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

Exam in FFR 105 (Stochastic optimization algorithms), 2015-10-28, 14.00-18.00, M.
The examiner will visit the exam rooms twice, around 15.00 and around 17.00.

It will be possible to review your results (for the exam and the home problems) during the week starting Nov. 16.

In the exam, 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 course book, or any lecture notes 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 four problems in the exam, and the maximum number of points is 25.

1. (a) *Premature convergence* is a common problem when using genetic algorithms. Define and describe this problem in detail. Next, assuming that (for a particular genetic algorithm) the population size is set to a given, fixed value, and that tournament selection is used, with a fixed tournament selection probability, introduce and describe *two* different ways of preventing premature convergence. (You may describe more than two ways, but you are *not* required to do so: Note that incorrect, additional descriptions may also result in a deduction of points). (3p)

   (b) In genetic algorithms, the concept of *genes* is central. In the biological counterpart, genes serve the purpose of providing the necessary information for generating proteins by means of a process that involves two major steps. Name and *describe* the two steps. (2p)

   (c) Ant colony optimization is based on the cooperative behavior of ants and, in particular, a special form of communication used by ants. Name and describe, in as much detail as possible, this form of communication. (1p)

Continued on the back of this page.

(d) In the Ant system (AS) algorithm, in every iteration, paths are generated probabilistically for each ant. Once the paths have been generated, the pheromone levels are updated. For the case of the standard travelling salesman problem (TSP), describe in detail (with equations and clear descriptions of each equation) exactly how the pheromone levels are updated. Note: You should *only* describe the pheromone update equations, not the entire AS algorithm! (2p)

(e) Gradient descent is a classical optimization method, in which one follows the negative gradient from a given starting point towards a (local) minimum. In this method, starting from a given point $\mathbf{x}_j$ (where $\mathbf{x}$ is a vector and the index enumerates the iterations) one computes iterates such that, once the search direction has been determined, the next iterate $\mathbf{x}_{j+1}$ will depend only on the step length η, so that the *function value* at that point can be expressed as some function $\phi(\eta)$. Consider now the problem of minimizing the function $f(x_1, x_2) = 2x_1^2 + 3x_1x_2 + x_2^2 - 4$ using gradient descent, starting from the point $(x_1, x_2) = (1, 1)$. Find, and write down, the search direction (i.e. a vector with two components) and the expression for the next iterate, inserting numerical values. Then derive (and simplify as much as possible) the expression for $\phi(\eta)$, again with numerical values inserted. Give your answer in the form $a_2\eta^2 + a_1\eta + a_0$, where a_0, a_1, and a_2 are constants. Note: You do *not* have to carry out the line search. It is sufficient that you find the search direction, the next iterate, and $\phi(\eta)$! (2p)

2. In order to study GAs analytically, one often uses functions of unitation, i.e. objective functions in which the fitness depends only on the number (j) of ones in the chromosome.

   (a) Consider a simple GA, with a population size of 1, and where the (binary) chromosome is changed using mutations only. The new chromosome is kept if it is better (higher fitness) than the previous one, otherwise it is discarded. Using the Onemax function ($f(j) = j$) as the fitness function, and setting the mutation rate p_{mut} to k/m, where m is the chromosome length, and $k \ll m$ is a positive integer, *derive* an estimate for the runtime of this GA, i.e. the number of iterations required to reach the global optimum. (3p)

   (b) In some cases, one makes the further assumption that the population size is infinite. Consider such a case, in which the (binary) chromosomes are initialized randomly, and where a GA with selection only (i.e. no crossover or mutations) is applied to the problem of maximizing the function (of unitation)

$$f(j) = j\left(1 - \frac{j}{m}\right), \tag{1}$$

   where m, again, is the chromosome length. Compute

   i. The average fitness of the initial population. (1p)

   ii. The probability distribution $p_2(j)$ in the second generation (i.e. after one fitness-proportional selection step). (1p)

3. (a) The Lagrange multiplier method depends on a particular relation involving the objective function $f(x_1, x_2, \ldots)$ and the equality constraint function (assuming, here, that there is only one such constraint) $h(x_1, x_2, \ldots)$. Write down this relation (in equation form) and also explain *why* the (local) optima of the objective function, subject to the constraint, occur at the points where this relation holds. You *should* draw a figure (for the case of two dimensions) as a part of your explanation, but you must also describe the figure clearly. (2p)

