Chalmers University of Technology, Department of Applied Mechanics
Teacher: Mattias Wahde, tel. 772 3727

Exam in FFR 105 (Evolutionary computation), 2007-10-25, 14.00-18.00, V.
It is allowed to use a calculator, as long as it cannot store any text. Furthermore, mathematical tables (such as Beta, Standard Math etc.) are allowed, provided that no notes have been added. However, it is *not* allowed to use the handouts (including the problems) from the course during the exam.

Note! In problems involving computation, show *clearly* how you arrived at your answer, i.e. include intermediate steps etc. Only giving the answer will result in zero points on the problem in question.

There are 5 problems in the exam, and the maximum number of points is 25.

1. (a) Many different selection operators have been defined in connection with evolutionary algorithms. Describe, in detail, the following operators (it is *not* necessary to write Matlab code, but your description should be sufficiently detailed to make it possible to write such code based on your text!) (3p):

   Roulette-wheel selection
   Tournament selection
   Boltzmann selection

   (b) Particle swarm optimization (PSO) is based on the properties of animal swarms, such as e.g. bird flocks. One of the most central parts of PSO is the method for updating particle velocities. Write down and describe the equation for updating particle velocities in standard PSO! (2p)

   (c) A common problem in applications of evolutionary algorithms is *premature convergence*. Name (and describe briefly) at least two different methods for preventing premature convergence. (1p)

   (d) In optimization problems, convex objective functions constitute a very important special case. How are convex functions defined (mathematically)? Give also a geometric interpretation (in one dimension) of the definition. (1p)

   (e) Is the function

   $$f(x_1, x_2) = 4x_1^2 + 2x_2^2 - 2x_1 x_2 \qquad (1)$$

   convex or not? Motivate your answer clearly! (1p)

2. Many different analytical and numerical methods have been developed in the field of classical optimization.

   (a) Using the method of Lagrange multipliers, find the minimum value of the function

   $$f(x_1, x_2) = x_1^2 x_2 \qquad (2)$$

on $\mathbf{R}^2$, subject to the equality constraint

$$2x_1^6 + 3x_1^4 x_2^2 + 8x_2^6 - 36 = 0. \tag{3}$$

(3p)

(b) Consider the function

$$g(x_1, x_2) = x_1^4 + x_1 x_2 + x_2^2. \tag{4}$$

Find (and report) the gradient of $g(x_1, x_2)$ at the point $(1, 1)^{\mathrm{T}}$. Next, starting from this point, take one step of gradient descent (including the line search needed to find the minimum along the search direction). Which point (x_1^*, x_2^*) is reached after this step? Finally, show that the search direction obtained in the next gradient descent step is perpendicular to the first search direction determined above (a numerical demonstration is sufficient, i.e. it is *not* required to provide an analytical proof). (3p)

3. Ant colony optimization is an important stochastic optimization method, of which several different versions have been defined.

   (a) One of the first ant-inspired algorithms was *ant system* (AS). Describe the AS algorithm in detail, i.e. in the form of a list of steps, with a clear description of each step. You may use the traveling salesman problem (TSP) as a concrete example. (2p)

   (b) Min-Max Ant System (MMAS) differs from AS in the sense that only the best ant is allowed to deposit pheromone. Furthermore, in MMAS, pheromone levels are explicitly constrained to a given range $[\tau_{\min}, \tau_{\max}]$. Let $p(K)$ be the probability that the optimal solution (for the case of TSP) is encountered (using MMAS) at least once in the first K generations. Prove that

$$\lim_{K \to \infty} p(K) = 1. \tag{5}$$

   (2p)

4. Consider a case where a genetic algorithm with a binary encoding scheme is to be used in a problem where the fitness function for a chromosome with k ones is given by

$$f(k) = k\frac{2 + (-1)^k}{3}. \tag{6}$$

The population size is assumed to be infinite, and it is further assumed that the chromosomes are initialized randomly so that the probability distribution for the initial population will be

$$p_1(k) = 2^{-n}\binom{n}{k}, \tag{7}$$

where n is the length of the chromosomes.

