Stochastic optimization methods (FFR 105), 2011
Solutions to the exam (20111017)

1. (a) Roulette-wheel selection and tournament selection, see pp. 48-50 in the course book. For roulette-wheel selection, the quantities

$$\phi_j = \frac{\sum_{i=1}^{j} F_i}{\sum_{i=1}^{N} F_i} \tag{1}$$

are generated for $j = 1, 2, \ldots, N$. Next a random number $r \in [0, 1[$ is drawn and the selected individual is taken as the individual with the smallest j that satisfies $\phi_j > r$. For tournament selection (with tournament size 2), two individuals are picked randomly from the population. Next a random number $r \in [0, 1[$ is generated. If $r < p_{\text{tour}}$ (typically around 0.7-0.8), the better of the two individuals is chosen, if not, the worse individual is chosen. Tournament selection can also be generalized to the case of tournaments with more than two participants. In that case, the best individual (of the j randomly picked individuals) is selected with probability p_{tour} as just described. If this individual is not selected, it is removed from the tournament, a new random number r is drawn, and the best of the *remaining* individuals is selected with probability p_{tour} etc. Note that both selection methods take place with replacement. That is, a given individual can be selected several times.

(b) The standard PSO is given in Algorithm 5.1 in the book *but with the inertia term added*, see Eq. (5.20) in the book. For full points, the description should include the five steps of the algorithm (initialization, evaluation, best position updates (particle best and swarm best), position and velocity updates, and the return to step 2. The indices i (enumerating particles) and j (enumerating dimensions) should be introduce correctly in all parts of the algorithm; for example, the swarm best vector x_j^{sb} should have only *one* index. Furthermore, the velocity update equation should be clearly described (the cognitive and social terms, with the two constants c_1 and c_2 and the random numbers q and r). The velocity restriction should be defined. The trade-off between exploration and exploitation is taken care of by the inertia term, which should vary from around 1.4 down to 0.3-0.4.

(c) Making a Taylor expansion of $f(x)$, one obtains

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 \equiv f_{[2]}(x). \tag{2}$$

Taking the derivative and setting it to zero, one finds the stationary point. Thus

$$f'_{[2]}(x) = 0 \Leftrightarrow f'(x_0) + (x - x_0)f''(x_0) = 0. \tag{3}$$

Solving this equation, one obtains

$$x^\star = x_0 - \frac{f'(x_0)}{f''(x_0)}. \tag{4}$$

Thus, the iteration rule takes the form

$$x_{j+1} = x_j - \frac{f'(x_j)}{f''(x_j)}. \tag{5}$$

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

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 8 & -5 \\ -5 & 6 \end{pmatrix}, \tag{6}$$

with eigenvalues $7 \pm \sqrt{26}$ which are both larger than zero. Thus, the function is convex.

2. (a) Decoding the chromosome, one obtains:

1214 $\Leftrightarrow$ $r_2 := r_1 + c_1$
3315 $\Leftrightarrow$ $r_3 := r_1 \times c_2$
3123 $\Leftrightarrow$ $r_1 := r_2 \times r_3$
3333 $\Leftrightarrow$ $r_3 := r_3 \times r_3$
1323 $\Leftrightarrow$ $r_3 := r_2 + r_3$
4213 $\Leftrightarrow$ $r_2 := r_1 / r_3$

With the initial values $r_1 = x$, $r_2 = r_3 = 0$, and with $c_1 = 1, c_2 = 2$, and $c_3 = -1$, one obtains:

Step 1: $r_1 = x, r_2 = x + 1, r_3 = 0$
Step 2: $r_1 = x, r_2 = x + 1, r_3 = 2x$
Step 3: $r_1 = 2x(x + 1), r_2 = x + 1, r_3 = 2x$
Step 4: $r_1 = 2x(x + 1), r_2 = x + 1, r_3 = 4x^2$
Step 5: $r_1 = 2x(x + 1), r_2 = x + 1, r_3 = 4x^2 + x + 1$
Step 6: $r_1 = 2x(x + 1), r_2 = 2x(x + 1)/(4x^2 + x + 1), r_3 = 4x^2 + x + 1$

Thus, the answer is

$$\hat{f}(x) = \frac{2x^2 + 2x}{4x^2 + x + 1}. \tag{7}$$

(b) The only difference compared to the case considered above is that the first instruction now takes the form

