ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΕΡΕΥΝΑ

ΚΕΦ. 2

ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ

Βλ. - Hillier-Lieberman, Introduction to Operations Research, 6th edition, Κεφάλαιο 13

-Ε.Φ. Μαγείρου, Σημειώσεις Μαθηματικού Προγραμματισμού

Ε.Φ. Μαγείρου

Επιμέλεια σημειώσεων: Δ. Κ. Βασιλάκης

ΕΛΛΑΔΑ 2008
Ανάπτυξη παντού. Ανάπτυξη για όλους.

ΥΠΟΥΡΓΕΙΟ ΕΘΝΙΚΗΣ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΕΠΕΑΕΚ
ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ
ΣΥΓΧΡΗΜΑΤΟΔΟΤΗΣΗ
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
ΕΥΡΩΠΑΪΚΟ ΤΑΜΕΙΟ ΠΕΡΙΦΕΡΕΙΑΚΗΣ ΑΝΑΠΤΥΞΗΣ

ΠΑΙΔΕΙΑ ΜΠΡΟΣΤΑ
2ο Επιχειρησιακό Πρόγραμμα
Εκπαίδευσης και Αρχικής
Επαγγελματικής Κατάρτισης

2.1 Βελτιστοποίηση χωρίς περιορισμούς (unconstrained optimization)

Έστω συνάρτηση N μεταβλητών $f:\mathbb{R}^N \to \mathbb{R}$

2.1.1 Βασικό Θεώρημα

$$f(x_1+\Delta x_1,\ldots,x_N+\Delta x_N) = f(x_1,x_2,\ldots,x_N) + \sum_{i=1}^{N}\frac{\partial f}{\partial x_i}(x_1,x_2,\ldots,x_N)\cdot\Delta x_i + $$
$$+\frac{1}{2}\cdot\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial^2 f}{\partial x_i \partial x_j}(x_1^*,x_2^*,\ldots,x_N^*)\cdot\Delta x_i \Delta x_j .$$

όπου $x_i \le x_i^* \le x_i + \Delta x_i,\ i=1,\ldots,N$

Για μικρά Δx_i εξετάζουμε <u>μόνο</u> το γραμμικό όρο.

Εναλλακτικά:

$$\Delta f = f(x_1+\Delta x_1,\ldots,x_N+\Delta x_N) - f(x_1,\ldots,x_N) \cong \sum_{i=1}^{N}\frac{\partial f}{\partial x_i}(x_1,\ldots,x_N)\cdot\Delta x_i$$

<u>Παράδειγμα 1:</u>

$$f(x_1,x_2) = e^{x_1+x_2^2}$$

$$\frac{\partial f}{\partial x_1} = e^{x_1+x_2^2} \qquad \frac{\partial f}{\partial x_2} = 2x_2\, e^{x_1+x_2^2}$$

Με ακριβείς υπολογισμούς έχουμε:

$$f(0,0) = e^0 = 1$$

$$f(0.1,-0.1) = e^{0.1+0.01} = 1.1163$$

ενώ προσεγγιστικά:

$$\frac{\partial f}{\partial x_1}(0,0) = 1 \quad \frac{\partial f}{\partial x_2}(0,0) = 0$$

Άρα:

$$f(0.1,-0.1) \cong f(0,0) + 1\cdot 0.1 + 0\cdot(-0.1) = 1.1$$

Παρατηρούμε ότι το σφάλμα είναι ≈ 0.01, δηλαδή $\sim \Delta x^2$

$\square$

Πώς αποδεικνύεται; (π.χ. για 2 μεταβλητές)

$$\begin{aligned}\Delta f &= f(x_1+\Delta x_1, x_2+\Delta x_2) - f(x_1, x_2) = \\ &= f(x_1+\Delta x_1, x_2+\Delta x_2) - f(x_1+\Delta x_1, x_2) + f(x_1+\Delta x_1, x_2) - f(x_1, x_2) \\ &\cong \frac{\partial f}{\partial x_2}(x_1+\Delta x_1, x_2)\cdot \Delta x_2 + \frac{\partial f}{\partial x_1}(x_1, x_2)\cdot \Delta x_1\end{aligned}$$

Αν η $\frac{\partial f}{\partial x_2}$ είναι συνεχής, τότε

$$\frac{\partial f}{\partial x_2}(x_1+\Delta x_1, x_2) \cong \frac{\partial f}{\partial x_2}(x_1, x_2) + \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1^*, x_2)\cdot \Delta x_1$$

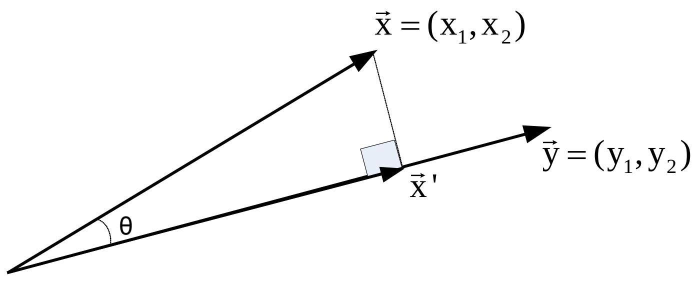
οπότε έχουμε:

$$\Delta f = \frac{\partial f}{\partial x_1}(x_1, x_2)\cdot \Delta x_1 + \frac{\partial f}{\partial x_2}(x_1, x_2)\cdot \Delta x_2 + \underbrace{\frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1^*, x_2)\cdot \Delta x_1 \Delta x_2}_{\text{ΑΜΕΛΗΤΕΟ!}}$$

$\square$

2.1.2 Ερμηνεία θεωρήματος

Έστω ότι έχουμε δύο διανύσματα $\vec{x} = (x_1, x_2)$ και $\vec{y} = (y_1, y_2)$ που σχηματίζουν μεταξύ τους γωνία θ.



Εσωτερικό γινόμενο διανυσμάτων:

$$\vec{x}\cdot\vec{y} = |\vec{x}|\cdot|\vec{y}|\cdot \sigma\upsilon\nu\theta \overset{\text{θεώρημα}}{=} x_1 y_1 + x_2 y_2$$

Γενικά: αν $\vec{x}, \vec{y} \in \mathbb{R}^N$, $\vec{x}\cdot\vec{y} = \sum_{i=1}^{N} x_i \cdot y_i$

Στο MS Excel το εσωτερικό γινόμενο διανυσμάτων υλοποιείται με την συνάρτηση SUMPRODUCT. Π.χ. η SUMPRODUCT(A1:A5;B1:B5) υπολογίζει το A1*B1+A2*B2+...+A5*B5.

<u>Καθετότητα διανυσμάτων:</u>

$$\vec{x} \perp \vec{y} \Leftrightarrow \ \vec{x} \cdot \vec{y} = 0 \Leftrightarrow \sum_{i=1}^{N} x_i \cdot y_i = 0$$

<u>Βαθμίδα ή κλίση (Gradient):</u>

Ορίζεται ως το διάνυσμα:

$$\nabla f(x_1,\ldots,x_N) = \left(\frac{\partial f}{\partial x_1}(x_1,\ldots,x_N), \frac{\partial f}{\partial x_2}(x_1,\ldots,x_N), \ldots, \frac{\partial f}{\partial x_N}(x_1,\ldots,x_N) \right)$$

Π.χ. αν $f(x_1, x_2) = e^{x_1+x_2^2}$ τότε $\nabla f(0,0) = (1,0)$ (γιατί;)

Με τί ισούται το $\nabla f(-1,1)$;

<u>Έκφραση βασικού θεωρήματος προσέγγισης:</u>

$$\boxed{\Delta f \cong \sum_{i=1}^{N} \left\{ \frac{\partial f}{\partial x_i}(x_1,\ldots,x_N) \cdot \Delta x_i \right\} = \nabla f \cdot \vec{\Delta} x}$$

<u>Παρατήρηση:</u> αν $\nabla f(x_1,\ldots,x_N) \neq 0$, τότε το σημείο $(x_1,\ldots,x_N)$ δεν είναι βέλτιστο, ΔΙΟΤΙ αν εξετάσουμε μια οικογένεια μεταβολών $\Delta x(\varepsilon) = \left(\varepsilon \frac{\partial f}{\partial x_1}, \varepsilon \frac{\partial f}{\partial x_2}, \ldots \right) = \varepsilon \cdot \nabla f$ για $\varepsilon \in \mathbb{R}$, τότε $\Delta f \cong \varepsilon \cdot \nabla f \cdot \nabla f = \varepsilon \cdot \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2$. Άρα για $\varepsilon > 0$ μικρό (αν π.χ. θέλουμε να βρούμε μέγιστο) έχουμε $\Delta f > 0$ και άρα μπορούμε να αυξήσουμε την τιμή της συνάρτησης.

<u>Άρα</u> αναγκαίο για την ύπαρξη βέλτιστου στο σημείο $\vec{x}^* = (x_1^*, \ldots, x_N^*)$: $\Delta f(\vec{x}^*) = 0$

2.1.3 Διαδικασίες αναζήτησης

Αλγόριθμοι που όταν $\Delta f \neq 0$ βελτιώνουν την τιμή της f έως ότου βρεθεί ακρότατο (μέγιστο ή ελάχιστο).

<u>Αναζήτηση κατά διεύθυνση</u>

Έστω κάποια διεύθυνση $\vec{d} = (d_1, d_2, \ldots, d_N)$ και αρχικό σημείο $\vec{x}$. Πώς βρίσκουμε το μέγιστο της f πάνω στα σημεία $\vec{x} + t\vec{d}$ για αριθμό $t \in \mathbb{R}$; ($\vec{x}, \vec{d}$ σταθερά)

Παρατηρήστε ότι:

$$f(\vec{x} + (t + \Delta t)\vec{d}) - f(\vec{x} + t\vec{d}) \cong \nabla f(\vec{x} + t\vec{d}) \cdot \Delta t \cdot \vec{d} \quad \text{(γιατί;)}$$

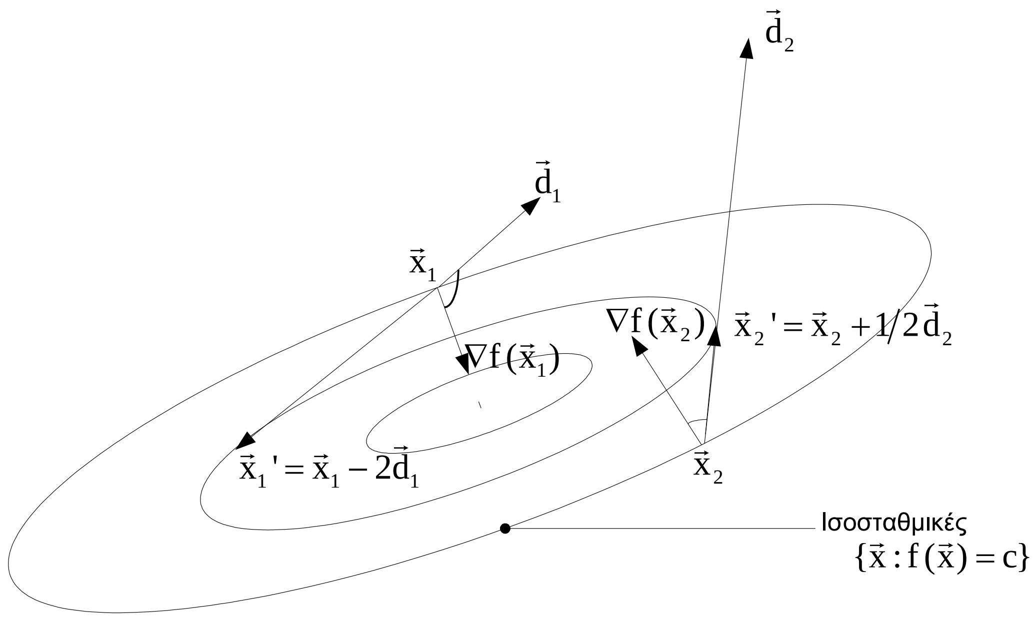
άρα $\dfrac{\Delta f}{\Delta t} \cong \nabla f(\vec{x} + t\vec{d}) \cdot \vec{d}$ (Δt: αριθμός)

και έτσι $\lim\limits_{\Delta t \to 0} \dfrac{\Delta f}{\Delta t}(\vec{x} + t\vec{d}) = \nabla f(\vec{x} + t\vec{d}) \cdot \vec{d} = \dfrac{df}{dt}$

Άρα, αν $\nabla f \cdot \vec{d} \neq 0$, μπορούμε να αυξήσουμε την τιμή της f θέτοντας:

- $\Delta t > 0$ αν $\nabla f \cdot \vec{d} > 0$
- $\Delta t < 0$ αν $\nabla f \cdot \vec{d} < 0$

<u>Γεωμετρικό παράδειγμα:</u> (μεγιστοποίηση με αναζήτηση στη διεύθυνση $\vec{d}$)



Στο σημείο $\vec{x}_1$ το γινόμενο $\nabla f(\vec{x}_1) \cdot \vec{d}_1$ είναι αρνητικό (γιατί;). Επομένως το μέγιστο πάνω στη διεύθυνση $\vec{d}_1$ είναι στο σημείο $\vec{x}_1' = \vec{x}_1 - 2\vec{d}_1$, δηλαδή $t < 0$.

