Conjugate Direction Methods

Video: 5A

Reading:

Chong, An Introduction to Optimization, 2011, Chapter 10 – Sections 10.1-10.2

https://onlinelibrary-wiley-com.proxy1.library.jhu.edu/doi/book/10.1002/9781118033340

The conjugate direction and conjugate gradient methods discussed in this module are applicable to any unimodal differentiable objective function. It's not practically useful to apply them to a quadratic function because there is a much easier way to solve a quadratic optimization problem (setting the gradient to zero). However, applying the conjugate direction and conjugate gradient methods to a quadratic optimization problem helps to illustrate the ideas behind these methods and so this is how we will proceed.

A conjugate direction method finds a basis where the components of vector x in the quadratic optimization problem

$$\min f(x) = \frac{1}{2} x^T Q x + b^T x$$

can be separated and optimized separately.

To illustrate the idea, let's start with a case when Q is a diagonal matrix,

$$Q = \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix} \quad b = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$$

The same optimization problem in the scalar form is

$$\min f(x) = 1?x_1^2 + 2?x_2^2 - 3?x_1 - 4?x_2$$

where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Question 1: Replace the 1?, 2? ,3? and 4? with the correct numbers.

x_1 and x_2 are separated in the expression for f(x),

$$f(x) = f(x_1) + f(x_2)$$

there is no x_1x_2 term, and so we can optimize $f(x_1)$ and $f(x_2)$ separately.

Question 2: The optimal solution to $\min f(x_1)$ is .3

Question 3: (continued from the previous question) and the optimal solution to $\min f(x_2)$ is .5

Question 4: (continued from the previous question) Therefore, the optimal solution to $\min f(x)$ is (x_1, x_2) where x_1 = .3

Question 5: (continued from the previous question) and x_2 = .5

If Q is not a diagonal matrix, for example,

$$Q = \begin{bmatrix} 10 & 1 \\ 1 & 4 \end{bmatrix} \quad b = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$$

separating x_1 and x_2 is not possible due to the x_1x_2 term in the expression for f(x). However, separation of variables becomes possible if x is expressed not as a linear combination of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $x = x_1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, but as a linear combination of specially selected vectors d_1 and d_2,

$x = \alpha_1 d_1 + \alpha_2 d_2$ where α_1 and α_2 are scalar variables.

Before we attempt to find specific d_1 and d_2 that allow for separating x_1 and x_2, let's investigate why x_1 and x_2 could be easily separated in the example with the diagonal matrix. x_1 and x_2 are the components of vector x in basis $\left\{ u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$.

u and v are conjugate with respect to $\begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix}$ (or any other diagonal matrix), that is,

$u^T \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix} v = 0$ and this results in a zero coefficient for the x_1x_2 term in $x^T \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix} x$:

$$x^T \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix} x = (x_1 u + x_2 v)^T \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix} (x_1 u + x_2 v)$$

$$= (x_1 u)^T \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix} (x_2 v) + (x_2 v)^T \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix} (x_1 u) + (x_1 u)^T \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix} (x_1 u) + (x_2 v)^T \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix} (x_2 v).$$

$$= (x_1 u)^T \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix} (x_1 u) + (x_2 v)^T \begin{bmatrix} 10 & 0 \\ 0 & 4 \end{bmatrix} (x_2 v)$$

By analogy, for d_1 and d_2 to allow for separating x_1 and x_2 when $Q=\begin{bmatrix}10 & 1\\ 1 & 4\end{bmatrix}$, d_1 and d_2 must be conjugate with respect to Q, $d_1^T Q d_2 = 0$.

Let's arbitrary choose d_1 to be $\begin{bmatrix}1\\ 0\end{bmatrix}$ and the first element of d_2 to be 1.

