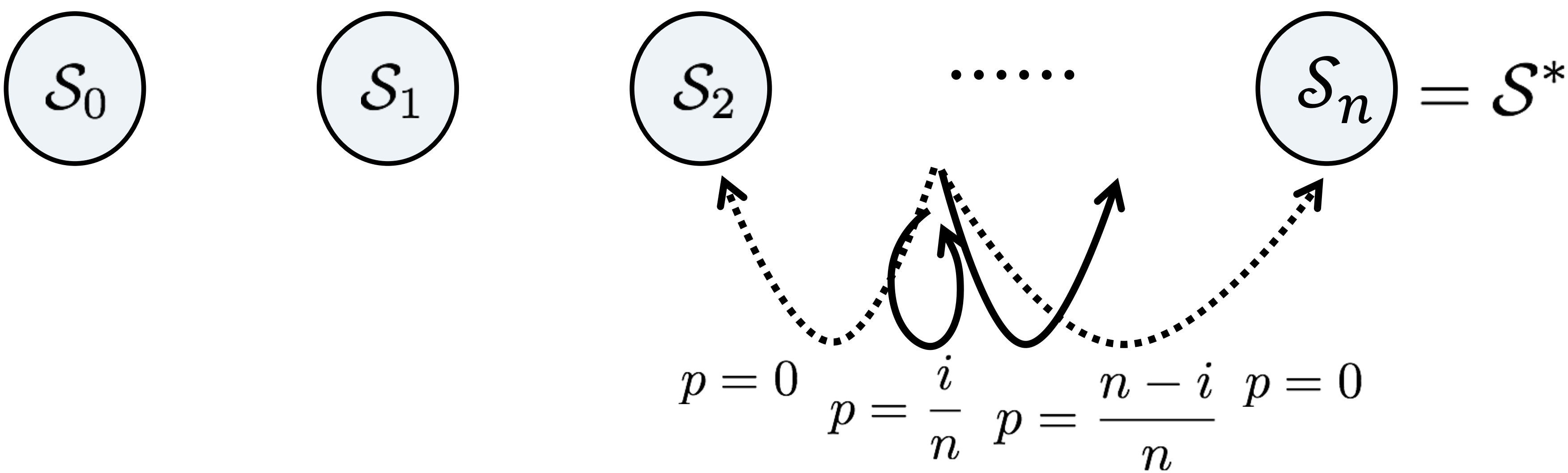
Upper bound analysis

(1+1)-EA with one-bit mutation

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

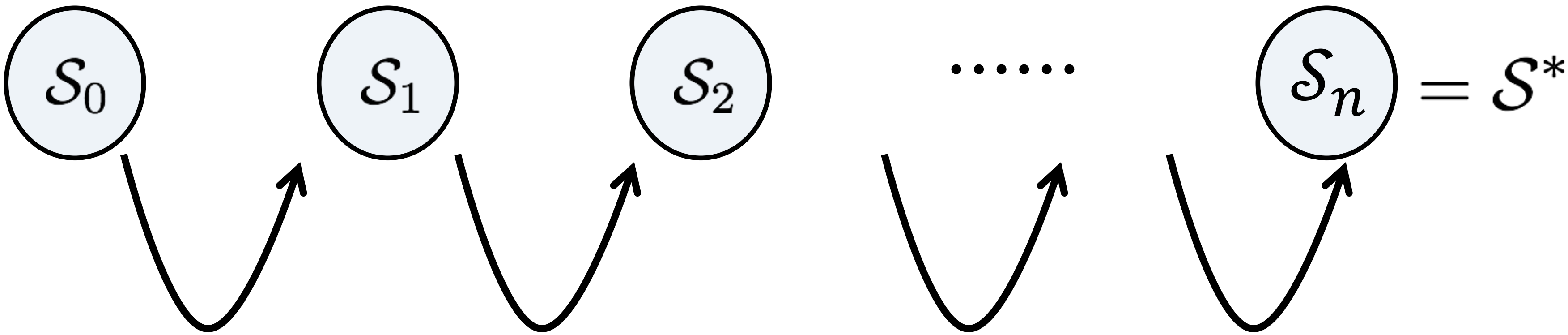
the solutions with the same number of 1-bits share the same *f* value

solutions with 0 1-bits | solutions with 1 1-bits | solutions with 2 1-bits | solutions with *n* 1-bits