   (a) Compute the average fitness for the initial population. (2p)

   (b) Find the probability distribution $p_2(k)$ after evaluation and (roulette-wheel) selection, assuming that no mutations take place. (1p)

5. The *Schema theorem* indicates how so-called building blocks spread in the population of a genetic algorithm.

   (a) Derive the Schema theorem. Make sure to include all relevant definitions and intermediate steps in the derivation. In particular, motivate *clearly* the expression showing how the expected number of copies of a schema S is expected to vary from one generation to the next. (2p)

   (b) Consider a population of seven individuals with the six–digit binary chromosomes

$$
\begin{aligned}
c_1 &= \; 010111 \\
c_2 &= \; 001101 \\
c_3 &= \; 100010 \\
c_4 &= \; 101000 \\
c_5 &= \; 000111 \\
c_6 &= \; 110110 \\
c_7 &= \; 110000
\end{aligned}
$$

   In this case, the chromosomes are decoded by multiplying gene i by 2^{6-i} so that $c_1 \to x_1 = 0 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 23$ etc. The chromosomes are used in a GA in order to find the maximum of the simple function $f(x) = \frac{1}{(1+x)}$ for $x \geq 0$. Consider the schema S_1 =10xxxx. Assuming $p_c = 0.75$ and $p_{\text{mut}} = 0.01$, what is the expected number of copies of S_1 in the next generation, according to the schema theorem? (2p)

Evolutionary computation (FFR 105), 2007
Solutions to the exam (2007-10-25)

1. (a) See Chapter 3 of the book. It is important to list the equations showing how the selection is carried out in the different cases.

   (b) See Chapter 5. The velocity update is made according to

$$v_{i,j} \leftarrow w v_{i,j} + c_1 q \left(\frac{x_{i,j}^{\text{pb}} - x_{i,j}}{\Delta t} \right) + c_2 r \left(\frac{x_j^{\text{sb}} - x_{i,j}}{\Delta t} \right), \; i = 1, \ldots, N, \; j = 1, \ldots, n \tag{1}$$

   where w is the inertia weight that regulates the trade-off between exploitation and exploration. The term proportional to c_1 is the cognitive component, which measures the degree to which a particle trusts its own previous performance as a guide towards obtaining better results. The term proportional to c_2 is the social component of the velocity update, and it measures the degree to which a particle trusts the ability of the other members of the swarm to find good solutions. Commonly, velocities are also limited to a specific range $[-v_{\max}, v_{\max}]$.

   (c) See Chapter 3. Examples include the use of sub-populations or a varying mutation rate.

   (d) Convex functions fulfill the inequality

$$f\left(a\mathbf{x}_1 + (1-a)\mathbf{x}_2\right) \leq a f(\mathbf{x}_1) + (1-a) f(\mathbf{x}_2), \tag{2}$$

   for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{S}$ (where $\mathbf{S}$ is a convex set) and for all $a \in [0,1]$. The geometric interpretation is that the function values lie "below" the straight line joining $\mathbf{x}_1$ and $\mathbf{x}_2$.

   (e) The convexity of a function can be investigated by considering the properties of the Hessian. For the function in question, the Hessian equals

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 8 & -2 \\ -2 & 4 \end{pmatrix}, \tag{3}$$

   with eigenvalues $6 \pm 2\sqrt{2}$ which are both larger than zero. Thus, the function is convex.

2. (a) The function $L(x_1, x_2, \lambda)$ takes the form

$$L = x_1^2 x_2 + \lambda(2x_1^6 + 3x_1^4 x_2^2 + 8x_2^6 - 36). \tag{4}$$

   Setting the derivatives of L to zero, one obtains the equations

$$\frac{\partial L}{\partial x_1} = 2x_1 x_2 + \lambda(12x_1^5 + 12x_1^3 x_2^2) = 0 \tag{5}$$

$$\frac{\partial L}{\partial x_2} = x_1^2 + \lambda(6x_1^4 x_2 + 48x_2^5) = 0 \tag{6}$$

$$\frac{\partial L}{\partial \lambda} = 2x_1^6 + 3x_1^4 x_2^2 + 8x_2^6 - 36 = 0. \tag{7}$$

Now, multipliying the first equation by x_1 and the second one by $2x_2$, and subtracting the results, the equation

$$12x_1^6 - 96x_2^6 = 0. \tag{8}$$

is obtained. Assuming that x_1 and x_2 are different from zero (the exceptions will be handled later), the solution to this equation is

$$x_1 = \pm\sqrt{2}x_2. \tag{9}$$

Inserting this expression into the constraint equation one obtains

$$36x_2^6 - 36 = 0, \tag{10}$$

so that $x_2 = \pm 1$. Thus, the four points that must be considered are $(\sqrt{2}, 1)$, $(-\sqrt{2}, 1)$, $(\sqrt{2}, -1), (-\sqrt{2}, -1)$. Examining these four points, it easy to see that the minimum value is -2. The cases $x_1 = 0$ and $x_2 = 0$ can be discarded since they would give a value of $0 > -2$.