Στο σημείο $\vec{x}_2$ το γινόμενο $\nabla f(\vec{x}_2)\cdot\vec{d}_2$ είναι θετικό (γιατί;). Επομένως το μέγιστο πάνω στη διεύθυνση $\vec{d}_2$ είναι στο σημείο $\vec{x}_2{}' = \vec{x}_2 + 1/2\vec{d}_2$, δηλαδή $t > 0$.

□

Αριθμητικό παράδειγμα:

Έστω η συνάρτηση $f(x,y) = x^2 + y^2 - xy$ και θέλουμε να την ελαχιστοποιήσουμε, οπότε μεγιστοποιούμε την αρνητική της $g(x,y) = -\left[x^2 + y^2 - xy\right]$.

Έστω ότι αρχίζουμε από το σημείο $(x,y) = (3,5)$ και εξετάζουμε την διεύθυνση $(1,0)$. Δηλαδή τα σημεία $(3+t,5)$, οπότε αν κινηθούμε κατά t πάνω σε αυτή τη διεύθυνση έχουμε:

$$g(3+t,5) = -\left[(3+t)^2 + 5^2 - 5(3+t)\right] = g(t)$$

$$\frac{dg}{dt} = -\left[2(3+t) - 5\right] = 0 \Rightarrow t = -1/2$$

κι έτσι το νέο (x,y) είναι $(2.5,5)$ και το t είναι αρνητικό όπως και αναμέναμε (γιατί;).

Εξετάζουμε τώρα τη διεύθυνση $(0,1)$, δηλαδή σημεία $(2.5,5+t)$, οπότε:

$$g(t) = -\left[2.5^2 + (5+t)^2 - 2.5(5+t)\right]$$

$$\frac{dg}{dt} = -\left[2(5+t) - 2.5\right] = 0 \Rightarrow t = -7.5/2$$

με νέο $(x,y) = (2.5,1.25)$.

Η αρχική τιμή της f είναι $f(3,5) = 19$, ενώ τελικά έχει μειωθεί σε $f(2.5,1.25) = 4.69$. Το βέλτιστο προκύπτει τελικά για $x = y = 0$ μετά από παρόμοια βήματα.

□

Άσκηση:

Αρχίζοντας από το σημείο (x_n, y_n) και μεγιστοποιώντας ως προς την διεύθυνση $(1,0)$, αποδείξτε για την ίδια συνάρτηση ότι το επόμενο σημείο είναι το $(y_n/2, y_n)$. Τι ισχύει αν μεγιστοποιήσουμε ως προς $(0,1)$; Με βάση αυτά αναλύστε τον σχετικό αλγόριθμο με αναζητήσεις κατά $(1,0)$, $(0,1)$, $(1,0)$, $(0,1)$, κ.ο.κ.

2.1.3.1 Αλγόριθμος Αναζήτησης Συντεταγμένων

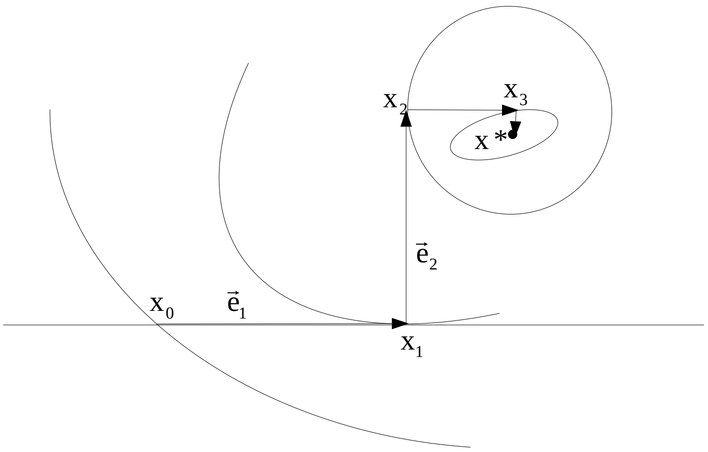
Στο προηγούμενο αριθμητικό παράδειγμα μεγιστοποιώντας διαδοχικά πάνω στις διευθύνσεις $(1,0)$ και $(0,1)$ εφαρμόσαμε ουσιαστικά τον <u>Αλγόριθμο Αναζήτησης Συντεταγμένων</u>:

- Έστω $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_N$ τα μοναδιαία διανύσματα.
- Αρχίζοντας από αυθαίρετο x_0 [1] μεγιστοποιούμε κατά την διεύθυνση $\vec{e}_1$ και παίρνουμε το x_1.
- Αρχίζοντας από το x_1 μεγιστοποιούμε κατά την διεύθυνση $\vec{e}_2$ και παίρνουμε το x_2 κ.ο.κ.
- Συνεχίζουμε εναλλάσσοντας κυκλικά τις διευθύνσεις, δηλαδή $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_N, \vec{e}_1, \ldots, \vec{e}_N, \vec{e}_1, \ldots$
- Σταματάμε όταν <u>δεν</u> πάρουμε βελτίωση N φορές, δηλαδή αν: $x_k = x_{k+1} = \ldots = x_{k+N} = x^*$.
- Το βέλτιστο είναι στο x^*.

<u>Άσκηση:</u>
Αποδείξτε ότι το κριτήριο τερματισμού είναι ορθό.

<u>Σχόλια:</u>

- Αν και ο Αλγόριθμος Αναζήτησης Συντεταγμένων είναι ορθός δεν είναι απαραίτητο να τερματίσει! Δώστε ένα παράδειγμα αναφερόμενοι στην άσκηση της ενότητας 2.1.3.
- Μία γεωμετρική εικόνα της συμπεριφοράς του αλγορίθμου δίνεται παρακάτω:



2.1.3.2 Αλγόριθμος Αναζήτησης Βαθμίδας

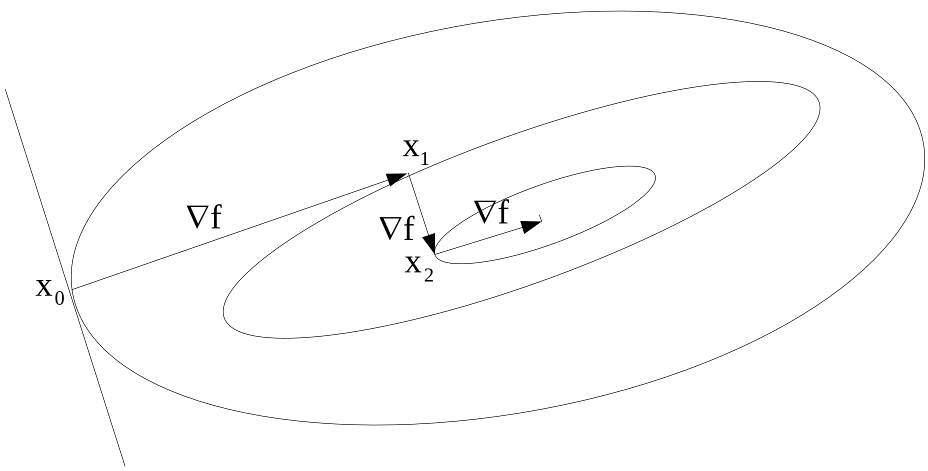
Ένας πολύ σημαντικός αλγόριθμος είναι ο Αλγόριθμος Αναζήτησης βαθμίδας:

- Διαλέγουμε αυθαίρετο σημείο x_0 και θέτουμε δείκτη $k=1$.
- Αναθεωρούμε:
  - Υπολογίζουμε τη διεύθυνση $\vec{d}=\nabla f(x_{k-1})$.
  - Μεγιστοποιούμε κατά την $\vec{d}$ από το σημείο x_{k-1}, υπολογίζοντας[1] έτσι το νέο σημείο x_k.
  - Εάν το $\nabla f(x_k)$ είναι «μικρό» (π.χ. αν $\sum_{i=1}^{N}\left[\frac{\partial f}{\partial x_i}(x_k)\right]^2<\varepsilon$ για $\varepsilon=10^{-3}$), τότε ο αλγόριθμος τερματίζει.
  - Διαφορετικά, θέτουμε $k\leftarrow k+1$ και επαναλαμβάνουμε την αναθεώρηση.

Σχόλια:

- Μειονέκτημα: απαιτεί υπολογισμό των παραγώγων κάθε φορά
- Πλεονέκτημα: χρειάζεται λιγότερα βήματα από ότι ο Αλγόριθμος Αναζήτησης Συντεταγμένων.

Γεωμετρική Συμπεριφορά:



Παρατηρούμε ότι οι διευθύνσεις είναι και πάλι κάθετες μεταξύ των σε 2 διαστάσεις (γιατί;). Ο αλγόριθμος αυτός εμφανίζει αρκετά συχνά παθολογική συμπεριφορά, δηλαδή πολλά αναποτελεσματικά βήματα.

Αναλυτικό παράδειγμα Αναζήτησης Βαθμίδας:

Έστω το ακόλουθο πρόβλημα βελτιστοποίησης:

$$\min f(x,y) = 1/2 \cdot (x^2 + 4y^2 - 2xy) = 1/2 \cdot [x, y] \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Υπολογίζουμε πρώτα:

$$\nabla f(x,y) = \begin{bmatrix} x - y \\ -x + 4y \end{bmatrix} = \left(= \begin{bmatrix} \partial f/\partial x \\ \partial f/\partial y \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \right)$$

Γνωρίζουμε εκ των προτέρων ότι το ελάχιστο της f βρίσκεται στο $x = y = 0$, καθώς $\nabla f(0,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, ενώ ο πίνακας $\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$ είναι θετικά ορισμένος.

1. Αρχίζουμε την αναζήτηση στον αλγόριθμο ξεκινώντας από το σημείο έστω $(x_0, y_0) = (1,1)$ όπου $\nabla f(1,1) = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$.

Εξετάζουμε το $\min\limits_t f(1, 1+3t)$ ή $\min\limits_t \left\{ 1/2 \left(1 + 4(1+3t)^2 - 2(1+3t) \right) \right\}$, οπότε $t^* = -1/4$ (γιατί;).

2. Επομένως, $(x_1, y_1) = (x_0, y_0 + 3t^*) = (1, 1 - 3/4) = (1, 1/4)$. Δηλαδή, οδηγούμαστε σε ένα νέο σημείο $(1, 1/4)$ όπου $\nabla f(1, 1/4) = (3/4, 0)$.

Παρατηρήστε ότι $\nabla f(x_0, y_0) \perp \nabla f(x_1, y_1)$!

3. Εξετάζουμε πλέον το $f(1 + 3/4t,\ 1/4)$ που έχει ελάχιστο για $t^* = -1$ (γιατί;), άρα $(x_2, y_2) = (1/4, 1/4)$ και $\nabla f(1/4, 1/4) = (0,\ 3/4)$ (δηλαδή $\nabla f(x_0, y_0) \parallel \nabla f(x_2, y_2)$!)

Ο παρακάτω πίνακας δείχνει πως εξελίσσεται ο αλγόριθμος.

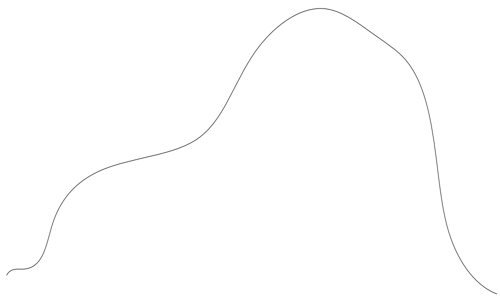
Αριθμός επανάληψης	t^*	x	y	f	$\partial f/\partial x$	$\partial f/\partial y$
0	t	1	1	3	0	3
1	-0.25	1.000	0.250	0.75000	0.750	0.000
2	-1	0.250	0.250	0.18750	0.000	0.750
3	-0.25	0.250	0.062	0.04687	0.187	0.000
4	-1	0.062	0.062	0.01172	0.000	0.187
5	-0.25	0.062	0.016	0.00293	0.047	0.000
6	-1	0.016	0.016	0.00073	0.000	0.047
7	-0.25	0.016	0.004	0.00018	0.012	0.000
8	-1	0.004	0.004	0.00005	0.000	0.012

$\square$

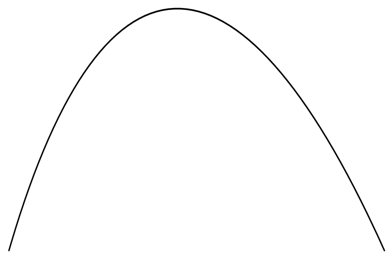
2.1.3.3 Βελτιστοποίηση συνάρτησης μίας μεταβλητής

Βασικό στοιχείο σε οποιαδήποτε αναζήτηση είναι η μεγιστοποίηση μίας συνάρτησης μίας μεταβλητής με σκοπό τον υπολογισμό του t^* κάθε φορά. Πώς μπορούμε να κάνουμε τη μεγιστοποίηση χωρίς υπολογισμό παραγώγων;

Έστω μία <u>μονόμορφη</u> (unimodal) συνάρτηση, δηλαδή μία συνάρτηση με ένα μοναδικό (τοπικό ή ολικό) μέγιστο ή ελάχιστο. Οι συναρτήσεις αυτές είναι γενικότερες από τις κοίλες συναρτήσεις και συναντώνται συχνά στην πράξη.