Upper bound analysis

solutions with 0 1-bits	solutions with 1 1-bits	solutions with 2 1-bits		solutions with *n* 1-bits



probability of transition

$p = 1$ $\quad$ $p = \frac{n-1}{n}$ $\quad$ $p = \frac{n-i}{n}$ $\quad$ $p = \frac{1}{n}$

expected #iterations until the transition happens

1 $\quad$ $\frac{n}{n-1}$ $\quad$ $\frac{n}{n-i}$ $\quad$ $\frac{n}{1}$

Upper bound analysis

(1+1)-EA with one-bit mutation

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

expected #iterations until the transition happens

$$1 \quad \frac{n}{n-1} \quad \dots \quad \frac{n}{i} \quad \frac{n}{1}$$

summed up

$$\sum_{i=1}^{n} \frac{n}{i} = nH_n \sim n \ln n$$

expected running time upper bound $O(n \log n)$

Lower bound analysis

(1+1)-EA with bit-wise mutation

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

Let τ denote the running time, and $|\boldsymbol{x}|_0$ denote the number of 0-bits of the initial solution

Law of total expectation

$$
\begin{aligned}
E[\tau] &= \sum_{i=0}^{n} E[\tau \mid |\boldsymbol{x}|_0 = i] \cdot P(|\boldsymbol{x}|_0 = i) \\
&\geq \sum_{i=n/3}^{n} E[\tau \mid |\boldsymbol{x}|_0 = i] \cdot P(|\boldsymbol{x}|_0 = i) \\
&\geq E[\tau \mid |\boldsymbol{x}|_0 = n/3] \cdot P(|\boldsymbol{x}|_0 \geq n/3) \\
&\geq E[\tau \mid |\boldsymbol{x}|_0 = n/3] \cdot 1/4
\end{aligned}
$$

$P(|\boldsymbol{x}|_1 \leq 2n/3) \geq 1/4$ by Markov's inequality

Lower bound analysis

(1+1)-EA with bit-wise mutation

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

$$E[\tau] \geq E[\tau \mid |\boldsymbol{x}|_0 = n/3] \cdot 1/4$$

In $(n-1)\ln n$ iterations, at least one of these $n/3$ 0-bits is never flipped

⇩

The optimum is not found

⇩

$$\tau > (n-1)\ln n$$

the probability is lower bounded by

Lower bound analysis

$$E[\tau] \geq E[\tau \mid |\boldsymbol{x}|_0 = n/3] \cdot 1/4$$

$$\geq (n-1)\ln n \cdot P(\tau > (n-1)\ln n) \cdot 1/4$$

lower bound

In $(n-1)\ln n$ iterations, at least one of these $n/3$ 0-bits is never flipped

- $1 - 1/n$: a specific 0-bit is not flipped
- $(1 - 1/n)^t$: a specific 0-bit is never flipped in t iterations
- $1 - (1 - 1/n)^t$: a specific 0-bit is flipped at least once in t iterations
- $(1 - (1 - 1/n)^t)^{n/3}$: any of these $n/3$ 0-bits is flipped at least once in t iterations
- $1 - (1 - (1 - 1/n)^t)^{n/3}$

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

Lower bound analysis

(1+1)-EA with bit-wise mutation

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

$$E[\tau] \geq E[\tau \mid |\boldsymbol{x}|_0 = n/3] \cdot 1/4$$

$$\geq (n-1)\ln n \cdot P(\tau > (n-1)\ln n) \cdot 1/4$$

$$\geq (n-1)\ln n \cdot \left(1 - \left(1 - (1-1/n)^{(n-1)\ln n}\right)^{n/3}\right) \cdot 1/4$$

$(1-1/n)^{n-1} \geq 1/e$

$$\geq (n-1)\ln n \cdot \left(1 - \left(1 - e^{-\ln n}\right)^{n/3}\right) \cdot 1/4$$

$$= (n-1)\ln n \cdot \left(1 - (1-1/n)^{n/3}\right) \cdot 1/4$$

$(1-1/n)^{n} \leq 1/e$

$$\geq (n-1)\ln n \cdot \left(1 - e^{-1/3}\right) \cdot 1/4 \in \Omega(n \log n)$$

Example of running time analysis

For (1+1)-EA solving OneMax $f(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$

If using one-bit mutation,

expected running time upper bound $O(n \log n)$

If using bit-wise mutation,

expected running time lower bound $\Omega(n \log n)$

Not asymptotically faster

Running time analysis tools

When facing new situations, analyses starting from scratch are quite difficult

We need general running time analysis tools to guide the analysis

- Fitness level
- Drift analysis
- Switch analysis

Summary

- Schema theorem
- Markov chain modeling
- Convergence
- Running time complexity
- Expectation and tail inequalities
- Example of running time analysis

References

- A. E. Eiben and J. E. Smith. Introduction to Evolutionary Computing. Chapter 16.
- K. A. De Jong. Evolutionary Computation – A Unified Approach. Chapter 6.
- G. Rudolph. Finite Markov chain results in evolutionary computation: A tour d'horizon. Fundamenta Informaticae, 1998, 35(1-4): 67-89.