Stochastic optimization methods (FFR 105), 2020
Solutions to the exam (2020-10-28)

1. (a) Newton-Raphson's method takes the form

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

so that, for this function,

$$x_{j+1} = x_j - \frac{x_j^3 - x_j - 1}{3x_j^2 - 1} \tag{2}$$

A plot of the function can be seen in Fig. 1 below. One need only plot a few points to realize that the stationary point must be in the interval $[0, 2]$. Starting with $x_0 = 1$, the following sequence is obtained: $x_1 = 1.5$, $x_2 = 1.347826087$, $x_3 = 1.325200399$, $x_4 = 1.324718174$, $x_5 = 1.324717957$. At this point, it can be seen that the first 5 decimal places remain the same. Thus $x^* = 1.32472$.

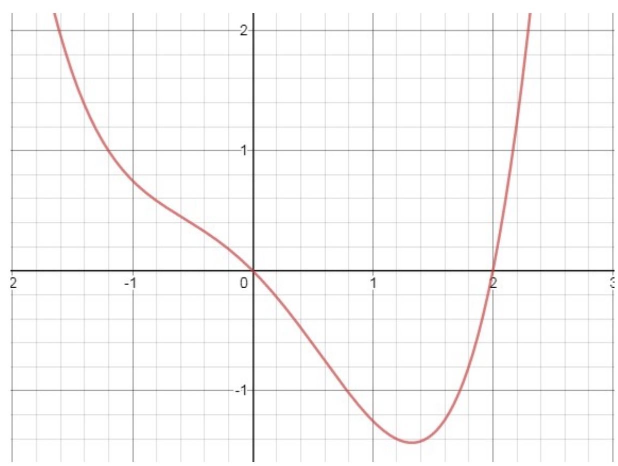

Figure 1: Problem 1a

(b) Here, the goal is to minimize the distance, but it is easier to work with the squared distance (to avoid involving a square root). The squared distance d^2 from any point $(x_1, x_2, x_3)^{\mathrm{T}}$ to the point $(4, 0, 0)^{\mathrm{T}}$ equals

$$d^2 = (x_1 - 4)^2 + x_2^2 + x_3^2. \tag{3}$$

The constraint can be written

$$h(x_1, x_2, x_3) = x_1^2 + x_2^2 - x_3^2 - 1 = 0, \tag{4}$$

so that

$$L(x_1, x_2, x_3, \lambda) = (x_1 - 4)^2 + x_2^2 + x_3^2 + \lambda(x_1^2 + x_2^2 - x_3^2 - 1). \tag{5}$$

Taking the partial derivatives and setting them to zero, one obtains

$$\frac{\partial L}{\partial x_1} = 2(x_1 - 4) + 2\lambda x_1 = 0, \tag{6}$$

$$\frac{\partial L}{\partial x_2} = 2x_2 + 2\lambda x_2 = 0, \tag{7}$$

$$\frac{\partial L}{\partial x_3} = 2x_3 - 2\lambda x_3 = 0, \tag{8}$$

$$\frac{\partial L}{\partial \lambda} = x_1^2 + x_2^2 - x_3^2 - 1 = 0. \tag{9}$$

From Eq. (7) we see that this equation is solved (for any value of x_2) if $\lambda = -1$. However, with $\lambda = -1$, Eq. (6) gives $-8 = 0$, which is a contradiction, meaning that $\lambda = -1$ can safely be excluded.

Now, with $\lambda \neq -1$, Eq. (7) leads to the requirement that $x_2 = 0$. Consider now Eq. (8). This equation is solved (for any value of x_3) if $\lambda = 1$. With $\lambda = 1$, Eq. (6) gives $4x_1 - 8 = 0$, so $x_1 = 2$. From the constraint, i.e. Eq. (9), one then gets $2^2 + 0^2 - x_3^2 = 1$, so that $x_3 = \pm\sqrt{3}$. Thus, the two points $P_1 = (2, 0, \sqrt{3})^{\mathrm{T}}$ and $P_1 = (2, 0, -\sqrt{3})^{\mathrm{T}}$ are obtained.

If instead $\lambda \neq 1$, Eq. (8) is solved only for $x_3 = 0$. The constraint equation then gives $x_1^2 + 0^2 + 0^2 = 1$, so that $x_1 = \pm 1$. Hence, the two points $P_3 = (1, 0, 0)^{\mathrm{T}}$ and $P_4 = (-1, 0, 0)^{\mathrm{T}}$ are obtained.