   (b) Use the Lagrange multiplier method to find the minimum value and the maximum value of the function

$$f(x_1, x_2) = x_1^2 x_2 + 2x_2, \tag{2}$$

   subject to the constraint

$$x_1^2 + x_2^2 - 1 = 0. \tag{3}$$

   (2p)

4. (a) In particle swarm optimization (PSO), there is a specific mechanism for handling the tradeoff between exploration and exploitation. Write down the general equation for the velocity updates in PSO and describe, in detail, the mechanism just mentioned. (2p)

   (b) Consider now a simple one-dimensional application of PSO, in which one is trying to minimize the function $f(x) = (x - \frac{1}{4})^2$, using a swarm size of three. Initially the three particles are located at $x = -1/3$ (particle 1), $x = 0$ (particle 2), and $x = 3/4$ (particle 3), and their speeds are $v = 3$ (particle 1), $v = 1/4$ (particle 2) and $v = -1$ (particle 3) The parameters α and Δt are both equal to 1, w is (here) kept *constant* at the value 1, and $c_1 = c_2 = 2$. Moreover, assume (somewhat unrealistically) that the random numbers q and r are always equal to 1. The initial range $[x_{\min}, x_{\max}]$ is equal to $[-2, 2]$, and the particle speeds are thus restricted to a maximum of 4. Given these parameters, determine, under the PSO algorithm

      i. the velocities and positions of all particles after one iteration (i.e. one updating step for both velocities and positions). (2p)

      ii. the velocities and positions of all particles after two iterations (2p)

Stochastic optimization methods (FFR 105), 2015
Solutions to the exam (2015-10-28)

1. (a) Premature convergence occurs when the population converges to a suboptimal solution. This can happen when, in the early generations, one or a few individuals have much higher fitness than the others, but still well below the best possible fitness. In many cases, the fitness landscape may have very narrow peaks, making it difficult to find those better solutions.

   Through selection and crossover, the (relatively speaking) high-fitness individuals quickly spread their genetic material in the population and, in some cases, the population may then become stuck near the suboptimal solution, unless one or a few individuals happen to stumble upon a path towards a better solution before premature convergence has occurred.

   Premature convergence can be prevented in many ways, for example using varying mutation rates (such that the mutation rate is increased whenever the diversity of the population becomes too low, and vice versa) or some form of mating restriction (for example, by means of diffusion models, in which the individuals are placed on an imaginry grid and where each individuals is only allowed to mate with the nearest neighbours).

   (b) The two steps are called transcription and translation. In transcription, the information in a gene (in the form of a sequence of bases, from the alphabet A, C, G, and T) is read by RNA polymerase, resulting in an mRNA molecule, containing the same information (albeit coded slightly differently) as the gene. In translation, the mRNA molecule is used as a template when forming a chain of amino acids (i.e. a protein). Each codon, i.e. a sequence of three bases in the mRNA molecule, e.g. CAA, encode a particular amino acid. Some codons encode the start and stop command. Once the stop command has been reached the amino acid chain is complete.

   (c) The form of communication is referred to as stigmergy. This is a process of indirect communication by means of local modification of the environment, in which an ant deposits a volatile hydrocarbon (a pheromone) that other ants can perceive. Ants tend to move in the direction of highest pheromone scent. Note that the pheromones will evaporate after a while, unless the path is replenished by additional ants.

   (d) The pheromones are updated as follows: Let D_k denote the length of the tour generated by ant k. The pheromone level on edge e_{ij} is then modified as

$$\Delta\tau_{ij}^{[k]} = \frac{1}{D_k}, \tag{1}$$

   if ant k traversed the edge e_j. If not, $\Delta\tau_{ij}^{[k]} = 0$. Once all ants have been considered, the total change in the pheromone level on edge e_{ij} is computed as

$$\Delta\tau_{ij} = \sum_{k=1}^{N} \Delta\tau_{ij}^{[k]}, \tag{2}$$

   where N is the number of ants. Finally, evaporation is applied, so that

$$\tau_{ij} \leftarrow (1-\rho)\tau_{ij} + \Delta\tau_{ij}, \tag{3}$$

where ρ is the evaporation rate (typically set to 0.5).