Μονόμορφη

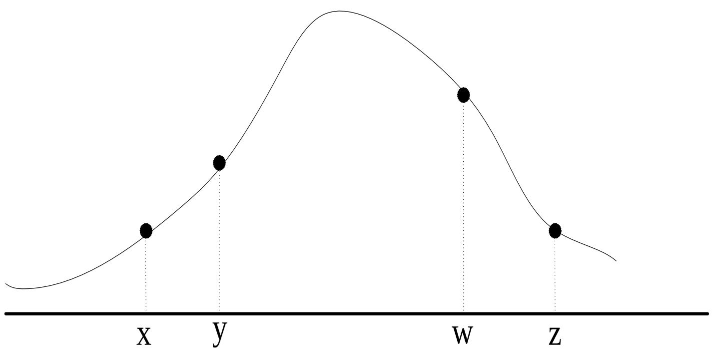


Κοίλη

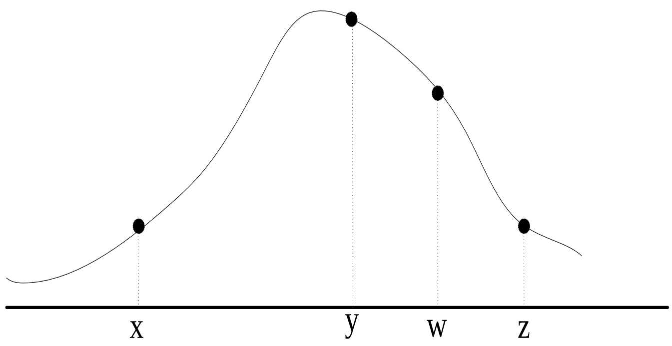
Σε τέτοιες μονόμορφες συναρτήσεις η αναζήτηση είναι πολύ αποτελεσματική και βασίζεται στην εξής παρατήρηση:
Αν $x \le y \le z$ και $f(x), f(z) \le f(y)$, το μέγιστο της f είναι μεταξύ x και z. (αποδείξτε το!)

Έστω τώρα ότι υπολογίζουμε την f στο σημείο w με $x < y < w < z$. Ισχύει ότι $f(w) > f(z)$ (γιατί;).

- Αν ισχύει ότι $f(w) > f(y)$, τότε το μέγιστο βρίσκεται στο διάστημα $[y, z]$.
- Αν ισχύει ότι $f(w) < f(y)$, τότε το μέγιστο βρίσκεται στο διάστημα $[x, w]$.
-



$f(w) > f(y) \Rightarrow \max \in [y, z]$



$f(w) < f(y) \Rightarrow \max \in [x, w]$

Αναζήτηση χρυσής τομής

Έστω ότι έχουμε διαπιστώσει ότι το $\max$ είναι στο $[0,1]$, χωρίς απώλεια γενικότητας.

- Σχεδιάζουμε μετρήσεις στα διάφορα x, y με $0 < x < y < 1$.
- Ανάλογα με τα αποτελέσματα το $\max$ θα βρίσκεται είτε στο $[x,1]$ είτε στο $[0, y]$.
- Το $[x,1]$ περιλαμβάνει το y όπου έχει ήδη γίνει μέτρηση της f.
- Το $[0, y]$ περιλαμβάνει το x όπου έχει ήδη γίνει μέτρηση της f.
- Καλό θα είναι:
  - Τα διαστήματα αβεβαιότητας που θα προκύψουν να είναι ίσα.
  - Να μπορούμε να χρησιμοποιήσουμε όλες τις προηγούμενες μετρήσεις.

Από το 1. προκύπτει η συνθήκη $y = 1 - x$.

Από το 2. ότι $\frac{x}{y} = \frac{y}{1}$

Δηλαδή η θέση του x στο $[0, y]$ είναι ό,τι και του y στο $[x,1]$.

Άρα $x = y^2$ και $y = 1 - x$ ή $y = 1 - y^2$ ή $y^2 + y - 1 = 0$ και άρα $y^* = \frac{-1 + \sqrt{5}}{2}$ που είναι ο λόγος της χρυσής τομής.

Δηλαδή $y^* = 0.61803...$ επίσης $x^* = y^{*2} = 0.3820...$

Άρα με κάθε υπολογισμό της f το διάστημα αβεβαιότητας πολλαπλασιάζεται με $y^* (< 1)$ και έτσι μειώνεται εκθετικά!

Αν π.χ. το αρχικό διάστημα είναι 10 μονάδες, μετά από 20 υπολογισμούς της f θα έχει γίνει ίσο με 0.0007.

2.2 Βελτιστοποίηση με περιορισμούς

2.2.1 Ισοτικοί περιορισμοί – μέθοδος Lagrange

Πρόβλημα:

$$\max(\min) f(x_1, x_2, \ldots, x_n)$$

για $x_1, x_2, \ldots, x_n$ τέτοια ώστε

$$\left.\begin{array}{l} g_1(x_1, x_2, \ldots, x_n) = 0 \\ \vdots \\ g_k(x_1, x_2, \ldots, x_n) = 0 \end{array}\right\} k \text{ περιορισμοί}$$

2.2.1.1 Απλή λύση με αντικατάσταση

Αν μπορούμε να «λύσουμε» τις εξισώσεις των g_j θα προκύψουν $n-k$ ελεύθεροι άγνωστοι, τους οποίους μπορούμε να αντικαταστήσουμε στην f και να έχουμε ένα πρόβλημα με $n-k$ ελεύθερους αγνώστους.

Π.χ. $\min\ x^2 + 2y^2 + 3z^2$

$$x + y + z = 0$$

$$2x - y + z = 3$$

Γράφουμε τους περιορισμούς ως

$$x + y = -z$$

$$2x - y = 3 - z$$

Λύνοντας ως προς x, y έχουμε:

$$x = 1 - 2/3z \quad y = -1 - z/3$$

Οπότε το πρόβλημα γράφεται ως προς τον ελεύθερο άγνωστο z:

$$f(x, y, z) = x^2 + 2y^2 + 3z^2 = f(x(z), y(z), z) =$$

$$= g(z) = (1 - 2/3z)^2 + 2(1 + z/3)^2 + 3z^2$$

Που λύνεται θέτοντας απλά $\dfrac{\partial g}{\partial z} = 0$.

□

Η μέθοδος αυτή δεν είναι εφικτή αν δεν λύνεται το σύστημα $g_j(x_1, x_2, \ldots, x_n) = 0 \ \ j = 1, 2, \ldots, k$.

2.2.1.2 Έμμεση λύση με μέθοδο Lagrange

Εισαγωγή

Έστω $x, x+\Delta x$ (διανύσματα) ικανοποιούν τον περιορισμό $g=0$, δηλαδή $g(x)=g(x+\Delta x)=0$. Τότε, αν το Δx είναι «μικρό», αντιπροσωπεύει την εφαπτόμενη της ισοσταθμικής.

Επίσης: $g(x+\Delta x)=g(x)+\nabla g(x)\cdot\Delta x$

Αλλά: $g(x)=g(x+\Delta x)=0$, οπότε $\nabla g(x)\cdot\Delta x=0$

Άρα το $\nabla g(x)$ είναι <u>κάθετο</u> στην εφαπτόμενη της ισοσταθμικής άρα είναι και κάθετο στην ισοσταθμική.

Παράδειγμα:

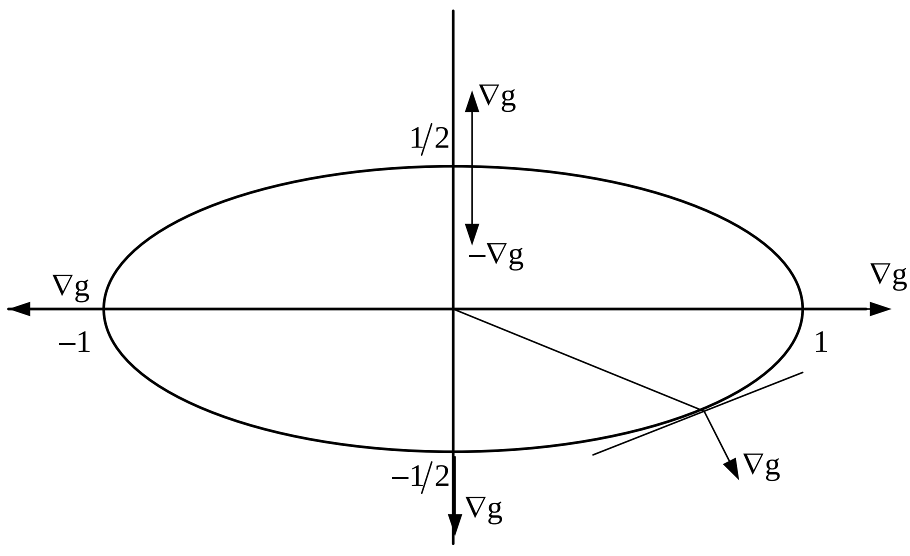
$$g(x,y)=x^2+y^2-1$$

και η ισοσταθμική $\{(x,y)|g(x,y)=0\}$, η οποία είναι περιφέρεια κύκλου ακτίνας 1.

Είναι $\nabla g=\left(\dfrac{\partial g}{\partial x},\dfrac{\partial g}{\partial y}\right)=(2x,2y)$ που προφανώς είναι ακτίνα του κύκλου, που είναι εξ' ορισμού κάθετη στην εφαπτόμενη σε οποιοδήποτε σημείο του.

$$g(x,y)=x^2+4y^2-1$$

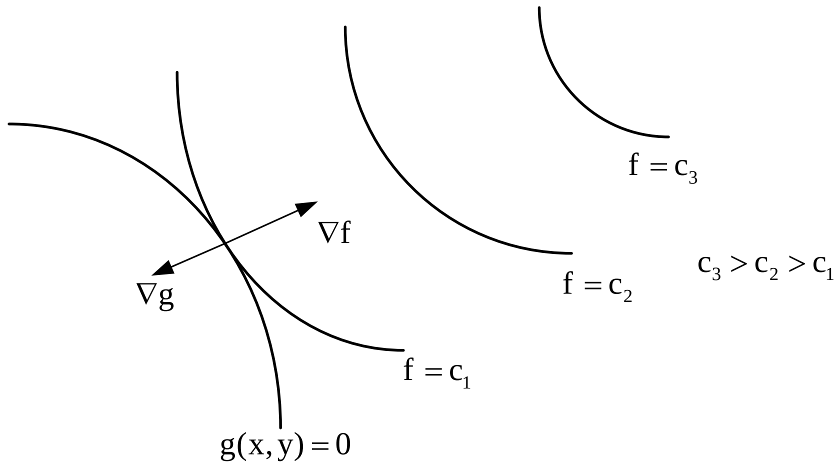
Η ισοσταθμική για $g=0$ είναι έλλειψη με κάθετη στο (x,y) ίση με $(2x,8y)$ (γιατί;)



2.2.1.3 Μέθοδος Lagrange με ένα περιορισμό και δύο μεταβλητές

Γεωμετρική εικόνα σε 2 διαστάσεις:

Έστω $\max f(x,y)$ με περιορισμό $g(x,y)=0$.



Όπως παρατηρούμε, στο βέλτιστο οι ισοσταθμικές $g(x, y) = 0$ και $f = c_1$ εφάπτονται, άρα τα ∇f και ∇g είναι <u>συγγραμικά</u>, δηλαδή $\nabla f = -\mu \cdot \nabla g$, όπου μ αριθμός!

