Stochastic optimization methods (FFR 105), 2008
Solutions to the exam (2008-10-22)

1. (a) i. Elitism is the process of transferring an unchanged copy of the best individual in the current generation, to the next generation. This is done, for example, by keeping track (during evaluation) of the index (in the population) of the best individual. Then, when making the new generation, one may start by inserting a copy of the best individual (it can also be inserted at the end of the procedure that generates the new individual). See also p. 55 in the course book.

   ii. In fitness ranking, one reassigns fitness values starting from the raw fitness values obtained during evaluation of individuals. The standard way to carry out ranking is to set new (ranked) fitness values as

$$F_i^{\text{rank}} = (N + 1 - R(i)), \tag{1}$$

   where F_i^{rank} is the new fitness value of individual i, N is the population size, and $R(i)$ is the ranking of individual i. The ranking is defined such that the best individual gets ranking $R(i) = 1$, the second best ranking $R(i) = 2$ etc. See also p. 51 in the course book.

   iii. Creep mutations are used in connection with real-number encoding. These mutations generally change the value (allele) of a gene by a smaller amount than the ordinary full-range mutations. In creep mutation, the new value of a gene is obtained based on a distribution centered on the old value, and with a range that is typically smaller than the (entire) allowed range of the gene. Thus

$$g \leftarrow \Psi(g), \tag{2}$$

   where g denotes the value of g, and Ψ the distribution. A common special case is to use a uniform distribution, in which case the mapping takes the form

$$g \leftarrow g - C_{\text{r}}/2 + C_{\text{r}}r, \tag{3}$$

   where C_{r} is the creep rate and r is a uniform random number in $[0, 1]$. In case the new value ends up outside the allowed range, it is modified to the nearest limit.

(b) i. Using roulette-wheel selection, the probability of selecting individual 4 can be written as

$$p_4 = \frac{F_4}{F_1 + F_2 + F_3 + F_4 + F_5} = \frac{16}{55} \approx 0.291. \tag{4}$$

   ii. In the case of tournament selection with tournament size 2, there are $5 \times 5 = 25$ possible tournaments, since the individuals are chosen (for the tournament) with replacement. Thus the possible pairs of individuals are $(1, 1), (1, 2), \ldots (5, 5)$. Of these 25 pairs (which occur with equal probability, namely 1/25), 9 involve individual 4: (1,4), (2,4), (3,4), (4,4), (4,5), (4,1), (4,2), (4,3), (5,4). For six of the pairs individual 4 is the better individual (and is thus selected with probability p_{tour}) whereas for two of the pairs ((4,5) and (5,4)) the other individual is better, so that individual 4 is selected only with probability $1 - p_{\text{tour}}$. For the pair (4,4), individual 4 is obviously selected with probability 1. Thus, summarizing, the probability of selecting individual 4 equals

$$\frac{1}{25}(6p_{\text{tour}} + 2(1 - p_{\text{tour}}) + 1) = 0.24. \tag{5}$$

(c) The proof can be found on p. 173 of the course book.

(d) Convexity of functions can be studied using the Hessian matrix. More specifically, a function $f(x_1, x_2)$ is convex if the Hessian

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

is positive definite, i.e. has positive eigenvalues. In this particular case, the Hessian becomes

$$H = \begin{pmatrix} 8 & -3 \\ -3 & 4 \end{pmatrix}. \tag{7}$$

The eigenvalues are obtained from the equation

$$(8 - \lambda)(4 - \lambda) - 3 \times 3 = 0. \tag{8}$$

Solving this equation, one obtains

$$\lambda_{1,2} = 6 \pm \sqrt{13} > 0. \tag{9}$$

Thus, the function is convex.

(e) In PSO, the tradeoff between exploration and exploitation is handled using the inertia weight w. The velocities change according to

$$v_{ij} \leftarrow w v_{ij} + c_1 q \left(\frac{x_{ij}^{\text{pb}} - x_{ij}}{\Delta t} \right) + c_2 r \left(\frac{x_j^{\text{sb}} - x_{ij}}{\Delta t} \right), \; j = 1, \ldots, n, \tag{10}$$

where x_{ij} denotes position component j of particle i, v_{ij} denotes velocity component j of particle i, c_1 and c_2 are constants, x_{ij}^{pb} are the components of the best position found by particle i and x_j^{sb} are the components of the best position found by any particle in the swarm. If $w > 1$, the search puts more emphasis on exploration, since the cognitive and social components (the terms involving c_1 and c_2) then play a less significant role than if $w < 1$, in which case the PSO algorithm tries to exploit the results already found, as encoded in the cognitive and social components. Initially, w is typically set to a value larger than 1 (1.4, say), and is then lowered down to a limit of around 0.3-0.4. A common procedure for reducing w is through multiplication by a factor $\beta \in]0, 1]$ (often very close to 1).

2. Local minima are found at stationary points, i.e. at points where the gradient of f is equal to the zero vector. For this particular function, the requirement that the gradient should vanish yields the two equations

$$\frac{\partial f}{\partial x_1} = 4x_1 - 4 = 0 \tag{11}$$

$$\frac{\partial f}{\partial x_2} = 2x_2 + 2 = 0 \tag{12}$$

with the solution $P_1 = (1, -1)^{\mathrm{T}}$. The boundary $2x_1^2 + x_2^2 = 12$ remains to be checked. This can be done using, for example, the method of Lagrange multipliers. However, even easier is to note that $\mathbf{S}$ is a convex set, and that $f(x_1, x_2)$ is a convex function (the eigenvalues of the Hessian are 4 and 2, i.e. both are positive), so that any local minimum must also be a global minimum. Thus, the minimum value of f over S is equal to f(1,-1) = -3.

3. (a) A detailed description of AS can be found on pp. 105-107 in the course book. For full points, the description should contain all the steps (1-4), as well as clear explanations of (1) pheromone initialization, (2) probabilistic path generation, (3) and the rules for updating pheromones.

   (b) The main differences between MMAS and AS are that

   - In MMAS, only the ant generating the best solution is allowed to deposit pheromone. The definition of the best solution is typically changes during a run, so that one uses *best so far* for some iterations, then *best in current iteration* for some iterations etc.
   - In MMAS, one introduces limits on the pheromone levels. Thus, if the pheromone level τ_{ij} on a given edge e_{ij} falls below $\tau_{\min}$, it is set to $\tau_{\min}$. Similarly, if the pheromone level τ_{ij} exceeds $\tau_{\max}$, it is set to $\tau_{\max}$.
   - In MMAS, pheromones are initialized to the maximum level, i.e. such that

     $$\tau_{ij} = \tau_{\max} \ \forall (i, j) \in \{1, n\}. \tag{13}$$

     $\tau_{\max}$ is set as $1/(\rho D_b)$, where ρ is the evaporation rate and D_b is the length of the current best tour.

   (c) The proof is given on p. 183 in the course book.

4. (a) The probability of not mutating any of the 1s equals $(1 - p_{\mathrm{mut}})^{m-l}$ (since mutations are independent of each other), and the probability of mutating at least one of the 0s equals $1 - (1 - p_{\mathrm{mut}})^l$. Thus, the probability for the combination of these two events (which can be taken as an approximation of the probability of an improvement as stated in the problem formulation) equals

   $$P(l, p_{\mathrm{mut}}) = (1 - p_{\mathrm{mut}})^{m-l}(1 - (1 - p_{\mathrm{mut}})^l). \tag{14}$$

   (b) The proof can be found on pp. 181-182 in the course book.

Mattias Wahde, 2008-10-22