Πολυτεχνική Σχολή ΑΠΘ

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τεχνικές Βελτιστοποίησης, 7ο εξάμηνο

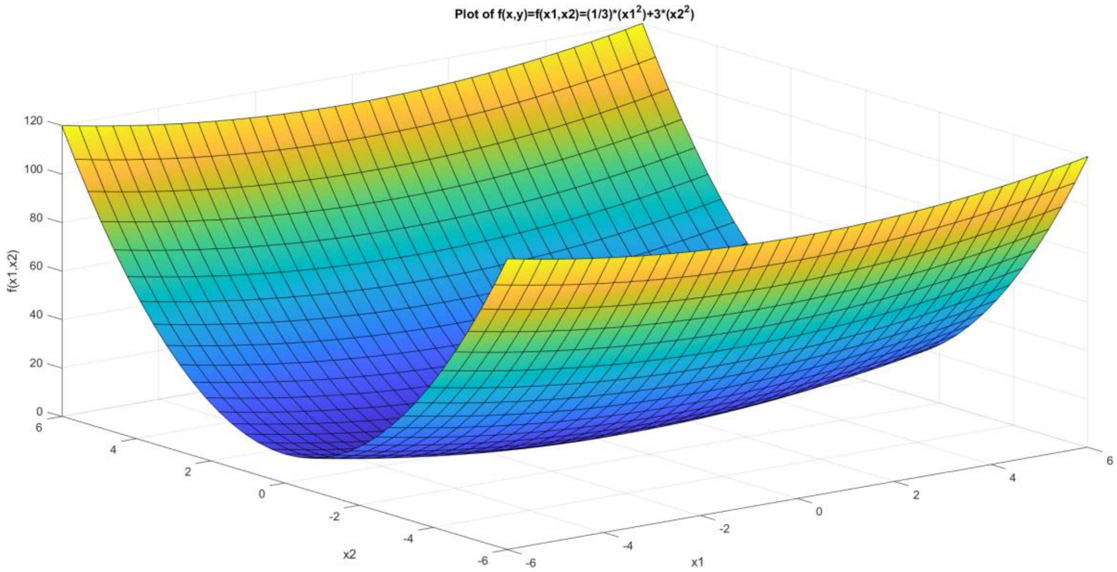
Δεκέμβριος 2024

Ινώ Χαριτίδη, 10043

$3^η$ Εργαστηριακή Άσκηση: Μέθοδος Μέγιστης Καθόδου με Προβολή

Θεωρήστε τη συνάρτηση:

f : $\mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x) = \frac{1}{3} * x_1^2 + 3 * x_2^2$, $x = [x_1, x_2]^T$



$\nabla f(x) = \left[\frac{2}{3} * x_1 \text{ , } 6 * x_2\right]$ $\qquad \nabla^2 f(x) = \begin{matrix} \frac{2}{3} & 0 \\ 0 & 6 \end{matrix}$

Ο Εσσιανός της $f(x)$ είναι θετικά ορισμένος, αφού $\frac{2}{3} > 0$ και η ορίζουσά του ισούται με $4>0$. Επομένως η $f(x)$ είναι γνήσια κυρτή.

Παρατηρώ ότι το gradient της $f(x)$ μηδενίζεται στο $(0,0)$. Επομένως το $(0,0)$ είναι κρίσιμο σημείο της, και αφού η $f(x)$ είναι γνήσια κυρτή, το $(0,0)$ αποτελεί ολικό της ελάχιστο.

Επομένως η δοθείσα συνάρτηση παρουσιάζει ολικό ελάχιστο ίσο με 0, στο σημείο (0,0).

<u>Θέμα 1</u>: Να χρησιμοποιηθεί η Μέθοδος Μέγιστης Καθόδου (προηγούμενη εργασία) με ακρίβεια ε =0.001 και βήμα i) γ_k= 0.1, ii) γ_k = 0.3, iii) γ_k = 3, iv) γ_k = 5 και οποιοδήποτε αρχικό σημείο εκκίνησης διαφορετικό του (0,0). Τι παρατηρείτε; Να αποδειχθούν τα αποτελέσματα αυτά με μαθηματική αυστηρότητα.