(b) The gradient is given by

$$\nabla f = \left(4x_1^3 + x_2, x_1 + 2x_2\right)^{\mathrm{T}}. \tag{11}$$

Thus, at $(1,1)^{\mathrm{T}}$, the gradient equals $(5,3)^{\mathrm{T}}$. The new point reached via gradient descent from $(1,1)^{\mathrm{T}}$ thus takes the form

$$\mathbf{x}^{\mathrm{new}} = \mathbf{x}^{\mathrm{old}} - \eta\nabla f = (1 - 5\eta, 1 - 3\eta)^{\mathrm{T}}. \tag{12}$$

Inserting this expression in the function $g(x_1, x_2)$, the function

$$\phi(\eta) = (1 - 5\eta)^4 + (1 - 5\eta)(1 - 3\eta) + (1 - 3\eta)^2 \tag{13}$$

is obtained. Using the bisection method starting from, say, the interval $[0,1]$, the minimum is found to occur at $\eta^* \approx 0.272125$. Thus, the point reached will be

$$(1 - 5\eta^*, 1 - 3\eta^*)^{\mathrm{T}} \approx (-0.3606, 0.1836)^{\mathrm{T}}. \tag{14}$$

At this point, the gradient equals $(-0.00397, 0.006625)^{\mathrm{T}}$. Finally, it is easy to verify that the scalar product of this vector and the original gradient $(5,3)^{\mathrm{T}}$ is equal to zero, showing that the two vectors are ortogonal, i.e. that the search directions are perpendicular to each other.

3. (a) Description of AS, see Chapter 4.
   (b) Proof: See the Appendix B.3.2

4. From the equation (the binomial distribution)

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \tag{15}$$

we get (with $a = -1$ and $b = x$)

$$(x-1)^n = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} x^k = (-1)^n \sum_{k=0}^{n} \binom{n}{k} (-1)^k x^k. \tag{16}$$

($(-1)^{-k}$ can be exchanged for $(-1)^k$). Inserting $x = 1$ we find

$$0 = \sum_{k=0}^{n} \binom{n}{k} (-1)^k \tag{17}$$

Differentiation of Eq. (16) w.r.t. x gives

$$n(x-1)^{n-1} \quad = \quad (-1)^n \sum_{k=0}^{n} k(-1)^k x^{k-1}. \tag{18}$$

Insertion of $x = 1$ gives

$$0 = n(1-1)^{n-1} = \sum_{k=0}^{n} k(-1)^k \binom{n}{k}. \tag{19}$$

Thus, the average fitness value for the initial population becomes

$$\begin{aligned} \overline{f}_1 &= \sum_{k=0}^{n} f(k)p_1(k) = \sum_{k=0}^{n} k \frac{2+(-1)^k}{3} 2^{-n} \binom{n}{k} = \\ &= \frac{2}{3} 2^{-n} \sum_{k=0}^{n} k 2^{-n} \binom{n}{k} + \frac{1}{2} 2^{-n} \sum_{k=0}^{n} k(-1)^k \binom{n}{k}. \end{aligned} \tag{20}$$

The last sum equals zero, and the first sum can be computed by inserting $a = x, b = 1$ in the binomial distribution, and then differentiating the result

$$\sum_{k=0}^{n} k \binom{n}{k} = n 2^{n-1}. \tag{21}$$

Insertion of Eq. (21) in Eq. (20) gives

$$\overline{f}_1 = \frac{n}{3}. \tag{22}$$

Thus, the probability distribution after selection is given by

$$p_2(k) = \frac{f(k)p_1(k)}{\sum_{k=0}^{n} f(k)p_1(k)}. \tag{23}$$

Insertion of Eq. (22) in Eq. (23) gives, finally

$$p_2(k) = 2^{-n} \left(2 + (-1)^k\right) \frac{k}{n} \binom{n}{k}. \tag{24}$$

5. (a) Derivation of the schema theorem, see Appendix B.2.1.
   (b) When decoded, the seven chromosomes give the following variable values: $x_1 = 23$, $x_2 = 13$, $x_3 = 34$, $x_4 = 40$, $x_5 = 7$, $x_6 = 54$, $x_7 = 48$, from which the fitness values easily can be computed as $f_i = 1/(1 + x_i)$. The average fitness $\overline{f}$ equals 0.04709, and the average fitness of S_1 equals $\overline{f}(S_1) = 0.02648$. Inserting these numbers, and the parameters given in the problem formulation, in the schema theorem, the resulting expected number of copies of S_1 becomes $0.93 \approx 1$.

Mattias Wahde, 2007-10-25