Άρα αναγκαίες συνθήκες για να είναι το (x^*, y^*) βέλτιστη λύση είναι:

- $\nabla f(x^*, y^*) = -\mu \nabla g(x^*, y^*)$
- $g(x^*, y^*) = 0$

Η πρώτη σχέση συνεπάγεται δύο συνθήκες:

- $\dfrac{\partial f}{\partial x}(x^*, y^*) + \mu \cdot \dfrac{\partial g}{\partial x}(x^*, y^*) = 0$
- $\dfrac{\partial f}{\partial y}(x^*, y^*) + \mu \cdot \dfrac{\partial g}{\partial y}(x^*, y^*) = 0$

που μαζί με την $g(x^*, y^*) = 0$ αποτελούν 3 εξισώσεις.

Άρα, αν <u>αναζητήσουμε</u> το βέλτιστο, αυτό είναι μεταξύ των λύσεων του συστήματος εξισώσεων:

$$\frac{\partial f}{\partial x}(x, y) + \mu \cdot \frac{\partial g}{\partial x}(x, y) = 0$$

$$\frac{\partial f}{\partial y}(x, y) + \mu \cdot \frac{\partial g}{\partial y}(x, y) = 0$$

$$g(x, y) = 0$$

ως προς 3 αγνώστους x, y, μ.

<u>Παράδειγμα:</u>

$$\min x^2 + y^2 + xy$$

με $g(x, y) = x + y - 1 = 0$

Οι συνθήκες για το βέλτιστο είναι:

$$\frac{\partial f}{\partial x}+\mu\cdot\frac{\partial g}{\partial x}=2x+y+\mu=0$$

$$\frac{\partial f}{\partial y}+\mu\cdot\frac{\partial g}{\partial y}=2y+x+\mu=0$$

$$x+y=1$$

Οι δύο πρώτες δίνουν λύνοντας ως προς μ:

$$x=y=-\mu/3$$

Εφόσον $x=y$ η τρίτη σχέση δίνει $x=y=1/2$ και άρα $\mu=-3/2$.
Η τιμή του μ είναι πολύ χρήσιμη όπως θα δούμε και στην συνέχεια.

□

Για περισσότερες μεταβλητές, π.χ.:

$$\min f(x,y,z)=2x^2+y^2+2z^2+xy-2xz$$

με $g(x,y,z): x+y-z=5$

Οι συνθήκες είναι:

- $\nabla f+\mu\cdot\nabla g=0$ (3 σχέσεις)
- $g=0$ (4η σχέση)

Δηλαδή:

- $\frac{\partial f}{\partial x}+\mu\cdot\frac{\partial g}{\partial x}=4x+y-2z+\mu=0$
- $\frac{\partial f}{\partial y}+\mu\cdot\frac{\partial g}{\partial y}=2y+x+\mu=0$
- $\frac{\partial f}{\partial z}+\mu\cdot\frac{\partial g}{\partial z}=4z-2x-\mu=0$
- $x+y-z=5$

Λύνοντας τις τρεις πρώτες ως προς μ έχουμε:

$$x=0 \qquad y=-\mu/2 \qquad z=\mu/4$$

Οπότε από την τέταρτη σχέση προκύπτει ότι $\mu=-20/3$, από όπου προκύπτουν τα

$$x=0 \qquad y=10/3 \qquad z=-5/3$$

□

2.2.1.4 Γενική διατύπωση – Πολλοί περιορισμοί

Έστω τώρα ένα υποτιθέμενο βέλτιστο στο πρόβλημα:

$$\max(\min)\ f(x_1, x_2, \ldots, x_n)$$
με περιορισμούς:

$$\left.\begin{array}{l} g_1(x_1, x_2, \ldots, x_n) = 0 \\ \vdots \\ g_k(x_1, x_2, \ldots, x_n) = 0 \end{array}\right\} k \text{ περιορισμοί}$$

Το υποτιθέμενο βέλτιστο συμβολίζεται με $(x_1^*, x_2^*, \ldots, x_n^*)$.
Εξετάζουμε μία μεταβολή, δηλαδή το $(x_1^* + \Delta x_1, x_2^* + \Delta x_2, \ldots, x_n^* + \Delta x_n)$.

Εφόσον το $(x_1^* + \Delta x_1, x_2^* + \Delta x_2, \ldots, x_n^* + \Delta x_n)$ ικανοποιεί όλους τους περιορισμούς, θα πρέπει :
$\nabla g_j(x_1^*, x_2^*, \ldots, x_n^*) \cdot \left(\Delta x_1, \Delta x_2, \ldots, \Delta x_n\right) = 0$ για κάθε περιορισμό (γιατί;)

Επιπλέον θα πρέπει για να είναι το $(x_1^*, x_2^*, \ldots, x_n^*)$ βέλτιστο:
$\nabla f(x_1^*, x_2^*, \ldots, x_n^*) \cdot \left(\Delta x_1, \Delta x_2, \ldots, \Delta x_n\right) = 0$.
Διότι, αν π.χ. $\nabla f \cdot \Delta x > 0$, τότε μία μεταβολή στο x κατά Δx θα οδηγούσε σε αύξηση της αντικειμενικής (τι θα συνέβαινε αν $\nabla f \cdot \Delta x < 0$;)

Μπορεί να αποδειχθεί ότι το παραπάνω συνεπάγεται ότι το $\nabla f(x_1^*, x_2^*, \ldots, x_n^*)$ θα πρέπει να είναι γραμμικός συνδυασμός των $\nabla g_j(x_1^*, x_2^*, \ldots, x_n^*)$, δηλαδή:
$$\nabla f(x_1^*, x_2^*, \ldots, x_n^*) + \mu_1 \nabla g_1(x_1^*, \ldots, x_n^*) + \ldots + \mu_k \nabla g_k(x_1^*, \ldots, x_n^*) = 0$$
για κάποιους αριθμούς $\mu_1, \ldots, \mu_k$.

Έτσι οι <u>συνθήκες βέλτιστου</u> είναι οι εξής $n + k$ σχέσεις:

$\nabla f + \sum_{j=1}^{k} \mu_j \cdot \nabla g_j = 0$	n σχέσεις, μία για κάθε x_i
$g_j = 0$	k σχέσεις, μία για κάθε περιορισμό

Τα υποψήφια βέλτιστα είναι οι λύσεις του παραπάνω συστήματος εξισώσεων ως προς αγνώστους $x_1, x_2, \ldots, x_n, \mu_1, \mu_2, \ldots, \mu_k$.

Οι συνθήκες γράφονται συνήθως με βάση τη <u>Λαγκραντζιανή</u> (<u>Lagrangean</u>) συνάρτηση:
$$L(x_1, x_2, \ldots, x_n, \mu_1, \mu_2, \ldots, \mu_k) = f(x_1, \ldots, x_n) + \sum_{j=1}^{k} \mu_j g_j(x_1, \ldots, x_n)$$
Θα πρέπει λοιπόν να ισχύουν οι ακόλουθες συνθήκες:

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{k} \mu_j \cdot \frac{\partial g_j}{\partial x_i} = 0 \quad i = 1, \ldots, n$$

$$\frac{\partial L}{\partial \mu_j} = g_j(x_1, \ldots, x_n) = 0 \qquad j = 1, \ldots, k$$

<u>Παράδειγμα:</u>

$$\min x + 2y - z = f(x, y, z)$$

$$x^2 + y^2 + z^2 - 1 = 0 = g_1(x, y, z)$$

$$2x + 3y + z = 0 = g_2(x, y, z)$$

Έχουμε:

$$L = x + 2y - z + \mu_1(x^2 + y^2 + z^2 - 1) + \mu_2(2x + 3y + z)$$

οπότε:

$$\frac{\partial L}{\partial x} = 1 + 2\mu_1 x + 2\mu_2 = 0$$

$$\frac{\partial L}{\partial y} = 2 + 2\mu_1 y + 3\mu_2 = 0$$

$$\frac{\partial L}{\partial z} = -1 + 2\mu_1 z + \mu_2 = 0$$

$$\frac{\partial L}{\partial \mu_1} = x^2 + y^2 + z^2 - 1 = 0$$

$$\frac{\partial L}{\partial \mu_2} = 2x + 3y + z = 0$$

Το σύστημα αυτό (5 εξισώσεις-5 άγνωστοι) είναι μη γραμμικό, όμως μπορεί να απλοποιηθεί:

Από τις 3 πρώτες σχέσεις έχουμε:

$$x = -\frac{1}{2\mu_1}(2\mu_2 + 1) \qquad y = -\frac{1}{2\mu_1}(3\mu_2 + 2)$$

$$z = -\frac{1}{2\mu_1}(\mu_2 - 1)$$

Οπότε αντικαθιστώντας στην τελευταία σχέση και απαλείφοντας το $-\frac{1}{2\mu_1}$ παίρνουμε:

$$2x + 3y + z = 0 \Rightarrow$$

$$-\frac{1}{2\mu_1} 2(2\mu_2 + 1) - \frac{3}{2\mu_1}(3\mu_2 + 2) - \frac{1}{2\mu_1}(\mu_2 - 1) \Rightarrow$$

$$4\mu_2 + 2 + 9\mu_2 + 6 + \mu_2 - 1 = 0 \Rightarrow \ \mu_2 = -1/2$$

Οπότε:

$$x = 0 \qquad y = \lambda \qquad z = -3\lambda \qquad (\lambda = -\frac{1}{4\mu_1})$$

Από την $4^η$ σχέση είναι:

$$x^2 + y^2 + z^2 - 1 = 0 \ \Rightarrow \ \lambda^2 + 9\lambda^2 = 1 \ \Rightarrow$$

$$\lambda = \pm\frac{1}{\sqrt{10}}$$

Ποιο λ είναι όμως το σωστό, το θετικό ή το αρνητικό;

Αντικαθιστώντας στην $f = x + 2y - z \Rightarrow f = 11\lambda$

Βλέπουμε ότι πρέπει να επιλεγεί το αρνητικό λ για ελάχιστο και το θετικό για μέγιστο! (Παρατήρηση: η λύση αυτή στον Solver έχει μικρότερη ακρίβεια…)

□

2.2.1.5 Ερμηνεία των πολλαπλασιαστών μ

Τα μ_j ονομάζονται <u>Πολλαπλασιαστές Lagrange</u> κι έχουν μεγάλη χρησιμότητα.

Έστω ότι είχαμε λύσει ένα πρόβλημα για το οποίο είχαν υπολογισθεί οι Πολλαπλασιαστές. Θεωρούμε το ίδιο πρόβλημα με νέους περιορισμούς $g_j(x_1, x_2, \ldots, x_n) = \varepsilon_j$ αντί για $g_j(x_1, x_2, \ldots, x_n) = 0$. Έστω ότι το προηγούμενο βέλτιστο ήταν στο $(x_1^*, x_2^*, \ldots, x_n^*)$ και το νέο βέλτιστο στο $(x_1^* + \delta x_1, x_2^* + \delta x_2, \ldots, x_n^* + \delta x_n)$. Τότε, αν τα ε_j είναι μικρά, θα πρέπει: $\nabla g_j(x_1^*, \ldots, x_n^*) \cdot (\delta x_1, \ldots, \delta x_n) = \varepsilon_j$ (γιατί;)

Η δε μεταβολή στο βέλτιστο είναι $\nabla f(x_1^*, \ldots, x_n^*) \cdot (\delta x_1, \ldots, \delta x_n)$.

Αλλά, $\nabla f = -\sum_{j=1}^{k} \mu_j \cdot \nabla g_j$ και άρα:

$$\nabla f \cdot \vec{\delta}x = -\sum_{j=1}^{k} \mu_j \cdot (\nabla g_j \cdot \vec{\delta}x) = -\sum_{j=1}^{k} \mu_j \cdot \varepsilon_j$$

Αυτό σημαίνει ότι η μεταβολή κατά ε_j του j-περιορισμού μεταβάλλει την αντικειμενική συνάρτηση στο βέλτιστο κατά $(-\mu_j) \cdot \varepsilon_j$, δηλαδή το μ_j δείχνει πόσο επηρεάζει ο j-περιορισμός την αντικειμενική συνάρτηση!

<u>Εφαρμογή:</u> Στο παράδειγμα ενός περιορισμού, $\min x^2 + y^2 + xy$ με $g(x, y) = x + y - 1 = 0$, το βέλτιστο ήταν $3/4$ ($x = y = 1/2, \ \mu = -3/2$).

Αν τώρα ο περιορισμός γίνει $x+y=1.1$, τα βέλτιστα που προκύπτουν (με τον Solver) είναι $x=y=0.55$ και η αντικειμενική συνάρτηση είναι 0.9075, δηλαδή μία μεταβολή (αύξηση) $\Delta f = 0.9075-0.75=0.1575$.
Η μεταβολή που προβλέπει η θεωρία είναι $(-\mu)\cdot\varepsilon = -(-3/2)\cdot 0.10 = 0.1500$.

