Stochastic optimization methods (FFR 105), 2016
Solutions to the exam (2016-10-26)

1. (a) In holdout validation one divides the data set into three subsets: A training set, a validation set, and a test set. The training set is used for giving feedback to the algorithm during training. At the same time, the performance over the validation set is also measured (but not provided to the training algorithm). Training continues until there is a significant drop in the validation performance. At that point, the training is stopped, and the system with the best *validation* performance is selected. Next, the performance over the previously unused test set is measured, and can be taken as the true performance of the system. A typical percentage division between training, validation, and test is 60-20-20.

   (b) i. 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.

   ii. 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.

   (c) Description of PSO, see the course book, pp. 121-123. The tradeoff between exploration and exploitation is handled by the inertia weight (w). If w takes a value above 1, exploration is favoured since, in that case, the particle's acceleration is dominated by a component its current direction of motion. If instead $w < 1$, exploitation is favoured. Typically w is initialized around 1.4, and is then allowed to drop by a constant factor (0.99, say) in each iteration, until it reaches around 0.3-0.4, after which point w is kept constant.

2. (a) Iterates are formed as

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

In this case,

$$f'(x) = 8 - \frac{1}{x}, \tag{2}$$

and

$$f"(x) = \frac{1}{x^2}. \tag{3}$$

Thus,

$$x_{j+1} = x_j - \frac{8 - \frac{1}{x}}{\frac{1}{x^2}} = 2x_j - 8x_j^2. \tag{4}$$

Starting from $x_0 = 0.075$, the following table is obtained

j	x_j
0	0.07500000
1	0.10500000
2	0.12180000
3	0.12491808
4	0.12499994
5	0.12499999

In the last step, the difference between the two successive iterates is less than 10^{-7}. From the value obtained, it easy to guess that the true minimum x^* is at $0.125 = 1/8$. One can easily check that indeed $f'(1/8) = 0$. Moreover, since $f"(1/8) = 64 > 0$, the optimum is a minimum.

(b) Rather than minimizing the distance, one can (equivalently) minimize the distance squared. Thus, function to minimize will be

$$f(x_1, x_2) = x_1^2 + x_2^2, \tag{5}$$

with the constraint

$$h(x_1, x_2) = x_1^2 + x_1 x_2 - 1 = 0. \tag{6}$$

Thus L takes the form

$$L(x_1, x_2, \lambda) = x_1^2 + x_2^2 + \lambda(x_1^2 + x_1 x_2 - 1), \tag{7}$$

so that

$$\begin{aligned}
\frac{\partial L}{\partial x_1} &= 2x_1 + \lambda(2x_1 + x_2) = 0, \\
\frac{\partial L}{\partial x_2} &= 2x_2 + \lambda x_1 = 0 \\
\frac{\partial L}{\partial \lambda} &= = x_1^2 + x_1 x_2 - 1 = 0. \qquad (8)
\end{aligned}$$

From the first two equations, one finds $2x_1 = -\lambda(2x_1 + x_2)$ and $2x_2 = -\lambda x_1$, so that

$$x_1(4 + 4\lambda - \lambda^2) = 0. \qquad (9)$$

The solution $x_1 = 0$ does not satisfy the constraint. Thus, instead, one must have $4 + 4\lambda - \lambda^2 = 0$, so that

$$\lambda_{1,2} = 2 \pm 2\sqrt{2} \qquad (10)$$

Considering first $\lambda_1 = 2 + 2\sqrt{2}$, using $2x_2 = -\lambda x_1$, one can then write (using the constraint)

$$x_1^2 - \frac{\lambda_1}{2} x_1^2 = 1, \qquad (11)$$

so that

$$x_1^2 = \frac{2}{2 - \lambda_1} \to x_1 = \pm \frac{1}{\sqrt[4]{2}} \approx \pm 0.8498964153. \qquad (12)$$

and

$$x_2 = \ldots = \pm \frac{-1 + \sqrt{2}}{\sqrt[4]{2}} = \pm 0.3483106995. \qquad (13)$$

If one instead tries with λ_2, one obtains imaginary values, so those solutions can be excluded. Thus, there are two possible solutions, namely

$$(x_1^\star, x_2^\star) = \pm\left(\frac{1}{\sqrt[4]{2}}, \frac{-1 + \sqrt{2}}{\sqrt[4]{2}}\right). \qquad (14)$$

Both points are equidistant from the origin.

