[*The exam text is in English. You should answer the questions in English.*

The exam consists of 23 questions (7 pages). For the open questions: Mark each answer with the question number *and use a* ***new page*** *for each of the answers.* The score is indicated per question.

The exam starts at 09.30 hrs and ends at 12.00 hrs. Participation in the exam requires being present for at least 1 hr.]

Depending on occasion:
Electronic devices are not allowed and must be stowed.
Points for each open question are indicated.

- *Please write readable and write your name and student identification number on every single sheet of paper for unique identification!*
- ***Don't spend much time writing long stories – that's NOT necessary. Don't panic! You CAN make it!***
- *Look at the whole exam first to get an overview and do those assignments first which you feel are doable for you.*
- *Good success and all the best for your studies!*

## Multiple choice – 20 questions with 60 points.

- ***Each problem has a variable number of correct statements***
- ***There is always at least one correct statement for each question***
- ***Write the letters of the correct statements for each question in the table (on the last page), e.g., "A, B" if statements A and B are correct for a question.***
- ***All correct statements are marked: 3 points***
- ***Correct statements(s) is/are partially marked without a wrong answer: 1.5 points***
- ***A wrong statement is marked: 0 points***

1. Which statement(s) are characteristics of evolutionary algorithms? (AB)
   A. Evolutionary algorithms require no fitness gradient information of any kind to proceed.
   B. Evolutionary algorithms are easy to process in parallel and can escape from local minima where deterministic optimization methods may fail.
   C. Evolutionary algorithms are local random search algorithms.
   D. Evolutionary algorithms require presumptions with respect to problem space.

2. Regarding exploration and exploitation in EAs, which statement(s) is/are correct? (BCD)
   A. Exploration relates to local search, and exploitation relates to a global search.
   B. Crossover operators have the ability to jump out of a local optimum.
   C. Mutation operators can search for an optimum near the parent.
   D. Go too far into exploration will get stuck in local optima.

3. Two state-of-the-art discrete problems are OneMax ( $f_{OM}(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$ ) and BinaryValue ($f_{BV}(\boldsymbol{x}) = \sum_{i=1}^{n} 2^{n-i} x_i$). Which statement(s) is/are correct regarding solving these two problems by EAs? (D)
   A. Compared to $f_{BV}(\boldsymbol{x})$ , $f_{OM}(\boldsymbol{x})$ is easier to optimize, because there is a strong fitness-distance correlation in $f_{OM}(\boldsymbol{x})$.

B. Compared to $f_{BV}(\boldsymbol{x})$ , $f_{OM}(\boldsymbol{x})$ is easier to optimize, because $f_{OM}(\boldsymbol{x})$ is a monotonically strictly increasing function.
C. Compared to $f_{OM}(\boldsymbol{x})$ , $f_{BV}(\boldsymbol{x})$ is easier to optimize, because each variable in a decision vector has a different weight in BinaryValue function
D. $f_{OM}(\boldsymbol{x})$ and $f_{BV}(\boldsymbol{x})$ have the same difficulty to optimize, as only the Hamming distance to the optimum determines the quality of search.

4. Suppose the computational complexity of an optimization algorithm is $\Theta(n^2 logn)$, where $n$ represents the number of elements in a decision vector. Which interpretation(s) is/are correct? Suppose $c$ and $C$ are two constants, and $0 < c < C \ll \infty$ in the following statements.
(AD)
A. There exists a constant $C > 0$ such that for all $n$ the expected runtime is at most $Cn^2 logn$
B. There exists a constant $C > 0$ such that for all $n$ the expected runtime is at least $Cn^2 logn$
C. There exists a constant $c > 0$ such that for all $n$ the expected runtime is at most $cn^2 logn$
D. There exists a constant $c > 0$ such that for all $n$ the expected runtime is at least $cn^2 logn$

5. We denote $T(n) = 2n \times n + 3C$ and $T'(n) = n + 1000$ as the running time of the algorithms *A* and *B, respectively,* where $n$ is the dimension of the given problem and $C$ is a constant. Which statement(s) is/are correct?
(BC)
A. The algorithm *A* is linear time $O(n)$
B. The algorithm *A* is quadratic time $O(n^2)$
C. The algorithm *B* is linear time $O(n)$
D. *A* is faster than *B* for a large $n$.