Question 6: (not for forum discussion) If d_1 and d_2 are conjugate with respect to $Q=\begin{bmatrix}10 & 1\\ 1 & 4\end{bmatrix}$ then the second element of d_2 is -10

d_1 and d_2 are linearly independent and so x can be expressed as a linear combination of them,

$$x = \alpha_1 d_1 + \alpha_2 d_2$$

Substituting x with $\alpha_1 d_1 + \alpha_2 d_2$ in $f(x) = \frac{1}{2}x^T Q x + b^T x$, we get

$$f(\alpha_1, \alpha_2) = 1?\alpha_1^2 + 2?\alpha_2^2 - 3?\alpha_1 + 4?\alpha_2$$

Question 7: Replace the 1?, 2? ,3? and 4? with the correct numbers. 5,195,3,17

Question 8: The optimal solution to $\min f(\alpha_1)$ is .3

Question 9: (continued from the previous question) and the optimal solution to $\min f(\alpha_2)$ is -.0436

Question 10: (continued from the previous question) which gives the optimal solution (x_1, x_2) to the original problem min f(x) with x_1 = .2564

Question 11: (continued from the previous question) and x_2 = .4359

Verifying the result in MATLAB:

$$Q=\begin{bmatrix}10 & 1\\ 1 & 4\end{bmatrix} \quad b=\begin{bmatrix}-3\\ -2\end{bmatrix}$$

```
Q = [10 1; 1 4];
b = [-3 -2];
quadprog(Q, b)
```

Note that we don't need to know d_2 in order to optimize the function in direction d_1, in other words, we don't need d_2 to find α_1. The conjugate gradient algorithm (described in the next section) takes advantage of this fact.

In general, a conjugate Gram–Schmidt process is used to find Q-conjugates $d_1, d_2, \ldots, d_n$

$$d_i^T Q d_j = 0, \quad for \;\; i \neq j, 1 \leq i, j \leq n$$

when Q is an n by n matrix. This process takes a set of n linearly independent vectors $\{y_1, y_2, \ldots, y_n\}$ as input and returns $d_1, d_2, \ldots, d_n$ according to the following equation:

$$d_{k+1} = -y_{k+1} + \sum_{j=1}^{k} \frac{y_{k+1}^T Q d_j}{d_j^T Q d_j} d_j$$

The $y_{k+1}^T Q d_j$ and $d_j^T Q d_j$ are scalars.

Question 12: (not for forum discussion) In the above example with $Q = \begin{bmatrix} 10 & 1 \\ 1 & 4 \end{bmatrix}$, if the set of linear independent vectors to be used by the conjugate Gram–Schmidt process is $\left\{ u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ then d_2 produced by this process (d_1 and the first element of d_2 are no longer arbitrary set) is $((d_2)_1, (d_2)_2)$ where $(d_2)_1 =$ .1

Question 13: (not for forum discussion, continued from the previous question) and $(d_2)_2 =$ -1

The Conjugate Gradient Algorithm

Video: 5B

Reading: Chong, An Introduction to Optimization, 2011, Chapter 10 – Sections 10.3-10.4

https://onlinelibrary-wiley-com.proxy1.library.jhu.edu/doi/book/10.1002/9781118033340

Rao, Engineering Optimization: Theory and Practice, 2009, Chapter 6 – Section 6.10

https://onlinelibrary-wiley-com.proxy1.library.jhu.edu/doi/book/10.1002/9780470549124

A conjugate Gram–Schmidt process can use any set of linear independent vectors $\{y_1, y_2, \ldots, y_n\}$ as an input, for example, the standard basis ({u, v} in the above example) can be used. However, one specific set of linear independent vectors is especially useful.

This set is based on the gradients of the objective function at consecutive iterations. Before proceeding, let's rename $\{d_1, d_2, \ldots, d_n\}$ into $\{d_0, d_1, \ldots, d_{n-1}\}$ and $\{y_1, y_2, \ldots, y_n\}$ into $\{y_0, y_1, \ldots, y_{n-1}\}$ because 1-based indexing is common for the conjugate Gram–Schmidt process, but zero-based indexing is used in the lectures and readings for the conjugate gradient algorithm.

