Algorithms for Optimal Decisions
Tutorial 3
Answers

Exercise 1 *Show that the steepest descent direction*

$$-\frac{\nabla f(x_k)}{\|\nabla f(x_k)\|_2} \tag{1}$$

is the solution of the constrained problem:

$$\begin{array}{ll} \min\limits_{d} & \nabla f(x_k)^t d \\ s.t. & \|d\|_2^2 = 1. \end{array} \tag{2}$$

Solution : We need to show that the solution of the constrained problem

$$\begin{array}{ll} \min\limits_{d} & F(d) = \nabla f(x_k)^t d \\ s.t. & G(d) = \|d\|_2^2 - 1 = 0 \end{array} \tag{3}$$

is equal to $d^* = -\frac{\nabla f(x_k)}{\|\nabla f(x_k)\|_2}$.

Observe, initially, that in problem (3) the objective function $F(d)$ is a linear function of $d^t = (d_1, d_2, ..., d_n)$ since

$$\begin{aligned} F(d) &= \left(\frac{\partial f(x_k)}{\partial x_1}, \frac{\partial f(x_k)}{\partial x_2}, \ldots, \frac{\partial f(x_k)}{\partial x_n}\right) \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} = \\ &= \frac{\partial f(x_k)}{\partial x_1} d_1 + \frac{\partial f(x_k)}{\partial x_2} d_2 + \ldots + \frac{\partial f(x_k)}{\partial x_n} d_n = \\ &= \sum_{i=1}^{n} \frac{\partial f(x_k)}{\partial x_i} d_i. \end{aligned}$$

Also, the constraint of the problem (3) is quadratic:

$$\begin{aligned} G(d) &= \|d\|_2^2 - 1 = d^t d - 1 = (d_1, d_2, \ldots, d_n) \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} - 1 = \\ &= d_1^2 + d_2^2 + \ldots d_n^2 - 1 = \sum_{i=1}^{n} d_i^2 - 1. \end{aligned}$$

The Lagrangian of problem (3) is given by:

$$\begin{aligned} L(d, \lambda) &= F(d) + \lambda G(d) = \\ &= \nabla f(x_k)^t d + \lambda(\|d\|_2^2 - 1). \end{aligned} \tag{4}$$

The KKT conditions for optimality of problem (3) are:

$$\nabla_d L(d, \lambda) = \nabla f(x_k) + 2\lambda d = 0 \tag{5}$$

$$\nabla_\lambda L(d, \lambda) = \|d\|_2^2 - 1 = d^t d - 1 = 0. \tag{6}$$

Assuming that $\lambda \neq 0$[1] and solving (5) for d we have:

$$d = -\frac{1}{2\lambda} \nabla f(x_k). \tag{7}$$

Substituting (7) into (6) we have:

$$\begin{aligned} & (-\frac{1}{2\lambda} \nabla f(x_k))^t (-\frac{1}{2\lambda} \nabla f(x_k)) - 1 = 0 \\ \Leftrightarrow \quad & \frac{1}{4\lambda^2} \nabla f(x_k)^t \nabla f(x_k) - 1 = 0 \Leftrightarrow \frac{1}{4\lambda^2} \|\nabla f(x_k)\|_2^2 = 1. \end{aligned} \tag{8}$$

Solving (8) for λ we obtain:

$$\lambda = \frac{1}{2} \|\nabla f(x_k)\|_2. \tag{9}$$

Substituting λ (from (9)) into (7) we have:

$$d = -\frac{1}{2\frac{1}{2}\|\nabla f(x_k)\|_2} \nabla f(x_k) = -\frac{\nabla f(x_k)}{\|\nabla f(x_k)\|_2}. \tag{10}$$

Hence, the pair

[1] otherwise $\nabla f(x_k) = 0$

$$(d_*, \lambda_*) = (-\tfrac{\nabla f(x_k)}{\|\nabla f(x_k)\|_2}, \tfrac{1}{2}\|\nabla f(x_k)\|_2)$$

is the optimum solution of the problem (3), or in other words the descent direction d^* is the optimum solution of (3).

Exercise 2 *Consider the following unconstrained problem:*

$$\max_x f(x) = 2x_1x_2 + 2x_2 - x_1^2 - 2x_2^2. \tag{11}$$

Find its solution using the steepest ascent method starting from the point

$$x^{(0)} = (x_1^{(0)}, x_2^{(0)}) = (0, 0).$$

Solution : The steepest ascent method is the same as the steepest descent method, but it uses the opposite direction. That is, the steepest ascent method moves from point x_k to the point $x_{k+1} = x_k + \tau\nabla f(x_k)$, while the steepest descent method moves from point x_k to the point $x_{k+1} = x_k - \tau\nabla f(x_k)$, where in both cases τ denotes the size of the step we take along the steepest ascent direction, $d_{sa} = \nabla f(x_k)$, and steepest descent direction $d_{sd} = -\nabla f(x_k)$ respectively. Steepest ascent method is used when we want to find the maximum of a function.

First find the gradient of $f(x)$:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_2 - 2x_1 \\ 2x_1 + 2 - 4x_2 \end{bmatrix}. \tag{12}$$

START OF ITERATION 1

At the initial point

$$x^{(0)} = (0,0)^t \text{ we have } \nabla f(x^{(0)}) = (0,2)^t.$$

To begin with first iteration we need to find the next point (which can be described as a better approximation than the initial point $x^{(0)}$ of the optimum point x^*).