□

2.2.2 Αριθμητική λύση συστημάτων εξισώσεων

Τα παραπάνω προβλήματα βελτιστοποίησης ανάγονται μέσω της μεθόδου Lagrange, στην επίλυση ενός συστήματος n εξισώσεων με n αγνώστους. Δίνουμε μία επισκόπηση των βασικών αριθμητικών μεθόδων για λύση τέτοιων προβλημάτων.

Ας θεωρήσουμε αρχικά μία απλή εξίσωση f(x)=0, f:R→R.

2.2.2.1 Μέθοδος Διχοτόμησης

Ο αλγόριθμος έχει ως εξής:

0. Αρχίζουμε με δύο σημεία x_i^{min} και x_i^{max} ($x_i^{min} < x_i^{max}$) τέτοια ώστε $f(x_i^{min}) < 0$ και $f(x_i^{max}) > 0$ (ερώτηση: πώς βρίσκουμε αρχικά τέτοια σημεία;). Γνωρίζουμε ότι υπάρχει τουλάχιστον ένα x^* με $f(x^*)=0$ και $x^* \in [x_i^{min}, x_i^{max}]$.

1. Θέτουμε $\bar{x}_i = \dfrac{x_i^{min} + x_i^{max}}{2}$

- Αν $f(\bar{x}_i) > 0$, τότε $x^* \in [x_i^{min}, \bar{x}_i]$ και άρα θέτουμε $x_{i+1}^{min} \leftarrow x_i^{min}$ και $x_{i+1}^{max} \leftarrow \bar{x}_i$.
- Αν $f(\bar{x}_i) < 0$, τότε $x^* \in [\bar{x}_i, x_i^{max}]$ και άρα θέτουμε $x_{i+1}^{max} \leftarrow x_i^{max}$ και $x_{i+1}^{min} \leftarrow \bar{x}_i$.

Τώρα $\Delta x_{i+1} = [x_{i+1}^{max} - x_{i+1}^{min}] = \dfrac{\Delta x_i}{2}$.

Θέτουμε $i \leftarrow i+1$.

2. Αν $\Delta x_i < \varepsilon$ σταματάμε, διαφορετικά επιστρέφουμε στο βήμα 1.

Η μέθοδος είναι σχετικά απλή, δεν χρειάζεται υπολογισμούς παραγώγων, όμως είναι σχετικά αργή: αν αρχίσουμε με αβεβαιότητα $\Delta x = 1$, τότε σε 10 επαναλήψεις έχουμε ακρίβεια μόνο ενός χιλιοστού ($=\dfrac{1}{2^{10}}$)!

2.2.2.2 Μέθοδος Νεύτωνα

Η μέθοδος του Νεύτωνα είναι σαφώς ταχύτερη, αλλά απαιτεί υπολογισμούς παραγώγων. Μία προσέγγιση στην f ικανοποιητική αν η f'' είναι μικρή (γιατί;) είναι η $f(x+\Delta x)\cong f(x)+f'(x)\Delta x$. Υπολογίζουμε ένα Δx ώστε η γραμμική προσέγγιση να μηδενίζεται, δηλαδή $0=f(x_{n+1})\cong f(x_n)+f'(x_n)(x_{n+1}-x_n)$ οπότε $x_{n+1}=x_n-\dfrac{f(x_n)}{f'(x_n)}$.

Η μέθοδος είναι ταχύτατη, καθώς αν x^* είναι ρίζα και το x_n είναι κοντά στο x^* ισχύει ότι:

$$|x_{n+1}-x^*|<\beta|x_n-x^*|^2$$

Έτσι ανεξάρτητα από το β το διάστημα $|x_{n+1}-x^*|$ μειώνεται ταχύτατα.

<u>Παράδειγμα:</u>

Υπολογίστε τον φυσικό λογάριθμο του 10 χρησιμοποιώντας την ενσωματωμένη συνάρτηση του MS Excel e^x και όχι την ενσωματωμένη $\ln x$.

Θέλουμε να βρούμε ουσιαστικά τη ρίζα της $f(x)=e^x-10$.

Η μέθοδος Νεύτωνα δίνει (ρίζα της γραμμικής προσέγγισης στο x_n):

$$x_{n+1}=x_n-\frac{e^{x_n}-10}{e^{x_n}}=x_n-1+10\cdot e^{-x_n}$$

Τα αποτελέσματα αν $x_0=1$ δίνονται από τον παρακάτω πίνακα.

Αριθμός Επανάληψης	x_n	$\exp(x_n)$
0	1.00000000	2.71828183
1	3.67879441	39.59862564
2	2.93132843	18.75252535
3	2.46458994	11.75865940
4	2.31502702	10.12519652
5	2.30266217	10.00077084
6	2.30258510	10.00000003
7	2.30258509	10.00000000
8	2.30258509	10.00000000

□

<u>Πολλές διαστάσεις</u>

Η μέθοδος Νεύτωνα επεκτείνεται και σε πολλές μεταβλητές. Ενδεικτικά για δύο μεταβλητές ισχύει:

$$f(x+\Delta x,y+\Delta y)\cong f(x,y)+f_x(x,y)\Delta x+f_y(x,y)\Delta y$$

$$g(x+\Delta x,y+\Delta y)\cong g(x,y)+g_x(x,y)\Delta x+g_y(x,y)\Delta y$$

όπου $f,g:R^2\to R^2$ και $f_x=\dfrac{\partial f}{\partial x}$, $f_y=\dfrac{\partial f}{\partial y}$, $g_x=\dfrac{\partial g}{\partial x}$, $g_y=\dfrac{\partial g}{\partial y}$

Θέτοντας $f(x+\Delta x, y+\Delta y) = g(x+\Delta x, y+\Delta y) = 0$ έχουμε ότι:

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = -J(x,y)^{-1} \cdot \begin{bmatrix} f(x,y) \\ g(x,y) \end{bmatrix}$$

Όπου ο πίνακας $J(x,y) = \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \\ \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{bmatrix}$ ονομάζεται <u>Jacobian</u>.

Άρα η μέθοδος Νεύτωνα γράφεται ως:

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \end{bmatrix} - J(x_n, y_n)^{-1} \cdot \begin{bmatrix} f(x_n, y_n) \\ g(x_n, y_n) \end{bmatrix}.$$

και συνεπάγεται την αντιστροφή του πίνακα Jacobian, κάτι αρκετά επιβαρυντικό για προβλήματα πολλών μεταβλητών.

<u>Παράδειγμα:</u> Λύστε το σύστημα:

$$f(x,y) = ye^x - 10 = 0$$

$$g(x,y) = x + y - 4 = 0$$

Με τη μέθοδο Νεύτωνα δύο διαστάσεων είναι:

$$\begin{bmatrix} f \\ g \end{bmatrix} = \begin{bmatrix} ye^x - 10 \\ x + y - 4 \end{bmatrix} \quad \text{και} \quad J(x,y) = \begin{bmatrix} ye^x & e^x \\ 1 & 1 \end{bmatrix}$$

ενώ $J(x,y)^{-1} = \dfrac{1}{y-1} \cdot \begin{bmatrix} e^{-x} & -1 \\ -e^{-x} & 4 \end{bmatrix}$ κ.λ.π. $\Rightarrow$ πρέπει $y \neq 1$.

Η σειρά των επαναλήψεων για $x_0 = 4,\ y_0 = 0$ δίνεται από τον παρακάτω πίνακα.

					Jacobean			**InvJacobean**			**Dx, Dy**	
Αριθμός επανάληψης	**x**	**y**	**f-g**		**x**	**y**		**x**	**y**			
0	4	0	-10	**f**	0	54.59815		-0.0183156	1		0.1831564	Dx
			0	**g**	1	1		0.0183156	0		-0.1831564	Dy
1	3.8168436	0.1831564	-1.6736207	**f**	8.3263793	45.460491		-0.0269294	1.2242245		0.0450696	Dx
			0	**g**	1	1		0.0269294	-0.2242245		-0.0450696	Dy
2	3.771774	0.228226	-0.0819617	**f**	9.9180383	43.457089		-0.029816	1.2957161		0.0024438	Dx
			0	**g**	1	1		0.029816	-0.2957161		-0.0024438	Dy
3	3.7693302	0.2306698	-0.0002296	**f**	9.9997704	43.351019		-0.0299839	1.299832		6.885E-06	Dx
			0	**g**	1	1		0.0299839	-0.299832		-6.885E-06	Dy
4	3.7693233	0.2306767	-1.818E-09	**f**	10	43.350721		-0.0299844	1.2998436		5.451E-11	Dx
			0	**g**	1	1		0.0299844	-0.2998436		-5.451E-11	Dy
5	3.7693233	0.2306767	0	**f**	10	43.350721		-0.0299844	1.2998436		0	Dx
			0	**g**	1	1		0.0299844	-0.2998436		0	Dy
6	3.7693233	0.2306767	0	**f**	10	43.350721		-0.0299844	1.2998436		0	Dx
			0	**g**	1	1		0.0299844	-0.2998436		0	Dy

2.2.3 Ανισοτικοί περιορισμοί

Θυμηθείτε ότι αν το πρόβλημα $\max f(\vec{x})$ με $g(\vec{x})=0$, έχει λύση με πολλαπλασιαστή μ ($L=f+\mu g$), τότε, αν ο περιορισμός γίνει $g(\vec{x})=\varepsilon$ η αξία του μέγιστου θα αλλάξει κατά $-\mu\cdot\varepsilon$.

Θεωρήστε τώρα το πρόβλημα:

$$\begin{array}{lcl} \max f(x_1,x_2,\ldots,x_n) & & \min f(x_1,x_2,\ldots,x_n) \\ g_1(x_1,x_2,\ldots,x_n)\geq 0 & & g_1(x_1,x_2,\ldots,x_n)\geq 0 \\ \vdots & \textbf{ή} & \vdots \\ g_k(x_1,x_2,\ldots,x_n)\geq 0 & & g_k(x_1,x_2,\ldots,x_n)\geq 0 \end{array}$$

(Τυπική μορφή μεγίστου) (Τυπική μορφή ελαχίστου)

Παρατηρήστε ότι η έκφραση αυτή είναι γενικότερη από την ισοτική περίπτωση, καθώς οι ισοτικοί περιορισμοί π.χ $g(\vec{x})=0$ είναι δύο ανισοτικοί: $g(\vec{x})\geq 0$ και $-g(\vec{x})\geq 0$.

Σε αυτήν την περίπτωση των ανισοτικών περιορισμών οι αναγκαίες συνθήκες είναι περίπου ίδιες με τις Lagrange και οφείλονται στους Karush (1930) και Kuhn-Tucker (1950) γι’αυτό και συνήθως τις αποκαλούμε Συνθήκες KuhnTucker (K-T).

Στην περίπτωση των ισοτικών περιορισμών οι Συνθήκες Lagrange δίνουν:

$$L=f+\mu_1 g_1+\mu_2 g_2+\ldots+\mu_k g_k$$

$$\frac{\partial L}{\partial x_i}=0 \qquad i=1,\ldots,N$$

$$\frac{\partial L}{\partial \mu_j}=0 \qquad j=1,\ldots,k$$

και τετριμμένα:

$$\mu_j\cdot g_j=0$$

Στην περίπτωση ανισοτικών περιορισμών και συγκεκριμένα για την τυπική μορφή μεγίστου οι Συνθήκες Kuhn-Tucker γράφονται ως:

Για βέλτιστο $\vec{x}$ υπάρχουν $\mu_j\geq 0$ $j=1,2,\ldots,k$ ώστε να ισχύουν οι Συνθήκες Lagrange:

$$\frac{\partial L}{\partial x_i}=0 \qquad i=1,\ldots,N$$

$$\mu_j\cdot g_j(\vec{x})=0 \quad j=1,2,\ldots,k$$

$g_j(\vec{x})\geq 0 \qquad j=1,2,\ldots,k$ (ή $\dfrac{\partial L}{\partial \mu_j}\geq 0$ όπου L η Lagrangean συνάρτηση)

Η επιβεβαίωση του αν ένα δεδομένο $\vec{x}$ ικανοποιεί τις Συνθήκες K-T είναι να βρεθεί αν υπάρχει λύση θετική των ισοτήτων $\frac{\partial L}{\partial x_i} = 0$.

2.2.3.1 Αλγεβρική δικαιολόγηση

Ιδέα: οι ανισοτικοί περιορισμοί $g_j \geq 0$ γράφονται ως $g_j - \varepsilon_j^2 = 0$ με πρόσθετη μεταβλητή ε_j (τα ε_j είναι αντίστοιχα με τις μεταβλητές απόκλισης στο Γραμμικό Προγραμματισμό).