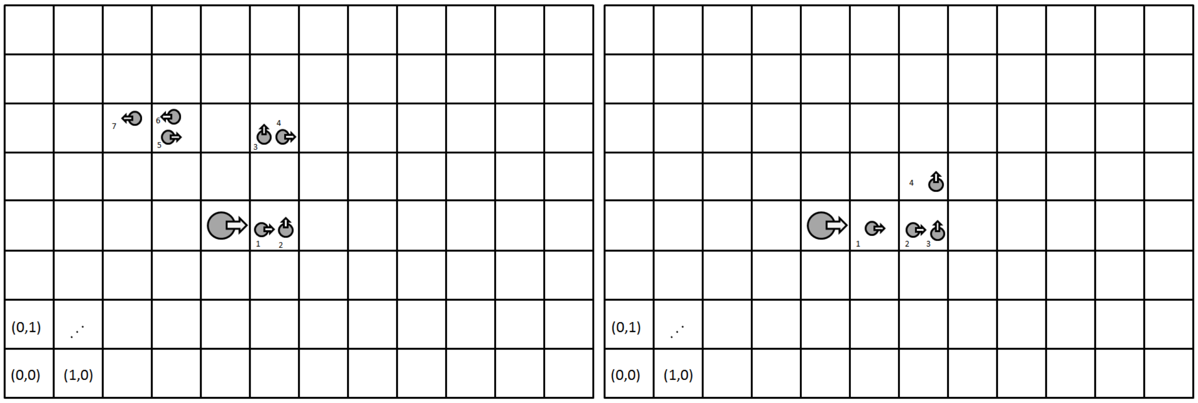

Figure 1: Left panel: The trajectory for Problem 3(a). Right Panel: The trajectory for Problem 3(b).

3. (a) The chromosome $c_1 = 11311241223211$ can be decoded as
   11 $\rightarrow$ Move forward one step
   31 $\rightarrow$ Turn left, 90 degrees
   12 $\rightarrow$ Move forward two steps
   41 $\rightarrow$ Turn right 90 degrees
   22 $\rightarrow$ Move backward, two steps
   32 $\rightarrow$ Turn left $2 \times 90 = 180$ degrees
   11 $\rightarrow$ Move forward one step

   (b) The crossover results in the two chromosomes $c_3 = 11113111$ and $c_4 = 22311241223221$. When decoded, the shorter of the two new chromosomes results in the following sequence of movements:
   11 $\rightarrow$ Move forward one step
   11 $\rightarrow$ Move forward one step
   31 $\rightarrow$ Turn left, 90 degrees
   11 $\rightarrow$ Move forward one step
   The trajectories are shown in Fig. 1.

4. (a) The probabilities are obtained using Eq. (4.3) in the book. Since the pheromone levels are equal on all edges, they will disappear from the equation. The resulting probabilities are thus

$$p(I \rightarrow II) = \frac{3}{3+2+2} = \frac{3}{7}, \tag{15}$$

since the ant can move to any other node (II, III, or IV) from node 1. Continuing, one finds in a similar way

$$p(II \to III) = \frac{3}{4+3} = \frac{3}{7}, \tag{16}$$

$$p(III \to IV) = 1, \tag{17}$$

and

$$p(IV \to I) = 1. \tag{18}$$

Thus, the probability of obtaining the desired path is equal to $9/49 \approx 0.18367$.

(b) The pheromone levels will be

$$\tau_{ij} \leftarrow \tau_{ij}(1-\rho) + f = 1 \times 0.5 + 0.5 = 1, \tag{19}$$

for those (note!) four edges that the ant traversed (listed above). For all other edges (including the reverse edges, e.g. I $\to$ IV etc.), only evaporation will occur, so that $\tau_{ij} \leftarrow 0.5$.