The distances d can now easily be computed, giving $d(P_1) = d(P_2) = \sqrt{7}$, $d(P_3) = 3$ and $d(P_4) = 5$. Thus, the minimum distance $d_{\min} = \sqrt{7}$ is found for P_1 and P_2.

(c) Here, one can use the analytical method described in the course book, where one first finds all stationary points in the interior of the set S, and then any stationary points of the restriction of f to the boundary ∂S.

Starting with the interior of S, the partial derivatives of $f(x_1, x_2)$ are

$$\frac{\partial f}{\partial x_1} = (1 - x_1^2)e^{\left(x_1 - \frac{1}{3}x_1^3 - x_2^2\right)} \tag{10}$$

and

$$\frac{\partial f}{\partial x_2} = -2x_2 e^{\left(x_1 - \frac{1}{3}x_1^3 - x_2^2\right)} \tag{11}$$

Since the exponential term cannot be zero in this bounded region, the two partial derivatives only take the value 0 if $1 - x_1^2 = 0$ and $-2x_2 = 0$, giving the two points $(1, 0)^{\mathrm{T}}$ and $(-1, 0)^{\mathrm{T}}$.

The next step is to investigate the boundary ∂S. In this case, the set S is a square, so there are four edges. Starting with the edge $x_2 = -2$, $-2 < x_1 < 2$, one finds

$$f(x_1, -2) = \mathrm{e}^{\left(x_1 - \frac{1}{3}x_1^3 - 4\right)} \equiv g(x_1). \tag{12}$$

Taking the derivative of $g(x_1)$ and setting it to zero, one gets

$$g'(x_1) = (1 - x_1^2)\mathrm{e}^{\left(x_1 - \frac{1}{3}x_1^3 - 4\right)} = 0, \tag{13}$$

with the solutions $x_1 = 1$ and $x_1 = -1$. Thus, from this part of the boundary, the points $(1, -2)^{\mathrm{T}}$ and $(-1, -2)^{\mathrm{T}}$ are obtained as points of interest (for further investigation later; see below). Similarly, for $x_2 = 2$, one obtains the two points $(1, 2)^{\mathrm{T}}$ and $(-1, 2)^{\mathrm{T}}$. For the edge $x_1 = -2$, $-2 < x_2 < 2$, the function takes the form

$$f(-2, 0) = \mathrm{e}^{\left(\frac{2}{3} - x_2^2\right)} \equiv h(x_2). \tag{14}$$

Setting the derivative to zero one finds

$$h'(x_2) = -2x_2\mathrm{e}^{\left(\frac{2}{3} - x_2^2\right)} = 0, \tag{15}$$

with the solution $x_2 = 0$. Thus, the point $(-2, 0)^{\mathrm{T}}$ is obtained. In the same way, for $x_1 = 2$ one finds the point $(2, 0)^{\mathrm{T}}$. Finally, one must also consider the corners, i.e. the points $(-2, -2)^{\mathrm{T}}$, $(-2, 2)^{\mathrm{T}}$,$(2, -2)^{\mathrm{T}}$, and $(2, 2)^{\mathrm{T}}$. Thus, in total, there are 12 candidates for the global extrema. The function values at these points are $f(1, 0) = \mathrm{e}^{\frac{2}{3}}$, $f(-1, 0) = \mathrm{e}^{-\frac{2}{3}}$, $f(1, -2) = \mathrm{e}^{-\frac{10}{3}}$, $f(-1, -2) = \mathrm{e}^{-\frac{14}{3}}$, $f(-2, 0) = \mathrm{e}^{\frac{2}{3}}$, $f(2, 0) = \mathrm{e}^{-\frac{2}{3}}$,$f(1, 2) = \mathrm{e}^{-\frac{10}{3}}$, $f(-1, 2) = \mathrm{e}^{-\frac{14}{3}}$,$f(-2, -2) = \mathrm{e}^{-\frac{10}{3}}$, $f(-2, 2) = \mathrm{e}^{-\frac{10}{3}}$,$f(2, -2) = \mathrm{e}^{-\frac{14}{3}}$, and $f(2, 2) = \mathrm{e}^{-\frac{14}{3}}$. Thus, the points $(-2, 0)^{\mathrm{T}}$ and $(1, 0)^{\mathrm{T}}$ are global maxima (where the function takes the value $\mathrm{e}^{\frac{2}{3}}$). The points $(-1, 2)^{\mathrm{T}}$, $(-1, -2)^{\mathrm{T}}$, $(2, 2)^{\mathrm{T}}$, and $(2, -2)^{\mathrm{T}}$ are global minima (where the function takes the value $\mathrm{e}^{-\frac{14}{3}}$).