Έχουμε λοιπόν για το νέο πρόβλημα:

$$L = f + \mu_1 \cdot (g_1 - \varepsilon_1^2) + \ldots + \mu_k \cdot (g_k - \varepsilon_k^2)$$

Οι συνθήκες είναι:

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \mu_1 \cdot \frac{\partial g_1}{\partial x_i} + \ldots + \mu_k \cdot \frac{\partial g_k}{\partial x_i}$$ (τα ε_j δεν εμφανίζονται!)

$$\frac{\partial L}{\partial \mu_j} = g_j - \varepsilon_j^2 = 0$$ (οι περιορισμοί)

$$\frac{\partial L}{\partial \varepsilon_j} = -2\mu_j \cdot \varepsilon_j = 0$$ που είναι ισοδύναμο με $\mu_j \cdot \varepsilon_j^2 = 0$ ή $\mu_j \cdot g_j = 0$.

Ποιό είναι όμως το πρόσημο του μ_j;

Έστω ότι έχουμε πρόβλημα μεγιστοποίησης. Αν γράψουμε $g_j - \varepsilon_j^2 = \eta$ με $\eta > 0$, τότε $g_j = \varepsilon_j^2 + \eta \geq \eta$. Άρα το νέο πρόβλημα έχει λιγότερες εφικτές λύσεις από το πρόβλημα με $g_j \geq 0$ (ή $g_j - \varepsilon_j^2 = 0$). Συνεπώς το νέο πρόβλημα πρέπει να έχει μικρότερη ή ίση λύση (βέλτιστη τιμή της αντικειμενικής συνάρτησης).
Δηλαδή, από τη Θεωρία Lagrange (βλ. ενότητα 2.2.1.5) πρέπει: $\left(-\mu_j\right) \cdot \eta \leq 0 \Rightarrow \mu_j \geq 0$.

Σε πρόβλημα ελαχιστοποίησης θα πρέπει:
$\left(-\mu_j\right) \cdot \eta \geq 0 \Rightarrow \mu_j \leq 0$ (γιατί;)

Παράδειγμα:

$\min x^2 + y^2$ με $x + y - 1 \geq 0$

$$L = x^2 + y^2 + \mu \cdot \left(x + y - 1\right)$$

Όπως προηγουμένως έχουμε:

$$x = y = -\mu/2$$

Αν $\mu \neq 0$ έχουμε $x = y = 1/2$ και $\mu = -1$.

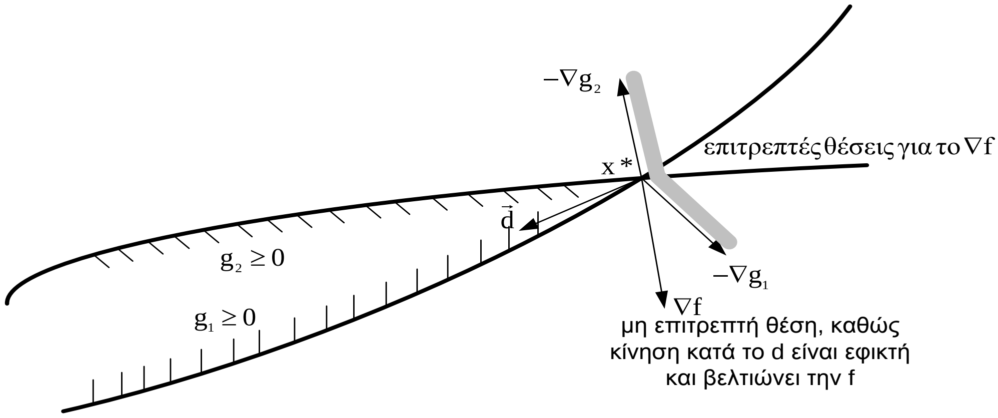
Δηλαδή το $x = y = 1/2$ ικανοποιεί τις συνθήκες K-T.

□

2.2.3.2 Γεωμετρική δικαιολόγηση

Έστω υποψήφιο βέλτιστο x^* (σε πρόβλημα μεγιστοποίησης) και έστω $g_i(x^*) = 0, i = 1,2$ και $g_i(x^*) > 0, i = 3,\ldots,k$ (δηλαδή δύο από τους περιορισμούς είναι ισότητες στο βέλτιστο). Για να είναι το x^* όντως βέλτιστο, δεν μπορεί να υπάρχει διάνυσμα $\vec{d}$ που να ικανοποιεί $\nabla f(x^*) \cdot \vec{d} > 0$ και $\nabla g_i(x^*) \cdot \vec{d} > 0, i = 1,2$.
[Διότι τότε για μικρό αριθμό $\lambda > 0$ εξετάζουμε το νέο υποψήφιο βέλτιστο $x^o = x^* + \lambda \cdot \vec{d}$. Για αυτό το υποψήφιο βέλτιστο είναι $\Delta f = \nabla f(x^*) \cdot \vec{d} \cdot \lambda > 0$ και $\Delta g_i = \nabla g_i(x^*) \cdot \vec{d} \cdot \lambda > 0,\ i = 1,2$, ενώ αν το λ είναι αρκετά μικρό ικανοποιούνται και οι υπόλοιποι ανισοτικοί περιορισμοί. Άρα το x^o είναι βελτίωση του x^* που δεν είναι συνεπώς βέλιστο!**]**

Εξετάζοντας το παραπάνω επιχείρημα προκύπτουν οι παρακάτω επιτρεπτές θέσεις του $\nabla f(x^*)$ που χαρακτηρίζονται από την συνθήκη:
Το ∇f είναι θετικός γραμμικός συνδυασμός των ∇g_i με για τα i με $g_i = 0$.



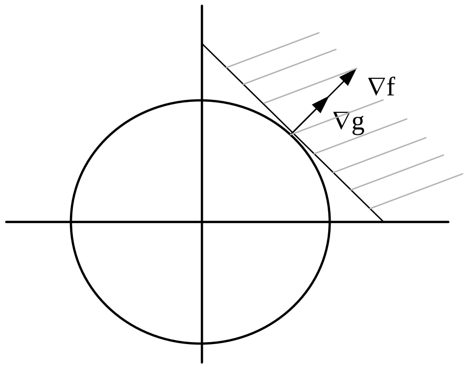
Άρα στο βέλτιστο x^* είναι $\nabla f(x^*) = \sum_{j=1}^{s} -\mu_j \cdot \nabla g_j(x^*)$, όπου $g_j(x^*) = 0,\ j = 1,2,\ldots s$.

Προφανώς παίρνουμε τις Συνθήκες Kuhn-Tucker $\nabla f(x^*) + \sum_{j=1}^{k} \mu_j \cdot \nabla g_j(x^*) = 0$ αν θέσουμε $\mu_j = 0$ για $j = s+1,\ldots,k$ όπου θεωρούμε ότι $g_j(x^*) > 0$.

2.2.3.3 Παραδείγματα εφαρμογής συνθηκών ΚΤ

Παράδειγμα 1:

$$\min\ x^2 + y^2$$
$$x + y \geq 1$$



Στο βέλτιστο, τα ∇f και ∇g πρέπει να είναι συγγραμικά και ίδιας κατεύθυνσης. Διαφορετικά, αν κινούμαστε κατά το ∇g θα είχαμε επιτρεπτή μετακίνηση που θα μείωνε την f. Άτοπο για βέλτιστο!

Γράφουμε:
$$\begin{aligned} &\min\ x^2 + y^2 = -\max\left(-x^2 - y^2\right) \\ &x + y - 1 \geq 0 \end{aligned}$$

Άρα:

$$L = -x^2 - y^2 + \mu \cdot \left(x + y - 1\right)$$

$$\left.\begin{aligned} \frac{\partial L}{\partial x} &= -2x + \mu = 0 \\ \frac{\partial L}{\partial y} &= -2y + \mu = 0 \end{aligned}\right\} x = y = \mu/2$$

Άρα αν $\mu > 0$ από την $\mu \cdot \left(x + y - 1\right)$=0 συνεπάγεται ότι

$$x + y - 1 = 0 \overset{x=y}{\Rightarrow} x = y = 1/2$$

Επομένως $\mu = 1 > 0$.

Δηλαδή, η λύση $x = y = 1/2$ και $\mu = 1$ ικανοποιεί τις Συνθήκες Κ-Τ.

□

Αν στο προηγούμενο πρόβλημα προσθέσουμε τον περιορισμό $x + 2y \geq 1$, τότε προφανώς η λύση δεν αλλάζει. Τότε ο πολλαπλασιαστής μ αυτής της σχέσης μπορεί να θεωρηθεί μηδενικός!

Παράδειγμα 2:

Το πρόβλημα $\max\ x^2 + y^2$ με $x + y \geq 1$

έχει $L = x^2 + y^2 + \mu \cdot (x + y - 1)$.

Οπότε $x = y = -\mu/2$ (γιατί;)

Αν $\mu > 0$, τότε $x = y = 1/2$ αλλά και $x = y = -\mu/2 < 0$. Άτοπο!

Άρα πρέπει $\mu = 0$, οπότε $x = y = 0$, αλλά και αυτό άτοπο εφόσον $x + y \geq 1$!

Επομένως, δεν υπάρχουν σημεία ικανοποίησης των συνθηκών Κ-Τ! (γιατί;)

□

Παράδειγμα 3:

$\min\ x^2 + y^2$ με $x + y \geq -1$

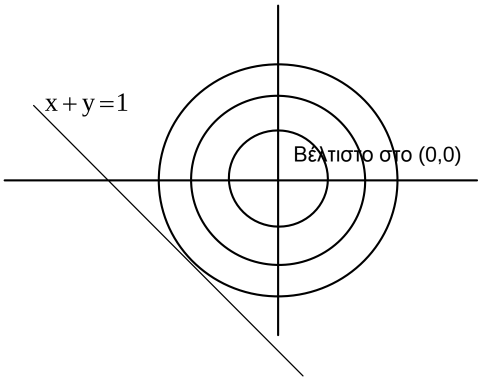
Ο περιορισμός γράφεται ισοδύναμα x+y+1≥0, και έχουμε πάλι **L=-x²-y²+μ(x+y+1)**

Οι συνθήκες Κ-Τ λένε ότι υπάρχει στο ελάχιστο x^o, y^o ένας αριθμός μ≥0 ώστε

(α) $\partial\mathcal{L}/\partial x = -2x+\mu=0$ και $\partial\mathcal{L}/\partial y = -2y+\mu=0$

(β) $\mu(x^o + y^o + 1) = 0$

Αν είναι μ>0, από το (α) έχουμε x=y=μ/2>0 και λόγω του (β) θα πρέπει x+y+1=0, πού όμως είναι άτοπο. Άρα πρέπει μ=0, και λόγω του (α), x=y=0= μ. Αυτό το αποτέλεσμα είναι το ίδιο με αυτό που θα προέκυπτε αν είχαμε να βρούμε ακρότατο χωρίς ανισοτικούς περιορισμούς. Έχουμε δηλαδή μία περίπτωση αναποτελεσματικών περιορισμών, οπότε οι συνθήκες ΚΤ ανάγονται σε ισοτικές συνθήκες Lagrange.



□

Παράδειγμα 4:

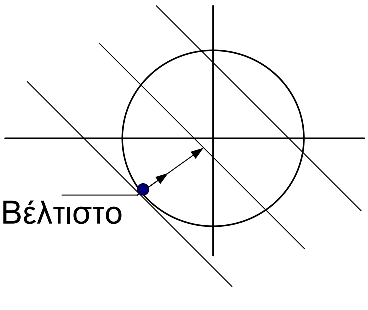
$\min\ x + y$ με $1 - (x^2 + y^2) \geq 0$

Θέτουμε $\mathcal{L} = -x - y + \mu[1-(x^2+y^2)]$ για μ≥0. Πρέπει να είναι για βέλτιστο:

(α) $\partial\mathcal{L}/\partial x = -1-2\mu x=0$ και $\partial\mathcal{L}/\partial y = -1-2\mu y=0$

(β) $\mu(1-(x^2+y^2)) = 0$

Από το (α) φαίνεται ότι δεν μπορεί να είναι μ=0. Άρα είναι μ>0, οπότε από το (α) x=y=-1/2μ που είναι αρνητικά. Εφόσον μ>0, πρέπει ο περιορισμός να είναι αποτελεσματικός δηλαδή $1= x^2+y^2$ και άρα x=y=-1/√2. Επομένως είναι μ=1/√2.