We'll start with an arbitrary selected point $x^{(0)}$ and the first vector in the linear independent set,

y_0, is $\nabla f(x^{(0)})$. Applying the conjugate Gram–Schmidt process equation,

$$d_0 = -\nabla f(x^{(0)})$$

Next, we optimize the quadratic objective function in the direction of d_0 and get a new point

$$x^{(1)} = x^{(0)} + \gamma_0 d_0$$

γ_0 is determined by an exact line search (in practice, an inexact line search is more common, but we'll assume it's exact to show the logic of the method)

The gradient of the quadratic objective function at the new point, $\nabla f(x^{(1)})$ is orthogonal to $\nabla f(x^{(0)})$ because the objective function is already fully optimized in direction d_0 (with the exact line search), there can no longer be any gradient along d_0. Therefore, $\nabla f(x^{(0)})$ and $\nabla f(x^{(1)})$ are linearly independent and so $\nabla f(x^{(1)})$ qualifies as y_1,

$d_1 = -\nabla f(x^{(1)}) + \dfrac{\nabla f^T(x^{(1)}) Q d_0}{d_0^T Q d_0} d_0$. Then the objective function is optimized along d_1 and so on.

$\nabla f(x^{(k)})$ is orthogonal to all the previous directions because there can no longer be any gradient along the previous directions (Lemma 10.2 on page 174 in Chong). It can be shown by multiplying both sides of the conjugate Gram–Schmidt process equation for d_{k+1} by $\nabla f(x^{(k+1+m)})$ where m=1, 2, etc. that the gradient is also orthogonal to all the previous gradients, so the set of consecutive gradients qualifies as a set of linearly independent vectors for the purpose of the conjugate Gram–Schmidt process.

This orthogonality also allows one to simplify the conjugate Gram–Schmidt process equation

$$d_{k+1} = -\nabla f(x^{(k+1)}) + \sum_{j=0}^{k} \frac{\nabla f^T(x^{(k+1)}) Q d_j}{d_j^T Q d_j} d_j$$

(the summation is from 0 now due to the zero-based indexing)

For the first step of the simplification, note that

$d_j = \dfrac{1}{\gamma_j}\left(x^{(j+1)} - x^{(j)}\right)$ and

$$Qd_j = \frac{1}{\gamma_j} Q\left(x^{(j+1)} - x^{(j)}\right). \qquad (1)$$

Also, since we're solving a quadratic optimization problem $f(x) = \frac{1}{2} x^T Qx + b^T x$, the gradient of f(x) is $\nabla f(x) = 1?\, x + 2?$

Question 14: (not for forum discussion) Replace the 1? with the correct expression

- $\frac{1}{2}Q$
- Q
- $\frac{1}{2}Q^T$
- Q^T
- none of the above

Question 15: (not for forum discussion) Replace the 2? with the correct expression

- $\frac{1}{2}b$
- b
- $\frac{1}{2}b^T$
- b^T
- none of the above

Therefore, $\nabla f(x^{(j+1)}) - \nabla f(x^{(j)}) = Q\left(x^{(j+1)} - x^{(j)}\right) \qquad (2)$

Combining (1) and (2) gives

$$Qd_j = \frac{1}{\gamma_j}\left(\nabla f(x^{(j+1)}) - \nabla f(x^{(j)})\right)$$

and the conjugate Gram–Schmidt process equation becomes

$$d_{k+1} = -\nabla f(x^{(k+1)}) + \sum_{j=0}^{k} \frac{\nabla f^T(x^{(k+1)})\left(\nabla f(x^{(j+1)}) - \nabla f(x^{(j)})\right)}{d_j^T\left(\nabla f(x^{(j+1)}) - \nabla f(x^{(j)})\right)} d_j$$

$\nabla f(x^{(k+1)})$ is orthogonal to $\nabla f(x^{(0)}), \nabla f(x^{(1)}), \ldots, \nabla f(x^{(k)})$, so only the last term of the above sum is not equal zero,

$$d_{k+1} = -\nabla f(x^{(k+1)}) + \frac{\nabla f^T(x^{(k+1)})\left(\nabla f(x^{(k+1)}) - \nabla f(x^{(k)})\right)}{d_k^T\left(\nabla f(x^{(k+1)}) - \nabla f(x^{(k)})\right)} d_k \qquad (3)$$