We have

$$\begin{aligned} x^{(1)} &= x^{(0)} + \tau \nabla f(x^{(0)}) \Rightarrow \\ \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \tau \begin{bmatrix} 0 \\ 2 \end{bmatrix} \Rightarrow \\ \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} &= \begin{bmatrix} 0 \\ 2\tau \end{bmatrix}. \end{aligned} \tag{13}$$

Substituting $(x_1^{(1)}, x_2^{(1)}) = (0, 2\tau)$ into $f(x)$ we obtain:

$$f(x) = f(x_1, x_2) = f(0, 2\tau) = 4\tau - 8\tau^2. \tag{14}$$

Next we need to find the value of the step size which <u>maximizes</u> (14). It is a univariable function so we can easily find its maximum:

$$\begin{aligned} \frac{\partial}{\partial \tau} f(x_1^{(1)}, x_2^{(1)}) &= \frac{\partial}{\partial \tau}(4\tau - 8\tau^2) = 0 \Rightarrow \\ \Rightarrow \quad 4 - 8 \cdot 2 \cdot \tau &= 0 \Rightarrow \tau = \frac{1}{4}. \end{aligned}$$

Hence the next point $x^{(1)}$ is:

$$x^{(1)} = x^{(0)} + \frac{1}{4} \nabla f(x^{(0)}) = (0, \frac{1}{2})^t. \tag{15}$$

END OF ITERATION 1

Since $\|\nabla f(x^{(1)})\|_2 \neq 0$ we carry on.

START OF ITERATION 2

At the new point

$$x^{(1)} = (0, \tfrac{1}{2})^t \text{ we have } \nabla f(x^{(1)}) = (1, 0)^t.$$

To begin with the next iteration we need to find the next point $x^{(2)}$. We have

$$\begin{aligned} x^{(2)} &= x^{(1)} + \tau \nabla f(x^{(1)}) \Rightarrow \\ \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \end{bmatrix} &= \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix} + \tau \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow \\ \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \end{bmatrix} &= \begin{bmatrix} \tau \\ \frac{1}{2} \end{bmatrix}. \end{aligned} \tag{16}$$

Substituting $(x_1^{(2)}, x_2^{(2)}) = (\tau, \frac{1}{2})^t$ into $f(x)$ we obtain:

$$f(x^{(2)}) = f(x_1^{(2)}, x_2^{(2)}) = f(\tau, \frac{1}{2}) = \tau - \tau^2 + \frac{1}{2}. \tag{17}$$

Next we need to find the value of the step size which <u>maximizes</u> (17). It is a univariable function so we can easily find its maximum:

$$\begin{aligned} \frac{\partial}{\partial \tau} f(x_1^{(2)}, x_2^{(2)}) = \frac{\partial}{\partial \tau}(\tau - \tau^2 + \frac{1}{2}) = 0 \Rightarrow \\ \Rightarrow \quad 1 - 2 \cdot \tau = 0 \Rightarrow \tau = \frac{1}{2}. \end{aligned}$$

Hence the next point $x^{(2)}$ is:

$$x^{(2)} = x^{(1)} + \frac{1}{2} \nabla f(x^{(1)}) = (\frac{1}{2}, \frac{1}{2})^t. \tag{18}$$

END OF ITERATION 2

Since $\|\nabla f(x^{(2)})\|_2 \neq 0$ we carry on.

START OF ITERATION 3

At the new point

$$x^{(2)} = (\tfrac{1}{2}, \tfrac{1}{2})^t \text{ we have } \nabla f(x^{(2)}) = (0,1)^t.$$

To begin with the next iteration we need to find the next point $x^{(3)}$. We have

$$\begin{aligned} x^{(3)} &= x^{(2)} + \tau \nabla f(x^{(2)}) \Rightarrow \\ \begin{bmatrix} x_1^{(3)} \\ x_2^{(3)} \end{bmatrix} &= \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} + \tau \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow \\ \begin{bmatrix} x_1^{(3)} \\ x_2^{(3)} \end{bmatrix} &= \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} + \tau \end{bmatrix}. \end{aligned} \tag{19}$$

Substituting $(x_1^{(3)}, x_2^{(3)}) = (\frac{1}{2}, \frac{1}{2} + \tau)^t$ into $f(x)$ we obtain:

$$f(x^{(3)}) = f(x_1^{(3)}, x_2^{(3)}) = f(\frac{1}{2}, \frac{1}{2} + \tau) = \tau - 2 \cdot \tau^2 + \frac{3}{4}. \tag{20}$$

Next we need to find the value of the step size which <u>maximizes</u> (20). It is a univariable function so we can easily find its maximum:

$$\begin{aligned} & \frac{\partial}{\partial \tau} f(x_1^{(3)}, x_2^{(3)}) = \frac{\partial}{\partial \tau}(\tau - 2 \cdot \tau^2 + \frac{3}{4}) = 0 \Rightarrow \\ \Rightarrow \quad & 1 - 4 \cdot \tau = 0 \Rightarrow \tau = \frac{1}{4}. \end{aligned}$$

Hence the next point $x^{(3)}$ is:

$$x^{(3)} = x^{(2)} + \frac{1}{4} \nabla f(x^{(1)}) = (\frac{1}{2}, \frac{3}{4})^t. \tag{21}$$

END OF ITERATION 3

Since $\|\nabla f(x^{(3)})\|_2 \neq 0$ we carry on.

You can try and do the rest of iterations by yourselves, but there are many. You can also write a computer program to perform those iterations.

A big disadvantage of the steepest descent method is that although it makes satisfactory progress during the initial iterations it may become very slow as it approaches the optimum.

However, it <u>always</u> guarantees to find a point where the value of the objective function is greater than the value of objective function at the previous point.