1211 $\Leftrightarrow$ $r_2 := r_1 + r_1$

With the same initial values as above, one obtains

Step 1: $r_1 = x, r_2 = 2x, r_3 = 0$
Step 2: $r_1 = x, r_2 = 2x, r_3 = 2x$
Step 3: $r_1 = 4x^2, r_2 = 2x, r_3 = 2x$
Step 4: $r_1 = 4x^2, r_2 = 2x, r_3 = 4x^2$
Step 5: $r_1 = 4x^2, r_2 = 2x, r_3 = 4x^2 + 2x$
Step 6: $r_1 = 4x^2, r_2 = 4x^2/(4x^2 + 2x), r_3 = 4x^2 + 2x$

Thus, the answer is

$$\hat{f}(x) = \frac{4x^2}{4x^2 + 2x} = \frac{2x}{2x + 1} \;(x \neq 0) \tag{8}$$

3. (a) It is easy to see that the nearest-neighbour path starting from node 1 is $1 \to 2 \to 3 \to 4 \to 5$ (with a return to node 1 as the final step). The equation for determining the probability of a move from node j to node i takes the following form

$$p(e_{ij}|S) = \frac{\tau_{ij}^{\alpha}\eta_{ij}^{\beta}}{\sum_{\nu_l \notin L_T(S)} \tau_{lj}^{\alpha}\eta_{lj}^{\beta}}, \tag{9}$$

where $\eta_{ij} = 1/d_{ij}$. Since the pheromone levels are equal on all edges, the τ_{ij} terms cancel out, and one is left with the expression

$$p(e_{ij}|S) = \frac{\eta_{ij}^{\beta}}{\sum_{\nu_l \notin L_T(S)} \eta_{lj}^{\beta}} \tag{10}$$

In node 1, there are four possible moves, with distances $d_{21} = 2$, $d_{31} = \sqrt{10}$, $d_{41} = \sqrt{13}$, and $d_{51} = \sqrt{5}$. With $\beta = 2$ the probability of moving to node 2 becomes

$$p(e_{21}|S = \{\nu_1\}) = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{10} + \frac{1}{13} + \frac{1}{5}} \approx 0.398773. \tag{11}$$

In node 2, there are three possible moves, with distances $d_{32} = \sqrt{2}$, $d_{42} = \sqrt{5}$, and $d_{52} = \sqrt{13}$. Thus

$$p(e_{32}|S = \{\nu_1, \nu_2\}) = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{5} + \frac{1}{13}} \approx 0.643564. \tag{12}$$

In node three, there are two possible moves, with distances $d_{43} = 1$ and $d_{53} = \sqrt{17}$. Thus

$$p(e_{43}|S = \{\nu_1, \nu_2, \nu_3\}) = \frac{1}{1 + \frac{1}{17}} \approx 0.944444. \tag{13}$$

The remaining steps, from node 4 to node 5 and then back to node 1, take place with probability 1. Thus, the probability of following the nearest-neighbour path, starting at node 1, becomes

$$p_{12345} = p_{21} \times p_{32} \times p_{43} \approx 0.242. \tag{14}$$

(b) The length of the nearest-neighbour path starting from node 1 equals

$$L_{12345} = 2 + \sqrt{2} + 1 + 4 + \sqrt{5} \approx 10.65 \tag{15}$$

length units. The initial pheromone level τ_{ij} is thus equal to

$$\tau_{ij} = \frac{1}{\rho D_{\text{nn}}} \approx 0.1878. \tag{16}$$

In MMAS, only the best ant is allowed to deposit pheromone. One can easily see that the path of the fourth ant is the shortest (i.e. the best). Since this path is the nearest-neighbour path considered above, we can write