(e) The search direction is the negative gradient ($-\nabla f$). Here, the gradient takes the form

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

where T denotes the transpose of the vector. Thus,

$$-\nabla f(x_1, x_2)|_{x_1=1, x_2=1} = -(7, 5)^T. \tag{5}$$

In gradient descent, the next iterate $\mathbf{x}_{i+1}$ is given by

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \eta \nabla f(\mathbf{x}_i). \tag{6}$$

At the point $(1, 1)^T$, one then obtains

$$\mathbf{x}_{i+1} = (1, 1)^T - \eta(7, 5)^T = (1 - 7\eta, 1 - 5\eta)^T. \tag{7}$$

At this point, the function thus becomes

$$\begin{aligned} \phi(\eta) &\equiv f(1 - 7\eta, 1 - 5\eta) = 2(1 - 7\eta)^2 + 3(1 - 7\eta)(1 - 5\eta) + (1 - 5\eta)^2 - 4 \\ &= 2 - 28\eta + 98\eta^2 + 3 - 15\eta - 21\eta + 105\eta^2 + 1 - 10\eta + 25\eta^2 - 4 = \\ &\quad 228\eta^2 - 74\eta + 2. \end{aligned} \tag{8}$$

2. (a) Proof: See Sect. B2.4 in the course book (pp. 181-182).

(b) i. With random initialization the initial probability distribution becomes

$$p_1(j) = 2^{-m} \binom{m}{j}. \tag{9}$$

The average fitness of the initial population can be computed as

$$\overline{F}_1 = \sum_{j=0}^{m} f(j) p_1(j) = 2^{-m} \sum_{j=0}^{m} j \binom{m}{j} - \frac{2^{-m}}{m} \sum_{j=0}^{m} j^2 \binom{m}{j}. \tag{10}$$

Starting from the binomial theorem, with $a = x, b = 1$, taking the derivative with respect to x, and then setting $x = 1$, one obtains

$$\sum_{j=0}^{m} j \binom{m}{j} = m 2^{m-1}. \tag{11}$$

If, instead of setting $x = 1$, one instead multiplies by x, takes the derivative again, and finally sets $x = 1$, one gets

$$\sum_{j=0}^{m} j^2 \binom{m}{j} = m(m+1) 2^{m-2}. \tag{12}$$

Thus, inserting the expressions for these two sums, one finally obtains

$$\overline{F}_1 = 2^{-m} m 2^{m-1} - \frac{2^{-m}}{m} m(m+1) 2^{m-2} = \frac{m}{2} - \frac{m+1}{4} = \frac{m-1}{4}. \tag{13}$$

ii. The probability distribution in the second generation is given by

$$p_2(j) = \frac{f(j)p_1(j)}{\sum_{j=0}^{m} f(j)p_1(j)} = \frac{f(j)p_1(j)}{\overline{F}_1} = \frac{4j}{m-1}\left(1 - \frac{j}{m}\right) 2^{-m} \binom{m}{j}. \quad (14)$$

3. (a) The condition is that the gradient of f should be parallel to the gradient of h, i.e. that $\nabla f + \lambda \nabla h = 0$, where λ (the Lagrange multiplier) is a parameter. This relation between the gradients can be understood by considering the level curves of f: By drawing a figure showing those levels curves, as well as the constraints, one can illustrate the fact that local optima occur where the gradient of f is parallel to the gradient of h. At those points, any movement along the constraint curve $h = 0$ will result in either an increase of f (at a local minimum) or a decrease of f (at a local maximum). See also Fig. 2.8 in the course book.