2. The best individual is individual 5, with fitness 30. Let F_j denote the fitness of individual j, and let p_j denote the probability of selecting that individual.

   (a) For roulette-wheel selection the probability is obtained as

   $$p_5 = \frac{F_5}{\sum_{j=1}^{5} F_j} = \frac{30}{45} = \frac{2}{3} \approx 0.667. \tag{16}$$

   (b) For tournament selection there are 5×5 possible tournaments (since the tournament size is equal to 2), all equally likely. Nine of those tournaments involve individual 5, namely $(1, 5), (2, 5), (3, 5), (4, 5), (5, 5), (5, 1)$,$(5, 2), (5, 3), (5, 4)$. For the tournament (5,5), the probability of selecting individual 5 is equal to 1. For all the other tournaments, the probability is equal to p_{tour} since individual 5 is the better individual in each pair. Thus

   $$p_5 = \frac{1}{25}\left(1 + 8p_{\mathrm{tour}}\right) = \frac{1}{25}(1 + 8 \times 0.75) = 0.280. \tag{17}$$

The main conclusion is that, with roulette-wheel selection, it is much more likely to select individuals with high relative fitness than with tournament selection. This, in turn, increases the risk of premature convergence if roulette-wheel selection is used (without fitness ranking).

3. The minimum of the simple quadratic function $f(x)$ is clearly at $x = 3/4$. Considering the decoding procedure specified in the problem, this corresponds to the chromosome 11000 (which will then be decoded to give $x = 2^{-1} + 2^{-2} = 3/4$). In order to generate this chromosome, an ant must traverse the construction graph such that it first makes two up-moves (at Nodes 1 and 4), and then three down-moves (at Nodes 7, 10, and 15). Note that, at Nodes $2, 3, 5, 6, \ldots$, the move to the next node (i.e. $4, 7, \ldots$) is deterministic and does not produce any output. Consider now the tour of one ant. Use a simplified notation, such that $p(e_{ij}|S)$ is denoted $p_{i,j}$.

   At Node 1, the ant can either go to Node 2 or to Node 3. Noting that the visibility (η_{ij}) is equal to 1 for all edges, the probability of making the required up-move (i.e. going to Node 2, so as to generate $a_0 = 1$) can be computed as

$$p_{2,1} = \frac{\tau_{21}^{\alpha}}{\tau_{21}^{\alpha} + \tau_{31}^{\alpha}} = \frac{0.25}{0.25 + 0.50} = \frac{1}{3}, \tag{18}$$

   where, in the second step, the fact that $\alpha = 1$ has been used. The ant then moves from Node 2 to Node 4, in preparation for the next bit generation step. At Node 4, the probability of making an up-move (to output $a_1 = 1$) equals

$$p_{5,4} = \frac{0.60}{0.60 + 0.45} = \frac{4}{7}. \tag{19}$$

   Next, after moving to Node 7, the ant should then move to Node 9 (to yield $a_2 = 0$). The probability for this move equals

$$p_{9,7} = \frac{0.50}{0.35 + 0.50} = \frac{10}{17}. \tag{20}$$

   Then, after reaching Node 10, the ant should move to Node 12,10 (to generate $a_3 = 0$). The probability for this move equals

$$p_{12,10} = \frac{0.75}{0.50 + 0.75} = \frac{3}{5}. \tag{21}$$

   After going to Node 13, the ant should then move to Node 15, to yield $a_4 = 0$. This probability for making this move is

$$p_{15,13} = \frac{0.90}{0.60 + 0.90} = \frac{3}{5}. \tag{22}$$

   Thus, the probability P of generating 11000 as output equals

$$P = p_{2,1} \times p_{5,4} \times p_{9,7} \times p_{12,10} \times p_{15,13} \approx 0.0403. \tag{23}$$

Now, the population consists of five ants that generate their paths independently of each other, and with the same pheromone levels, as specified in the problem formulation. For any given ant, the probability of *not* finding the required path is equal to $1 - P$. The probability that no ant finds this path thus equals $(1 - P)^5$ and therefore the probability Π that *at least* one ant finds the path is

$$\Pi = 1 - (1 - P)^5 = 0.1861, \tag{24}$$

and this is the answer.