6. Given a bitstring (i.e., individual) $\vec{a} = (01011100)$ and a mutation operator $m_p: \{0,1\}^8 \rightarrow \{0,1\}^8$ that inverts single bits independently of each other with probability $p \in [0, 1]$. Which statement(s) is/are correct? (D)
A. Total probability that 3 out of 8 bits are changed by mutation is: $15p^3(1-p)^5$
B. $P(\vec{a} \rightarrow\ 10101011)$ is $p^7$
C. $P(\vec{a} \rightarrow \vec{a})$ is 1
D. Total probability that $i$ $(0 \leq i \leq 8, i \in \mathbb{N})$ out of 8 bits are changed by mutation is: $\binom{8}{i}p^i(1-p)^{8-i}$

7. Given a parent population size $\mu$ and an offspring population size $\lambda$. Suppose one and only one individual is initialized as an optimum at the 0th generation, what is the number of generations needed for the optimum to fill the whole population? (Mutation and crossover are not considered) (AC)
A. $(\mu, \lambda)$-selection needs $\frac{\ln \lambda}{\ln \lambda/\mu}$ generations
B. $(\mu, \lambda)$-selection needs $\frac{\ln \mu}{\ln \mu/\lambda}$ generations
C. Proportional selection needs $\lambda \ln \lambda$ generations
D. Proportional selection needs $\frac{\ln \mu}{\ln \mu/\lambda}$ generations

8. To analyze a (1+1) genetic algorithm (GA) on a $n$-dimensional OneMax ( $f_{OM}(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$ ), we denote a $k$-step success probability $p_{\boldsymbol{x}}^{+}(k) = P\{f_{OM}(m(\boldsymbol{x})) = f_{OM}(\boldsymbol{x}) + k\}$, $m(\boldsymbol{x})$ is the solution candidate obtained from applying mutation on $\boldsymbol{x}$. Which formula can present the convergence velocity of the (1+1) GA? (B)
   A. $\varphi = \sum_{k=0}^{f_{OM}(x)} p_{\boldsymbol{x}}^{+}(k)$
   B. $\varphi = \sum_{k=0}^{n-f_{OM}(x)} k \cdot p_{\boldsymbol{x}}^{+}(k)$
   C. $\varphi = \sum_{k=0}^{n-f_{OM}(x)} p_{\boldsymbol{x}}^{+}(k)$
   D. $\varphi = \sum_{k=0}^{f_{OM}(x)} k \cdot p_{\boldsymbol{x}}^{+}(k)$

9. Which statement(s) is/are correct based on your knowledge of modern GA theory? (ABD)
   A. (1+1) evolutionary algorithm does not have the same optimization time complexity on all liner problems.
   B. Theory contributes to the understanding and use of EAs.
   C. Functions without local optima are always easier than functions with local optima to optimize.
   D. The (1+1) evolutionary algorithm finds the maximum of OneMax ( $f_{OM}(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$ ) in an expected number of $O(n\log n)$ iterations.

10. Which statement(s) is/are true about discrete and intermediary recombination? (ACD)
   A. Both of them are conserving common components of parent individuals.
   B. Discrete recombination only works for discrete search spaces, while intermediary recombination works well on continuous problems.
   C. The offspring generated by global intermediary recombination are the same.
   D. Recombination produces new individuals by combining the information contained in two or more parents.

11. Which statement(s) is/are correct for the learning rate $\tau$ of the self-adaptive ES with one step-size $\sigma$? (AD)
   A. If $\tau$ is bigger, $\sigma$ will adapt faster but less precise.
   B. If $\tau$ is bigger, $\sigma$ will adapt slower but more precise.
   C. If $\tau$ is smaller, $\sigma$ will adapt faster but less precise.
   D. If $\tau$ is smaller, $\sigma$ will adapt slower but more precise.

12. Regarding 1/5 rule in (1+1)-ES, which following statement(s) is/are correct? (BD)
   A. The number 5 is inspired by nature: every primate has 5 fingers in one hand.
   B. The step size should be increased/decreased if the success probability is bigger/smaller than 1/5.
   C. The step size should be increased/decreased if a random number, which is a realization of a standard normal distribution, is bigger/smaller than 1/5.
   D. The number 0.2 roughly equals the optimal success probability in corridor and sphere models.

13. Which statement(s) is/are correct about self-adapted mutations (single step-size mutation, individual mutation, correlated mutation) in ES?
(ABD)
    A. Correlation matrices used in correlated mutation must be positive definite.
    B. For a decision vector with $n$ variables, single step-size and individual mutations have 1 and $n$ parameters, respectively.
    C. For a population with $\mu$ individuals, single step-size and individual mutations have 1 and $\mu$ parameters, respectively.
    D. An individual mutation operator can work efficiently on problems of differently scaled object variables, while a single step-size mutation cannot.