The same equation for d_k,

$$d_k = -\nabla f(x^{(k)}) + \frac{\nabla f^T(x^{(k)})\left(\nabla f(x^{(k)}) - \nabla f(x^{(k-1)})\right)}{d_{k-1}^T\left(\nabla f(x^{(k)}) - \nabla f(x^{(k-1)})\right)} d_{k-1} \qquad (4)$$

Multiplying the transpose of both sides of (4) by $\nabla f(x^{(k)})$, we get

$$d_k^T \nabla f(x^{(k)}) = -\nabla f^T(x^{(k)}) \nabla f(x^{(k)}) = -\left\|\nabla f(x^{(k)})\right\|^2 \qquad (5)$$

because $d_{k-1}{}^T \nabla f(x^{(k)}) = 0$

Combining (3) and (5) and noticing that $d_k{}^T \nabla f(x^{(k+1)}) = 0$ gives

$$d_{k+1} = -\nabla f(x^{(k+1)}) + \frac{\nabla f^T(x^{(k+1)})\left(\nabla f(x^{(k+1)}) - \nabla f(x^{(k)})\right)}{\left\|\nabla f(x^{(k)})\right\|^2} d_k \qquad (6)$$

or

$$d_{k+1} = -\nabla f(x^{(k+1)}) + \frac{\left\|\nabla f(x^{(k+1)})\right\|^2}{\left\|\nabla f(x^{(k)})\right\|^2} d_k \qquad (7)$$

Equation (6) is used in the Polak-Ribiere Conjugate Gradient Method (PRCG) method and equation (7) is employed by the Fletcher-Reeves Conjugate Gradient Method (FRCG).

Equations (6) and (7) are equivalent when f(x) is a quadratic function and an exact line search is used. However, in practice, these equations are applied to non-quadratic optimization problems and exact line search is also replaced by inexact one. Therefore, the algorithm may take more than n steps to converge and this number of steps may be different for the two methods.

Question 16: Perform two FRCG iterations to solve $\min f(x_1, x_2) = \frac{1000}{x_1 + x_2} + (x_1 - 4)^2 + (x_2 - 10)^2$

Do not round any intermediate results. Use $x^{(0)} = (3, 1)$ as the starting point and backtracking line search with Armijo's rule and $\tilde{\alpha} = 2, c = 10^{-4}, \rho = 0.5$. $x^{(1)} = (x_1^{(1)}, x_2^{(1)})$ where $x_1^{(1)} =$ 11.0625

Question 17: (continued from the previous question) $x_2^{(1)} =$ 11.0625

Question 18: (not for forum discussion, continued from the previous question) $x^{(2)} = (x_1^{(2)}, x_2^{(2)})$ where $x_1^{(2)} =$ 5.4639

Question 19: (not for forum discussion, continued from the previous question) $x_2^{(2)} =$ 11.5736

Optimization with Constraints

Video: 5C-D

Consider the following optimization problem with an equality constraint,

$$\max f(x_1, x_2) = x_1 x_2^{\,2}$$

s.t.

$$2x_1^2 + 3x_2^2 = 1$$

Question 20: Solve this problem using the substitution method, an optimal solution is (x_1, x_2) with $x_1 =$ .4082

Question 21: (continued from the previous question) $x_2 = \pm$

Lagrange Multipliers

Video: 5E

Reading: Rao, Engineering Optimization: Theory and Practice, 2009– Section 2.4

https://onlinelibrary-wiley-com.proxy1.library.jhu.edu/doi/book/10.1002/9780470549124

Let's use a different method to solve the problem in the previous section and illustrate this method with a graph. For the plot to be more intuitive, we'll denote the decision variables by x and y instead of x_1 and x_2,

$$\max f(x, y) = xy^2$$

s.t.

$$2x^2 + 3y^2 = 1$$

Question 22: Plot the constraint and three contours of the objective function, $xy^2=0.1$, $xy^2=0.3$, $xy^2=0.5$, in MATLAB. Please upload a screenshot of your plot here.