(b) In this case, the function $L(x_1, x_2, \lambda)$ takes the form

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda h(x_1, x_2) = x_1^2 x_2 + 2x_2 + \lambda(x_1^2 + x_2^2 - 1). \quad (15)$$

Setting the gradient of L to zero, one finds

$$\frac{\partial L}{\partial x_1} = 2x_1 x_2 + 2\lambda x_1 = 0, \quad (16)$$

$$\frac{\partial L}{\partial x_2} = x_1^2 + 2 + 2\lambda x_2 = 0, \quad (17)$$

$$\frac{\partial L}{\partial \lambda} = x_1^2 + x_2^2 - 1 = 0. \quad (18)$$

The first equation gives $x_1 = 0$ or $\lambda = -x_2$. With $x_1 = 0$, the third equation gives $x_2 = \pm 1$. Thus, the two points $(0, 1)^T$ and $(0, -1)^T$ are obtained. If instead $\lambda = -x_2$, the second equation gives $x_1^2 = -2 + 2\lambda^2$. Inserting this into the third equation, one gets

$$-2 + 2\lambda^2 + \lambda^2 - 1 = 0, \quad (19)$$

so that $3\lambda^2 = 3$, i.e. $\lambda = \pm 1$. Thus $x_2 = -\lambda = \pm 1$ and $x_1^2 = -2 + 2\lambda^2 = 0$. This again gives the two points already considered above, namely $(0, 1)^T$ and $(0, -1)^T$. The function takes the value 2 at $(0, 1)^T$ and -2 at $(0, -1)^T$. Thus, the maximum value of 2 occurs at $(0, 1)^T$, and the minimun value of -2 occurs at $(0, -1)^T$.

4. (a) The velocity update for particle i is given by

$$v_{ij} \leftarrow w v_{ij} + c_1 q \left(\frac{x_{ij}^{\text{pb}} - x_{ij}}{\Delta t} \right) + c_2 r \left(\frac{x_j^{\text{sb}} - x_{ij}}{\Delta t} \right), \; j = 1, \ldots, n, \quad (20)$$

where w, the inertia weight, handles the tradeoff between exploration and exploitation. If $w > 1$, exploration is favored. If instead $w < 1$, the particle focuses on exploitation of the results already found. Normally, one starts with a value of w of around 1.4, then reduces w by a factor $\beta \approx 0.99$ until w reaches a lower limit of around $0.3 - 0.4$, where it is then kept constant.

(b) i. Initially, the function values are 49/144 (particle 1), 1/16 (particle 2), and 1/4 (particle 3). Thus, the swarm best position is equal to the position of particle 2 (i.e. $x = 0$). With the simplifications, the velocity update takes the form

$$v_i \leftarrow v_i + (x_i^{\text{pb}} - x_i) + (x^{\text{sb}} - x_i), \; i = 1, 2, 3. \tag{21}$$

One then obtains:

$$v_1 = 3 + 2(-1/3 - (-1/3)) + 2(0 - (-1/3)) = 11/3, \tag{22}$$

$$v_2 = 1/4 + 2(0 - 0) + 2(1/3 - 1/3) = 1/4, \tag{23}$$

and

$$v_3 = -1 + 2(3/4 - 3/4) + 2(0 - 3/4) = -5/2. \tag{24}$$

Thus, using the equation $x \leftarrow x + v$, the new positions become

$$x_1 = -1/3 + 11/3 = 10/3, \tag{25}$$

$$x_2 = 0 + 1/4 = 1/4, \tag{26}$$

$$x_3 = 3/4 - 5/2 = -7/4. \tag{27}$$

ii. In the second iteration, the swarm best position is $x = 1/4$, i.e. the position of particle 2 (which, of course, also is the particle best position for that particle). The particle best position is unchanged for particle 1 and particle 3, since the function values at their new positions exceeds those obtained at their initial positions. Using the same equations as above, one obtains

$$v_1 = 11/3 + 2(-1/3 - 10/3) + 2(1/4 - 10/3) = -59/6. \tag{28}$$

However, this value exceeds (in magnitude) the maximum (negative) speed of -4, meaning that the actual speed of the particle will be $v_3 = -4$ instead. For particle 2 one gets

$$v_2 = 1/4 + 2(1/4 - 1/4) + 2(1/4 - 1/4) = 1/4 \tag{29}$$

and for particle 3

$$v_3 = -5/2 + 2(3/4 - (-7/4)) + 2(1/4 - (-7/4)) = 13/2. \tag{30}$$

This value is larger than the limit of 4, so that the actual speed will be $v_3 = 4$ instead. Thus, finally, one obtains

$$x_1 = 10/3 - 4 = -2/3, \tag{31}$$

$$x_2 = 1/4 + 1/4 = 1/2, \tag{32}$$

and

$$x_3 = -7/4 + 4 = 9/4. \tag{33}$$