14. Why is a log-normal distribution used to update step size in ES? (AC)
    A. A log-normal distribution can guarantee that the updated step size is a positive number.
    B. A log-normal distribution can guarantee that a step size has 50% probability to stay the same.
    C. A log-normal distribution can guarantee the same probability to increase and decrease a step size.
    D. In log-normal distribution, large changes are more likely to happen than small ones.

15. Compared with "+" selection in ES, what statement(s) is/are true for "," selection? (ABC)
    A. A mis-adapted strategy parameter will be eliminated.
    B. A deterioration may occur.
    C. The offspring population size $\mu > 1$ is necessary.
    D. A recommended selective pressure is 5 because of the 1/5 rule.

16. Suppose the formula $\int_B^A zPDF(z)dz$ represents a convergency velocity of ES, what is/are the correct lower and upper integral domain(s)? (Hint: $z$ represents progress) (BC)
    A. For (1+1)-ES: $A = \infty, B = -\infty$
    B. For (1,1)-ES: $A = \infty, B = -\infty$
    C. For (1+1)-ES: $A = \infty, B = 0$
    D. For (1,1)-ES: $A = \infty, B = 0$

17. What are the differences between GA and ES? (AC)
    A. GA emphasizes the crossover operator, while ES emphasizes on mutation.
    B. GA and ES work efficiently with small population size and large population size, respectively.
    C. GA uses discrete representations, and ES is capable of dealing with mixed-integer variables.
    D. Crossover/recombination is a must in ES but is optional in GA.

18. Between $(\mu + \lambda)$-ES and $(\mu, \lambda)$-ES selections, which statement(s) is/are reasonable regarding a choice of $\mu$ and $\lambda$? (CD)
    A. In $(\mu, \lambda)$-ES selection, $\lambda < \mu$
    B. In $(\mu, \lambda)$-ES selection, $\frac{\mu}{\lambda} = 1$
    C. In $(\mu + \lambda)$-ES selection, $\lambda < \mu$
    D. In $(\mu + \lambda)$-ES selection, $\frac{\mu}{\lambda} = 1$

19. Why is a normal distribution used in ES for continuous problems? (ACD)

A. Because it can maximize the unbiasedness.
B. It is not necessary to use normal distribution, a uniform distribution with domain of $(-\infty, \infty)$ also works fine in ES for continuous problems.
C. Because of its infinite support.
D. Because the total probabilities of increasing and decreasing a decision variable are the same.

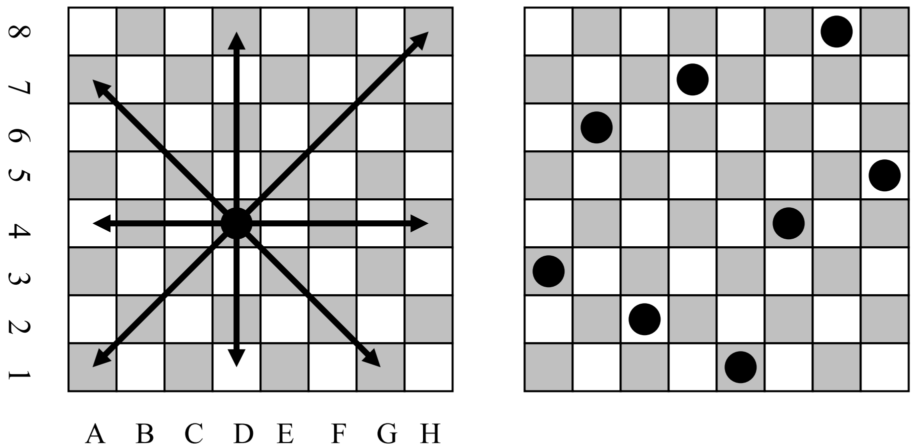
20. Which of the following step-size control mechanism(s) in ES is/are capable of self-adaption? ($\mathcal{N}(0,1)$ is a realization of a standard normal distribution, $\mathcal{U}(0,1)$ is a realization of a uniform distribution, $\alpha, \tau, \tau', a$ are learning rates)

(BC)

A. 1/5 step-size control
B. $\sigma' = \sigma(1 + \alpha\mathcal{N}(0,1))$
C. $\sigma_i' = \sigma_i \exp(\tau' + \tau\mathcal{N}(0,1))$
D. $\sigma' = \begin{cases} \sigma/a & if\ \ \mathcal{U}(0,1) > 0.5 \\ \sigma \cdot a & if\ \ \mathcal{U}(0,1) \leq 0.5 \end{cases}$