Next, let's plot the contour $xy^2=a$ that touches the constraint. At any point where the constraint and the contour touch, the constraint and the contour share the tangent line. At such a point, the gradient of $f(x,y)=xy^2$ is perpendicular to this shared tangent line and the gradient of $h(x,y)=2x^2+3y^2-1$ is also perpendicular to the shared tangent line. Therefore, these two gradients are colinear, one of them is equal to a constant multiplied by the other gradient,

$$\nabla f(x,y)=\lambda\nabla h(x,y)$$

The constant λ is called a Lagrange multiplier.

Applying the above equation to our constraint and the objective function,

$$\begin{bmatrix} y^2 \\ 2xy \end{bmatrix}=\lambda\begin{bmatrix} 4x \\ 6y \end{bmatrix}$$

There are three variables (λ is unknown) and two equations, but, if we add the constraint $2x^2+3y^2-1=0$, there'll be three equations. Solving them gives two candidates for the optimal solution.

Question 23: λ =? .1361

Question 24: An optimal solution is (x, y) with x = .4082

Question 25: (continued from the previous question) y = ± .4714

Question 26: What is the value of the objective function at an optimal solution? .0907

We can write $\nabla f(x,y)=\lambda\nabla h(x,y)$ and $h(x,y)=2x^2+3y^2-1=0$ all together as one equation by defining the *Lagrangian*,

$$L(x,y,\lambda)=f(x,y)-\lambda h(x,y)$$

The gradient of the Lagrangian is

$$\begin{bmatrix} \frac{\partial L}{\partial x} \\ \frac{\partial L}{\partial y} \\ \frac{\partial L}{\partial \lambda} \end{bmatrix}=\begin{bmatrix} \frac{\partial f}{\partial x}-1?\frac{\partial h}{\partial x} \\ \frac{\partial f}{\partial y}-2?\frac{\partial h}{\partial y} \\ 3? \end{bmatrix}$$

Questions 27: Replace the 1? above with the correct expression

- λ
- $\frac{\partial \lambda}{\partial x}$
- $\frac{\partial \lambda}{\partial y}$
- $\frac{\partial \lambda^2}{\partial x \partial y}$
- none of the above

Questions 28: Replace the 2? above with the correct expression

- λ
- $\frac{\partial \lambda}{\partial x}$
- $\frac{\partial \lambda}{\partial y}$
- $\frac{\partial \lambda^2}{\partial x \partial y}$
- none of the above

Questions 29: Replace the 3? above with the correct expression

- h(x, y)
- -h(x,y)
- $\frac{\partial h}{\partial \lambda}$
- $-\frac{\partial h}{\partial \lambda}$
- none of the above

Therefore, $\begin{cases} \nabla f(x,y) = \lambda \nabla h(x,y) \\ h(x,y) = 0 \end{cases}$ is equivalent to $\nabla L(x,y,\lambda) = 0$.

Aside from providing for a compact form of the optimality condition, Lagrange multipliers have another interesting application. Let's change the constraint of the original optimization problem

$\max f(x,y) = xy^2$

s.t.

$2x^2 + 3y^2 - 1 = 0$

into

$2x^2 + 3y^2 - 1 = a$

It can be shown that, if a is small, the corresponding change in the value of the objective function at the optimal point is approximately λa (this change approaches λa as a approaches 0).

Question 30: Use the above property of the Lagrange multiplier to find the approximate value of the objective function at an optimal solution to

$\max f(x, y) = xy^2$

s.t.

$2x^2 + 3y^2 = 1.01$ .0921

Question 31: Use the substitution method to find the exact value of the objective function at an optimal solution to the problem in the previous question. .0921

The Lagrangian for an Equality Constrained Problem

Chong, An Introduction to Optimization, 2011, Sections 19.1-19.4

https://onlinelibrary-wiley-com.proxy1.library.jhu.edu/doi/book/10.1002/9781118033340

In the case of more than two decision variables (x is a vector now), we can no longer easily plot the constraint and the contour, but their gradients are still colinear and the Lagrangian

$L(x, \lambda) = f(x) - \lambda h(x)$

can again be used to combine all the optimality conditions in one equation, $\nabla L = 0$.