4. Let $\mathbf{x}_i$ denote the position vector of particle i, i =A,B,C, with components x_{ij}, $j = 1, 2$. At initialization, for each particle, $\mathbf{x}_i^{\text{pb}} = \mathbf{x}_i$. Thus, $\mathbf{x}_A^{\text{pb}} = (0, 1)^{\text{T}}$, $\mathbf{x}_B^{\text{pb}} = (1, 1)^{\text{T}}$, $\mathbf{x}_C^{\text{pb}} = (1, 0)^{\text{T}}$. The function values at these points are $f_A = 2$, $f_B = 3$, and $f_C = 1$. Thus, since the objective is to minimize the function, $\mathbf{x}^{\text{sb}} = \mathbf{x}_C^{\text{pb}} = (1, 0)^{\text{T}}$.

   (a) The velocities are computed using the equation

$$v_{ij} \leftarrow w v_{ij} + c_1 q \left(\frac{x_{ij}^{\text{pb}} - x_{ij}}{\Delta t} \right) + c_2 r \left(\frac{x_j^{\text{sb}} - x_{ij}}{\Delta t} \right), \tag{25}$$

   where $c_1 = c_2 = 2$, $w = 1$, $\Delta t = 1$, and $q = r = 0.5$, so that

$$v_{ij} \leftarrow v_{ij} + \left(x_{ij}^{\text{pb}} - x_{ij} \right) + \left(x_j^{\text{sb}} - x_{ij} \right) \tag{26}$$

   Moreover, in this initial step of the PSO algorithm $x_{ij}^{\text{pb}} = x_{ij}$ so that only the first and third term need to be considered. Thus, the new velocities for particle A become

$$v_{A1} = 0 + (1 - 0) = 1, \tag{27}$$

$$v_{A2} = -1 + (0 - 1) = -2. \tag{28}$$

   However, due to the componentwise velocity restriction (given in the problem formulation) v_{A2} is set to -1. Continuing with particle B, one obtains

$$v_{B1} = -1 + (1 - 1) = -1, \tag{29}$$

$$v_{B2} = 0 + (0 - 1) = -1. \tag{30}$$

   Finally, for particle C,

$$v_{C1} = 0 + (1 - 1) = 0, \tag{31}$$

$$v_{C2} = 1 + (0 - 0) = 1. \tag{32}$$

   (b) The positions are obtained as $x_{ij} \leftarrow x_{ij} + v_{ij}$, since $\Delta t = 1$. Thus,

$$x_{A1} = 0 + 1 = 1, \tag{33}$$

$$x_{A2} = 1 - 1 = 0, \tag{34}$$

$$x_{B1} = 1 - 1 = 0, \tag{35}$$

$$x_{B2} = 1 - 1 = 0, \tag{36}$$

$$x_{C1} = 1 + 0 = 1, \tag{37}$$

$$x_{C2} = 0 + 1 = 1. \tag{38}$$

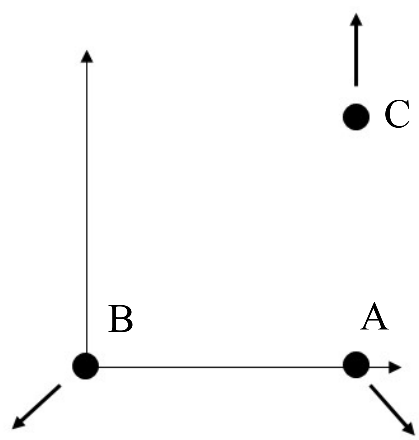

Figure 2: Problem 4c

(c) The new positions and velocities are plotted in the figure above. The function values at the new points are $f_A = 1$, $f_B = 0$, and $f_C = 3$. Thus, comparing with the original function values (see (a) above), one finds that, now, $\mathbf{x}_A^{\text{pb}} = (1, 0)^{\text{T}}$ (i.e. its new position), $\mathbf{x}_B^{\text{pb}} = (0, 0)^{\text{T}}$ (also its new position), $\mathbf{x}_C^{\text{pb}} = (1, 0)^{\text{T}}$ (its previous position). The swarm best (which is also the global minimum of the function) is now at $(0, 0)^{\text{T}}$, i.e. the position of Particle B.