## Open Questions (to be answered on separate sheet of paper):

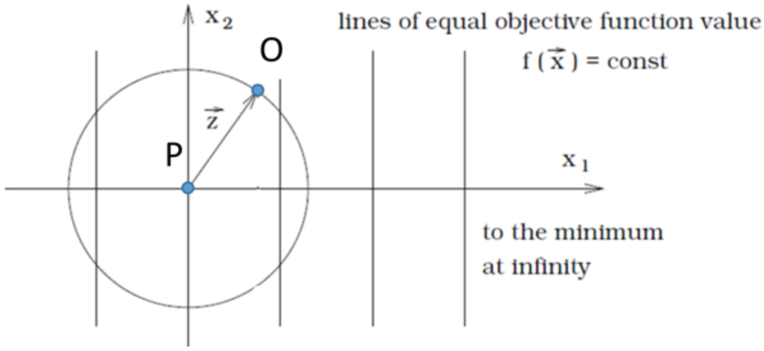
1. **(Total: 15)** The so-called 8-queens problem is defined as the problem to place 8 queens on an 8 by 8 chessboard in such a way that the queens cannot check each other (the figure below, left, shows all possible fields a queen can move to from position d4, while the figure to the right shows a feasible solution). Design an evolutionary algorithm for the generalized n-queens problem by answering the following questions, and give reasons, if possible, for your design choices.




   a. **(2)** Develop a representation of solution candidates for the n-queens problem (on an n-by-n board) to allow for an evolutionary approach towards solving the problem.
   b. **(2)** Develop a mutation operator working on the solution candidates.
   c. **(2)** Develop a crossover operator.
   d. **(4)** Propose an objective function for the n-queens problem.
   e. **(2)** Propose a selection operator for the n-queens problem.
   f. **(3)** Explain how your approach handles infeasible solution candidates.

2. **(Total: 15)** Answer the following questions about the Evolution Strategy (ES).
   a. **(6)** Describe the three different mutation operators of ES and give the corresponding representation of search points.
   b. **(3)** Suppose that we run an ES with $(\mu, \lambda)$-selection without crossover and mutation (i.e., there is no genetic variation, only selection and reproduction). Let $x^*$ be the best solution in generation $t = 0$. At time $t = 0$ there is one instance of $x^*$in the population (denoted by $N_0 = 1$). What is the expected number of copies of $x^*$ in the generation $t = 1$ (i.e., compute $N_1$)?
   c. **(6)** The takeover time $\tau^*$of a selection strategy is defined as the expected number of generations needed for one individual to take over the entire population when there is no genetic variation, that is, it denotes the expected generation in which all individuals are copies of $x^*$. Show that the takeover time of the $(\mu, \lambda)$-selection is given by $\tau^*_{(\mu,\lambda)} = \frac{\ln\lambda}{\ln\frac{\lambda}{\mu}}$.

3. **(Total: 10)** In the picture below the progress rate analysis for the linear model $f(\boldsymbol{x}) = x_1 \rightarrow min, \boldsymbol{x} \in \mathbb{R}^2$ is sketched. Suppose $P = (p_1, p_2)$ is the parent and $O = (o_1, o_2)$ is the offspring.




a. **(2)** What is the progress of the offspring?

b. **(3)** What is the expression for the progress rate $\varphi(\boldsymbol{x},\sigma) = E(f_{t+1}(\boldsymbol{x}) - f_t(\boldsymbol{x}))$ of the (1+1)-ES on the linear model with a **constant** step size $\sigma$?
(**You don't have to derive the simplest expression in your result**, you can also use the probability density function of a standard normal distribution PDF($x$)=$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and an cumulative density function of a standard normal distribution CDF($x$)=$\frac{1}{2}(1+erf(\frac{x}{\sqrt{2}}))$ in your result.)

c. **(5)** For a (1+1)-evolution strategy, the so-called 1/5 success rule describes how $\sigma$ is modified over the course of the evolutionary optimization process. What was the reasoning of Rechenberg and Schwefel to choose the 'magic' number of 1/5th? (Hint: Explain the theoretical argument for the 1/5 success rule. Clarify the relationship between convergence velocity and success probability in words and by a figure.)

**Please write the alphabetic characters (e.g., CD) representing your answers of the multiple-choice questions into the table below.**

| No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Answers | | | | | | | | | | |
| No. | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Answers | | | | | | | | | | |