The minus sign in front of λ can be replaced by the plus sign because the constraint $h(x) = 0$ can be written as $-h(x) = 0$ or λ can be redefined to be minus the ratio of the gradients,

$-\dfrac{\nabla f(x)}{\nabla h(x)}$,

$L(x, \lambda) = f(x) + \lambda h(x)$

The Lagrangian is often written with a "–" in other applications, but the version with a "+" is more convenient in the context of the theory presented below and in further modules, so we will use a "+". For q equality constraints instead of one, we have

$$L(x, \lambda_1, \lambda_2, \ldots, \lambda_q) = f(x) + \sum_{i=1}^{q} \lambda_i h_i(x)$$

The Lagrangian for an Inequality Constrained Problem

Video: 5F

Reading: Chong, An Introduction to Optimization, 2011, Section 20.1

https://onlinelibrary-wiley-com.proxy1.library.jhu.edu/doi/book/10.1002/9781118033340

Rao, Engineering Optimization: Theory and Practice, 2009, Section 2.5

https://onlinelibrary-wiley-com.proxy1.library.jhu.edu/doi/book/10.1002/9780470549124

The inequality constrained problem

$\min f(x)$

s.t.

$g(x) \leq 0$

is equivalent to the equality constrained problem

$\min f(x)$

s.t.

$g(x) + s^2 = 0$

The Lagrangian for the equality constrained problem is

$L(x, \lambda) = f(x) + \lambda\left(g(x) + s^2\right)$

and the necessary optimality condition is $\nabla L(x, s, \lambda) = 0$.

The components of $\nabla L(x, s, \lambda)$ are

$$\begin{bmatrix} \dfrac{\partial L}{\partial x_i} \\ \dfrac{\partial L}{\partial s} \\ \dfrac{\partial L}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f}{\partial x_i} + \lambda \dfrac{\partial g}{\partial x_i} \\ 1? \\ 2? \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Questions 32: (not for forum discussion) Replace the 1? above with the correct expression
- λ
- $2s$
- $2\lambda s$

- $g(x)$
- $g(x)+2s$
- $g(x)+s^2$
- none of the above

Questions 33: (not for forum discussion) Replace the 2? above with the correct expression

- λ
- $2s$
- $2\lambda s$
- $g(x)$
- $g(x)+2s$
- $g(x)+s^2$
- none of the above

Combining $1?=0$ and $2?=0$ where 1? and 2? are the answers to the previous two questions, we get the *complementary slackness* rule for the optimal solution,

$$\lambda g(x)=0$$

Therefore, if the Lagrangian for an inequality constrained minimization problem is defined as

$$L(x,\mu)=f(x)+\mu g(x)$$

(λ is renamed to μ and the $+\mu s^2$ part is omitted), $\nabla L(x,s,\mu)=0$ will no longer hold, but, instead, we have

$$\frac{\partial L}{\partial x_i}=\frac{\partial f}{\partial x_i}+\mu\frac{\partial g}{\partial x_i}=0$$

and the complementary slackness rule is

$$\mu g(x)=0$$

By looking at the directions of the gradients, it is also possible to conclude that $\mu\geq 0$. In fact, at the optimal solution, if the constraint is not active then $\mu=0$ due to complementary slackness rule. If the constraint is active then f(x) and g(x) gradients should have the ? directions, otherwise we would be able to make a change to x that simultaneously decreases f(x) (which improves the objective function value) and decreases g(x) (which ensures that the solution stays in the feasible region). Therefore, $\mu=-\frac{\nabla f(x)}{\nabla h(x)}$ at the optimal point of the corresponding equality constrained problem is non-negative.

Question 34: (not for forum discussion) Please replace the question mark with the correct word

- same
- opposite
- none of the above

μ is called a *Karush-Kuhn-Tucker (KKT) multiplier* and, for m inequality constraints, we have

$$L(x, \mu_1, \mu_2, ..., \mu_m) = f(x) + \sum_{i=1}^{m} \mu_i g_i(x)$$