$$\Delta\tau_{ij}^{[\text{b}]} = \frac{1}{D_{\text{nn}}} \tag{17}$$

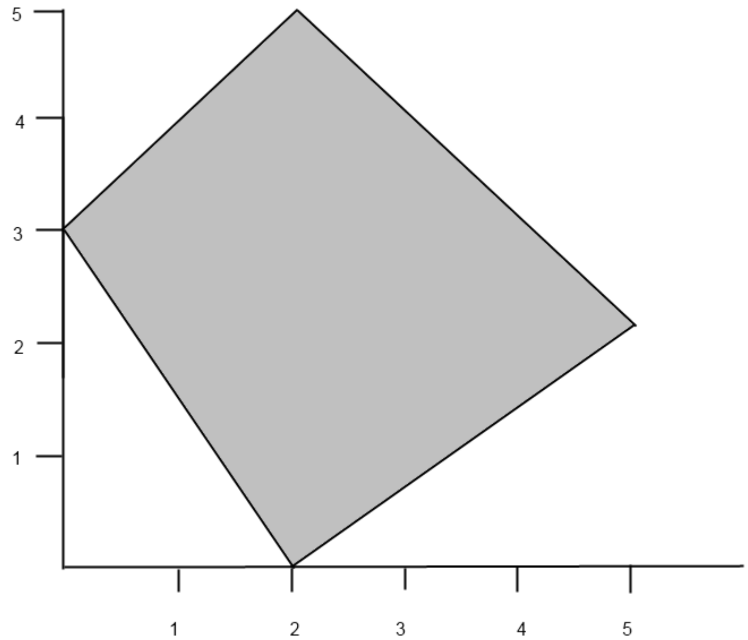

Figure 1: The feasible region for Problem 4.

for the edges in that path, namely e_{21}, e_{32}, e_{43}, e_{54} and e_{15}. the pheromones are updated as

$$\tau_{ij} \leftarrow \tau_{ij}(1-\rho) + \Delta\tau_{ij}^{[\mathrm{b}]} = \frac{1}{\rho D_{\mathrm{nn}}}(1-\rho) + \frac{1}{D_{\mathrm{nn}}} = \frac{1}{\rho D_{\mathrm{nn}}}. \tag{18}$$

That is, the pheromone levels remain unchanged on these edges. For all other edges, the pheromones change as

$$\tau_{ij} \leftarrow \tau_{ij}(1-\rho) = \frac{1}{\rho D_{\mathrm{nn}}}(1-\rho) \approx 0.0939. \tag{19}$$

However, since MMAS introduces a lower pheromone limit, in this case 0.1, the pheromone levels on those edges will be equal to 0.1 rather than 0.0939.

4. (a) The penalty term takes the form

$$p(\mathbf{x};\mu) = \mu\left(\sum_{i=1}^{m}(\max\{g_i(\mathbf{x}),0\})^2 + \sum_{i=1}^{k}(h_i(\mathbf{x}))^2\right), \tag{20}$$

where $g_i(\mathbf{x})$ and $h_i(\mathbf{x})$ are the (m) inequality and (k) equality constraints, respectively and μ is a positive parameter that determines the magnitude of the penalty.

(b) Using the four constraints, one can plot the feasible region, see the figure above. It is easy to see that the unconstrained minimum occurs at $(x_1, x_2) = (6, 7)$. Starting at this point, one can see that, in fact, only the third constraint is violated here. Thus, at this point, one can write objective function as

$$f_{\mathrm{p}}(\mathbf{x};\mu) = (x_1-6)^2 + (x_2-7)^2 + \mu(x_1+x_2-7)^2. \tag{21}$$

Setting the gradient to zero, one obtains

$$\frac{\partial f_{\mathrm{p}}}{\partial x_1} = 2x_1 - 12 + 2\mu(x_1 + x_2 - 7) = 0 \tag{22}$$

and

$$\frac{\partial f_{\mathrm{p}}}{\partial x_2} = 2x_2 - 14 + 2\mu(x_1 + x_2 - 7) = 0. \tag{23}$$

From these equations, one finds a single solution, namely

$$x_1(\mu) = \frac{6(1+\mu)}{1+2\mu}, \tag{24}$$

$$x_2(\mu) = 7 - \frac{6\mu}{1+2\mu}. \tag{25}$$

With $\mu \to \infty$ one obtains $(x_1, x_2) = (3, 4)$. Moreover, for any finite value of μ, the point $(x_1(\mu), x_2(\mu))$ violates the third constraint (and only that constraint), as can be seen by studying the constraints and the feasible region in the figure. One can also see that $f_{\mathrm{p}}(\mathbf{x}; \mu)$ is strictly convex for any $\mu > 0$, so the point (3,4) is the global minimum of f subject to the constraints.