□

Παράδειγμα 5:

$$\max\ x+y \ \ \text{με}\ 1-(x^2+y^2)\geq 0$$

Θέτουμε $\mathcal{L} = x + y + \mu[1-(x^2+y^2)]$ με $\mu\geq 0$ και οι συνθήκες είναι:

(α) $\partial \mathcal{L}/\partial x = 1-2\mu x=0$ και $\partial \mathcal{L}/\partial y = 1-2\mu y=0$

(β) $\mu(1-(x^2+y^2)) = 0$

Πάλι η περίπτωση μ=0 απορρίπτεται οπότε είναι μ>0, και από το (α): $x=y=1/(2\mu)$, που είναι θετικά. Εφόσον μ>0, πρέπει ο περιορισμός να είναι αποτελεσματικός δηλαδή $1= x^2+y^2$ και άρα $x=y=1/\sqrt{2}$. Επομένως είναι $\mu=1/\sqrt{2}$. Παράβαλε με το Παράδειγμα 4.

□

Παράδειγμα 6:

$$\max\ x+y \ \ \text{με}\ 1-(x^2+y^2)\geq 0\ \ \text{και}\ \ y\geq 0$$

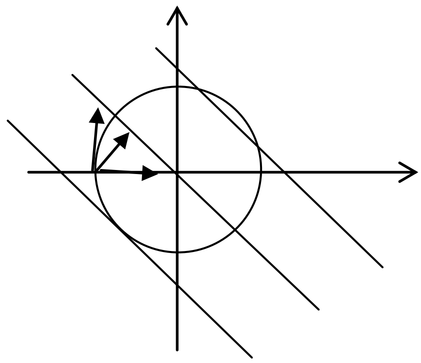
Έχουμε $\mathcal{L} = -x-y+ \mu_1[1-(x^2+y^2)] + \mu_2 y$, με μ_1 και $\mu_2 \geq 0$.

Οι συνθήκες είναι

(α) $\partial\mathcal{L}/\partial x = -1-2\mu_1 x=0$ και $\partial\mathcal{L}/\partial y = -1-2\mu_1 y+\mu_2=0$

(β) $\mu_1(1-(x^2+y^2)) = 0$ και $\mu_2 y = 0$

Είναι πάλι $\mu_1>0$, διαφορετικά -1=0 από την πρώτη σχέση της (α). Αν τώρα $\mu_2=0$, από την δεύτερη σχέση της (α) προκύπτει ότι $1 = -2\mu_1 y$ που όμως μας δίνει y αρνητικό και μη αποδεκτό λόγω του περιορισμού $y\geq 0$. Αρα το μ_2 είναι αυστηρά θετικό και επομένως y=0. Από την σχέση $1=-2\mu_1 x$ της (α) έπεται ότι $\mu_1>0$ και άρα $1-x^2-y^2 = 1-x^2 = 0$ οπότε $x^2 = 1$ και $x = \pm 1$. Αλλά $x =-1/2\mu_1 <0$ και επομένως x =-1. Επομένως είναι $\mu_1 =1/2$ και από την σχέση $1 = -2\mu_1 y+\mu_2$ έπεται ότι $\mu_2=1$. Βλέπε Σχήμα 6. Στο σημείο (-1,0) έχουμε $\nabla g_1=(-2x,-2y) = (2,0)$ ενώ $\nabla g_2=(0,1)$. Έχουμε τέλος $-\nabla f=(1,1)$ και ισχύει όντως ότι $\nabla f + \mu_1\nabla g_1+\mu_2\nabla g_2 =0$ για τις παραπάνω τιμές των $\mu_1,\nabla g_1$, μ_2 και ∇g_2. Παρατηρήστε ότι στο (-1,0) ικανοποιούνται ισοτικά και οι δύο περιορισμοί.



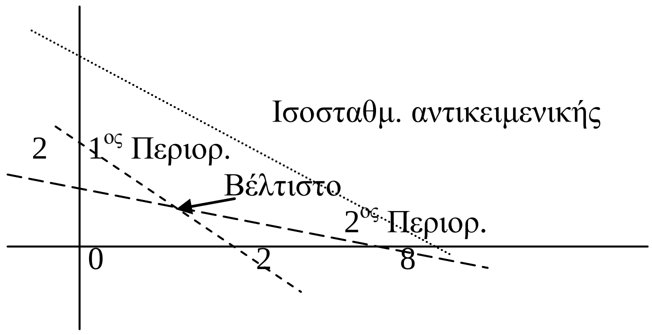
□

Παράδειγμα 7: (Επιβεβαίωση)

Οι συνθήκες Kuhn Tucker μπορούν να χρησιμοποιηθούν για να επιβεβαιώσουν ότι ένα δεδομένο σημείο είναι ή δεν είναι το βέλτιστο ενός προβλήματος. Η εργασία αυτή είναι αρκετά απλούστερη από το να εντοπισθούν όλες οι λύσεις των συνθηκών Kuhn Tucker.

Έστω λοιπόν το πρόβλημα **max 3x+4y με περιορισμούς $x+y\leq 2$, $x+8y\leq 8$, $x,y\geq 0$**.

Διαγραμματικά βλέπουμε πώς το βέλτιστο είναι στο x=8/7 y=6/7. Πώς επιβεβαιώνουμε ότι όντως το σημείο αυτό είναι βέλτιστο;



Θέτουμε $\mathcal{L}$ = 3x+4y+ μ_1 [2-x-y] + μ_2[8-x-8y]+ μ_3x + μ_4y με $\mu_1, \mu_2, \mu_3, \mu_4 \geq 0$.

Οι συνθήκες βελτιστοποίησης είναι:

(α) $\partial\mathcal{L}/\partial x = 3-\mu_1-\mu_2+\mu_3=0$ και $\partial\mathcal{L}/\partial y = 4-\mu_1-8\mu_2+\mu_4=0$

(β) $\mu_1[2-x-y]=0$ $\mu_2[8-x-8y]=0$ $\mu_3x=0$ $\mu_4y=0$

Αν x=8/7 y=6/7, θα είναι $\mu_3=\mu_4=0$ οπότε οι σχέσεις (α) γίνονται ένα σύστημα 2 εξισώσεων με δύο αγνώστους $\mu_1+\mu_2=3$ και $\mu_1+8\mu_2=4$ που έχει μοναδική λύση $\mu_1=20/7$ και $\mu_2=1/7$. Τα μ που εντοπίσθηκαν με τον τρόπο αυτό είναι μη αρνητικά, οπότε οι συνθήκες Kuhn Tucker ικανοποιούνται κατ' αρχήν – πράγμα που δεν θα συνέβαινε αν η λύση στο σύστημα προέκυπτε αρνητική. Πρέπει όμως επιπλέον λόγω της θετικότητας των μ και των σχέσεων (β) να ισχύει και 2-x-y = 8-x-8y =0, που επιβεβαιώνονται για τις συγκεκριμένες τιμές των x,y.

Μία άλλη κορυφή στο διάγραμμα των περιορισμών είναι η x=0, y=1. Εξετάζουμε αν στο σημείο αυτό ικανοποιούνται οι συνθήκες βελτίστου. Για τις τιμές αυτές θα είναι $\mu_4=0$ και επιπλέον καθώς η $x+y\leq 2$ ικανοποιείται ανισοτικά, θα πρέπει $\mu_1=0$ λόγω της πρώτης σχέσης στις (β). Έτσι οι σχέσεις (α) γίνονται πάλι ένα σύστημα δύο εξισώσεων με δύο αγνώστους και συγκεκριμένα $3-\mu_2+\mu_3=0$, $4-8\mu_2=0$. Η λύση των είναι $\mu_2=1/2$ και $\mu_3=-5/2$, πράγμα που σημαίνει ότι οι συνθήκες ΔΕΝ ικανοποιούνται.

□

<u>Παράδειγμα 8:</u> (Επιβεβαίωση)

Έστω το πρόβλημα **max x^2+3y^2 με περιορισμούς $x+y\leq 2$, $x+8y\leq 8$, $x,y\geq 0$**.

Οι περιορισμοί είναι οι ίδιοι όπως προηγουμένως, αλλά από την διαγραμματική παράσταση δεν είναι σαφές που θα είναι το βέλτιστο, δεδομένου ότι οι ισοσταθμικές της αντικειμενικής δεν είναι γραμμικές. Από το διάγραμμα φαίνεται πάντως ότι τα υποψήφια βέλτιστα είναι στις κορυφές του σχήματος, δηλαδή (i) x=y=0 (ii) x=8/7 y=6/7 (iii) x=2 y=0 και (iv) x=0 y=1.

Θέτουμε $\mathcal{L} = x^2+3y^2 + \mu_1[2-x-y] + \mu_2[8-x-8y]+ \mu_3x + \mu_4y$ με $\mu_1, \mu_2, \mu_3, \mu_4 \geq 0$.

Οι συνθήκες βελτιστοποίησης είναι:

(α) $\partial\mathcal{L}/\partial x = 2x-\mu_1-\mu_2+\mu_3=0$ και $\partial\mathcal{L}/\partial y = 6y-\mu_1-8\mu_2+\mu_4=0$

(β) $\mu_1[2-x-y]=0$ $\mu_2[8-x-8y]=0$ $\mu_3x=0$ $\mu_4y=0$

Στο (i) είναι $\mu_1=\mu_2=0$ και από την (α) έπεται ότι $\mu_3=\mu_4=0$ και οι συνθήκες ικανοποιούνται, αν και το σημείο (i) δεν έχει κανένα ενδιαφέρον από πλευράς βελτιστοποίησης.

Στο (ii) είναι $\mu_3=\mu_4=0$ οπότε η (α) γίνεται $16/7=\mu_1+\mu_2$ και $36/7=\mu_1+8\mu_2$. Λύνοντάς τις προκύπτει $\mu_1=92/49$, $\mu_2=20/49$ που είναι αποδεκτές καθώς οι γραμμικοί περιορισμοί ικανοποιούνται ισοτικά.

Στο (iii) είναι $\mu_2 = \mu_3 = 0$ οπότε η (α) γίνεται $4-\mu_1=0$ και $\mu_1=\mu_4$ οπότε είναι $\mu_1=\mu_4=4$, που είναι παραδεκτό.
Στο (iv) είναι $\mu_1 = \mu_4 = 0$ οπότε η (α) δίνει ένα σύστημα εξισώσεων ως προς μ_2 και μ_3 με λύση $\mu_2=\mu_3=3/4$ που επίσης ικανοποιεί τις συνθήκες βέλτιστου.

Από τα 4 σημεία αυτά, η τιμή της αντικειμενικής στο (i) είναι 0, στο (ii) είναι 3.510 , στο (iii) είναι 4, και τέλος στο (iv) είναι 3. Άρα το καθολικό βέλτιστο είναι το σημείο (iii). Τα παραπάνω μας δείχνουν ότι η ικανοποίηση των συνθηκών Kuhn Tucker είναι αναγκαία αλλά όχι ικανή για συνολικό βέλτιστο. Κάτω από συνθήκες κυρτότητας είναι εύκολο να αποδειχθεί ότι οι συνθήκες είναι και ικανές, ενώ αν υπάρξουν πρόσθετες συνθήκες δεύτερης τάξης τότε ένα οποιοδήποτε σημείο που ικανοποιεί τις συνθήκες είναι τοπικό αλλά όχι ολικό βέλτιστο. Το ίδιο θέμα εμφανίζεται και στο επόμενο παράδειγμα.

□

<u>Παράδειγμα 9</u>: (Οι συνθήκες K-T είναι αναγκαίες αλλά όχι ικανές)
Έστω το πρόβλημα **max x+y με περιορισμό $x^2+y^2 \geq 1$.**
Θέτουμε $\mathcal{L} = x+y+\mu[x^2+y^2-1]$ με $\mu\geq 0$ και οι συνθήκες είναι:

(α) $\partial \mathcal{L}/\partial x = 1+2\mu x=0$ και $\partial \mathcal{L}/\partial y = 1+2\mu y=0$

(β) $\mu(x^2+y^2-1) = 0$

Για να ικανοποιούνται οι K-T θα πρέπει να είναι $\mu>0$, και από το (α) $x = y = -1/(2\mu)$ που είναι αρνητικά. Εφόσον $\mu>0$, πρέπει ο περιορισμός να είναι αποτελεσματικός δηλαδή $1= x^2+y^2$ και άρα $x = y = -1/\sqrt{2}$. και $\mu=1/\sqrt{2}$. Όμως είναι τελείως προφανές ότι το σημείο $(-1/\sqrt{2}, -1/\sqrt{2})$ δεν είναι ούτε τοπικό ούτε και φυσικά ολικό βέλτιστο. Μάλιστα δεν υπάρχει ολικό βέλτιστο καθώς ο περιορισμός δεν αποκλείει απεριόριστη αύξηση των x,y και έτσι δίνουν στην αντικειμενικά απείρως μεγάλες τιμές. Επίσης, μετακινούμενοι κατά την περιφέρεια του μοναδιαίου κύκλου κοντά στο σημείο $(-1/\sqrt{2}, -1/\sqrt{2})$ επιτυγχάνουμε αύξηση της αντικειμενικής, και άρα δεν έχουμε ούτε καν τοπικό βέλτιστο.

□

<u>Παράδειγμα 10α:</u> (Οι συνθήκες K-T δεν είναι καν αναγκαίες σε ακραίες περιπτώσεις)
Έστω το πρόβλημα **max x-y με περιορισμούς $y\geq 0$ και $-x^3-y\geq 0$.**
Το x δεν μπορεί να πάρει θετικές τιμές (γιατί;) όποτε το προφανές βέλτιστο είναι στο x=y=0. Για το πρόβλημα αυτό οι συνθήκες K-T είναι $\mathcal{L} = x - y + \lambda y + \mu[-x^3-y]$ και $\partial\mathcal{L}/\partial x = 1-3\mu x^2 = 0$ και $\partial\mathcal{L}/\partial y = -1+(\lambda-\mu) = 0$. Προφανώς η πρώτη ισότητα δεν ικανοποιείται στο x=0, και άρα παρόλο που το σημείο (0,0) είναι βέλτιστο ΔΕΝ ικανοποιεί τις υποτιθέμενες αναγκαίες συνθήκες K-T.

Το πρόβλημα αυτό παρατηρείται διότι δεν ισχύει η επιπλέον "**ιδιότητα των περιορισμών**" που αναφέρθηκε στην θεωρία. Η ιδιότητα αυτή περίπου απαιτεί να είναι γραμμικά ανεξάρτητα τα διανύσματα κλίσης των αποτελεσματικών περιορισμών, και τότε όντως οι συνθήκες K-T είναι αναγκαίες. Αυτό φαίνεται και στο παράδειγμα 10β, όπου μία απειροελάχιστη αλλαγή στους περιορισμούς οδηγεί σε βέλτιστο όπου ικανοποιούνται οι συνθήκες K-T, καθώς τα σχετικά διανύσματα κλίσεως είναι ανεξάρτητα.

□

<u>Παράδειγμα 10β:</u> (Οι συνθήκες K-T είναι αναγκαίες αν ισχύει η ιδιότητα των περιορισμών*)*

Έστω το πρόβλημα **max x-y με περιορισμούς y≥εx και –x^3-y≥0**, με ε ένα πολύ μικρό θετικό αριθμό (οπότε ο πρώτος περιορισμός αντιστοιχεί περίπου με τον y≥0 που είχαμε στο προηγούμενο παράδειγμα).
Πάλι το x δεν μπορεί να πάρει θετικές τιμές (γιατί;) και το βέλτιστο είναι πάλι στο x=y=0. Για το πρόβλημα αυτό είναι $\mathcal{L}$ = x - y + λ[y-εx] + μ[–x^3-y] και οι συνθήκες K-T είναι:
$\partial\mathcal{L}/\partial x$ = 1-le-3mx^2= 0 και $\partial\mathcal{L}/\partial y$ = -1+(l-m) = 0.
Στο x=0 οι συνθήκες ικανοποιούνται με λ=1/ε και μ=-1+1/ε. Τα διανύσματα βαθμίδας στο πρόβλημα αυτό είναι (ε,1) και (0,-1) που είναι γραμμικά ανεξάρτητα. Αντίθετα, στο παράδειγμα 10α τα διανύσματα βαθμίδος είναι (0,1) και (0,-1) που είναι εξαρτημένα.

□

Μία πολύ καλή (αλλά προχωρημένη) παρουσίαση των ιδιοτήτων περιορισμών δίνεται στο βιβλίο του P. Varaiya Notes on Optimization που είναι διαθέσιμο στο διαδίκτυο (Κεφ. V.1).

2.2.3.4 Διαδικασίες Βελτιστοποίησης με αναζήτηση

Η γεωμετρική ερμηνεία των συνθηκών για βέλτιστο μας οδηγεί στον εξής απλό αλγόριθμο αναζήτησης:

0. Αρχίζουμε με αυθαίρετο x_0
1. Αν δεν είναι βέλτιστο, βρίσκουμε διάνυσμα $\vec{d}$ τέτοιο ώστε $\nabla f(x_0)\cdot\vec{d}>0$ και $\nabla g_j(x_0)\cdot\vec{d}>0$ για τους περιορισμούς που ικανοποιούνται ισοτικά, $g_j(x_0)=0$.
   Αυτό απαιτεί βέβαια λύση προβλήματος εξίσωσης για θετικές ρίζες, δηλαδή Γραμμικού Προγραμματισμού.
2. Εξετάζουμε το μονοδιάστατο πρόβλημα (με μία μεταβλητή λ)
   $\max\limits_{\lambda\geq 0} f(x_0+\lambda\vec{d})$ με λ τέτοιο ώστε $g_j(x_0+\lambda\vec{d})>0$.
3. Η λύση του μας δίνει ένα λ^*και ένα νέο υποψήφιο βέλτιστο $x_1 = x_0+\lambda^*\vec{d}$ που είναι καλύτερο από το προηγούμενο x$_o$.
4. Αν το νέο σημείο δεν είναι «ικανοποιητικό» επαναλαμβάνουμε το βήμα 1.

$|\lambda^*|<$

Προφανώς στην βιβλιογραφία έχουν αναπτυχθεί πολλοί αλγόριθμοι επίλυσης των προβλημάτων αυτών με αναζήτηση

Παράδειγμα αλγορίθμου αναζήτησης:
Στο πρόβλημα: $\max\ x+y$

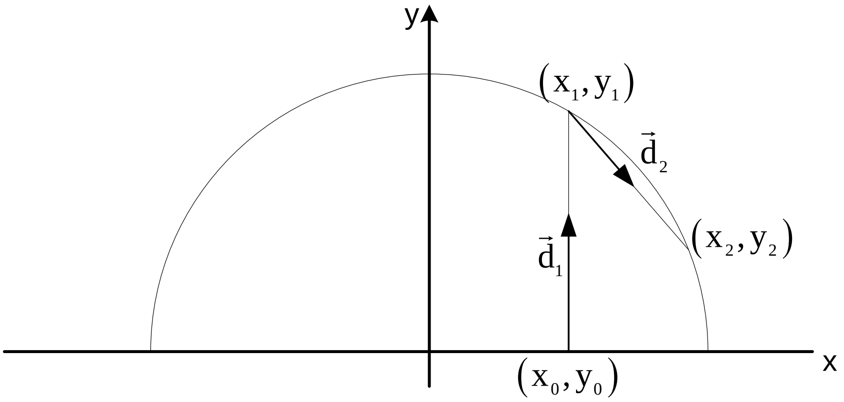
$$g_1 = 1-x^2-y^2 \geq 0$$
$$g_2 = y \geq 0$$

Αρχίζουμε με $(x_0, y_0) = (1/2, 0)$, όπου $g_1 > 0,\ g_2 = 0$.

Βρίσκουμε ένα $\vec{d}$ ώστε $\nabla f \cdot \vec{d} = (1,1)\cdot \vec{d} > 0$ και $\nabla g_2 \cdot \vec{d} = (0,1)\cdot \vec{d} > 0$. Διαλέγουμε το $\vec{d}_1 = (0,1)$ και εξετάζουμε τα σημεία $(1/2, \lambda)$ $\lambda \geq 0$. Το καλύτερο λ είναι $\sqrt{3}/2$ (γιατί;)

Θέτουμε $(x_1, y_1) = \left(1/2, \sqrt{3}/2\right)$ που πάλι δεν είναι βέλτιστο (γιατί;)

Μια νέα «καλή» διεύθυνση μεταβολής $\vec{d}$ είναι η $\vec{d}_2 = (3/2, -1)$ που μας οδηγεί σε νέο υποψήφιο βέλτιστο κ.ο.κ.



Εντοπίστε το (x_2, y_2). Είναι βέλτιστο; Προχωρήστε το αλγόριθμο για μερικά βήματα. □

Άσκηση-ερώτηση:

Περιγράψτε ένα αλγοριθμικό σχήμα που να στηρίζεται στην αλγεβρική απόδειξη των συνθηκών ΚΤ. Συγκεκριμένα έστω ότι λύσαμε τη μορφή Lagrange με τα ε^2 αλλά κάποιο μ δεν είναι μη αρνητικό. Πώς μεταβαίνουμε σε ένα πρόβλημα που έχει «καλύτερη» λύση;

2.2.3.5 Επιπλέον παραδείγματα

Παράδειγμα 1: Έστω το πρόβλημα:

$$\max\ x^2 + y^2 + 3z^2$$

με περιορισμούς:

$$x + y + 3z \leq 10$$

$$2x + y + z \leq 6$$

$$x, y, z \geq 0$$

Εισάγοντάς το στον Solver μάς προτείνεται από το λογισμικό η λύση $x = y = 0,\ z = 10/3$.

Επιβεβαιώστε ότι στο σημείο αυτό ισχύουν όλες οι συνθήκες βελτίστου Κ-Τ.

Έχουμε:

$$L = x^2 + y^2 + 3z^2 + \mu_1(10 - x - y - 3z) + \mu_2(6 - 2x - y - z) + \mu_3 x + \mu_4 y + \mu_5 z$$

Στο προτεινόμενο σημείο είναι $\mu_2 = \mu_5 = 0$ (γιατί;)

Οι συνθήκες παραγώγων είναι:

$$\frac{\partial L}{\partial x} = 2x - \mu_1 - 2\mu_2 + \mu_3 = 0$$

$$\frac{\partial L}{\partial y} = 2y - \mu_1 - \mu_2 + \mu_4 = 0$$

$$\frac{\partial L}{\partial z} = 6z - 3\mu_1 - \mu_2 + \mu_5 = 0$$

Αντικαθιστώντας τις τιμές των x, y, z, μ_2, μ_5 οι σχέσεις αυτές γίνονται:

$$-\mu_1 + \mu_3 = 0$$

$$-\mu_1 + \mu_4 = 0$$

$$20 - 3\mu_1 = 0$$

Άρα $\mu_1 = 20/3 = \mu_3 = \mu_4 > 0$ που είναι παραδεκτές τιμές. Άρα οι συνθήκες επιβεβαιώνονται.

□

<u>Παράδειγμα 2:</u> Έστω υποψήφια λύση $x = 3,\ y = z = 0$ για το ίδιο πρόβλημα. Εξετάστε αν ισχύουν και σ' αυτήν οι συνθήκες βελτίστου.

Είναι $\mu_1 = 0 = \mu_3$ λόγω συμπληρωματικότητας. Οι συνθήκες παραγώγων γίνονται:

$$\frac{\partial L}{\partial x} = 6 - 2\mu_2 = 0 \quad \Rightarrow \mu_2 = 3$$

$$\frac{\partial L}{\partial y} = 0 - \mu_2 + \mu_4 = 0 \ \Rightarrow \mu_4 = \mu_2 = 3$$

$$\frac{\partial L}{\partial z} = 0 - \mu_2 + \mu_5 = 0 \ \Rightarrow \mu_5 = \mu_2 = 3$$

Άρα αποδεκτή λύση είναι και αυτή, αφού οι συνθήκες επιβεβαιώνονται και εδώ!

□

<u>Σχόλια:</u>

- Και οι δύο λύσεις $(3,0,0)$ και $(0,0,10/3)$ είναι τοπικά βέλτιστα. Προφανώς η δεύτερη λύση δίνει μεγαλύτερη τιμή στην αντικειμενική συνάρτηση.
- Το λογισμικό Solver ενδέχεται να καταλήξει είτε στη μία λύση είτε στην άλλη, ανάλογα με τις αρχικές συνθήκες. Επιβεβαιώστε το. Ποιά είναι η ολική βέλτιστη λύση;
- Προβλήματα σαν το προηγούμενο είναι γνωστά ως «Τετραγωνικός Προγραμματισμός» (Quadratic Programming). Για τέτοια προβλήματα έχουν αναπτυχθεί ειδικοί αλγόριθμοι, παρόμοιοι με εκείνους του Γραμμικού Προγραμματισμού.

<u>Παράδειγμα 3:</u> Έστω η υποψήφια λύση $x=1, y=z=0$ για το ίδιο πρόβλημα. Εξετάστε αν ισχύουν και για αυτήν οι συνθήκες βελτίστου.

Είναι $\mu_1=\mu_2=\mu_3=0$ οπότε οι συνθήκες παραγώγων δίνουν:

$2+\mu_3=0 \quad \Rightarrow \quad 2=0$ Αδύνατη!

$\mu_4=0$

$\mu_5=0$

Άρα δεν ικανοποιούνται οι συνθήκες!
Τι ισχύει για την λύση $x=y=z=0$;

□