Όπως περιγράφει και η άσκηση 5.5.3 του βιβλίου, σελίδα 170:

Για να συγκλίνει ο αλγόριθμος στο σημείο ελαχίστου (0,0), εκκινούμενος από οποιοδήποτε άλλο σημείο, πρέπει, όσο προχωράει , απόλυτες τιμές των σημείων να τείνουν στο 0, δηλαδή να μειώνονται. Άρα πρέπει $x_{k+1} < x_k$ για κάθε k.

Ο αλγόριθμος της μέγιστης καθόθου είναι:

$$x_{k+1} = x_k - \gamma_k * \nabla f(x_k)$$

$$x_{k+1} = x_k - \gamma_k * \left[\frac{2}{3}, 6\right] * x_k$$

$$x_{k+1} = (I - \gamma_k * \left[\frac{2}{3}, 6\right]) * x_k$$

Για την κάθε μεταβλητή ξεχωριστά ισχύει:

$$x_{1\ k+1} = (1 - \gamma_k * \frac{2}{3}) * x_{1k}$$

$$x_{2\ k+1} = (1 - \gamma_k * 6) * x_{2k}$$

Για να συγκλίνει η μέθοδος στο (0,0), πρέπει :

$$|I - \gamma_k * \left[\tfrac{2}{3}, 6\right]| < I$$

$$-I < I - \gamma_k * \left[\frac{2}{3}, 6\right] < I$$

$$0 < \gamma_k * \left[\frac{2}{3}, 6\right] < 2 * I$$

Άρα πρέπει $0 < \gamma_k < 3$ για το x_{k1} και $0 < \gamma_k < \frac{1}{3}$ για το x_{k2}.

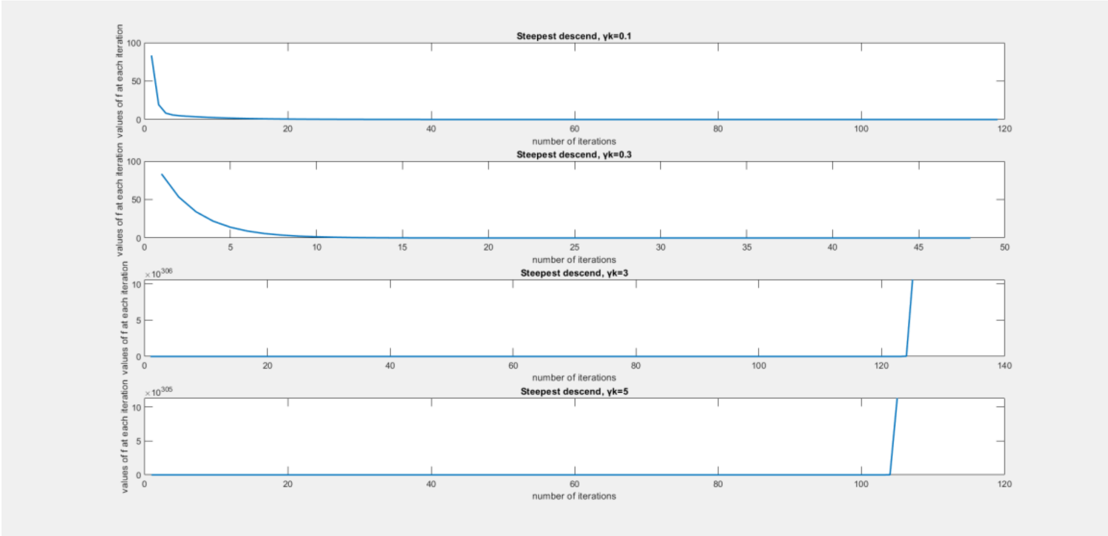
Τελικά για να είναι επιτυχής η σύγκλιση πρέπει $\gamma_k < \frac{1}{3}$.

Αυτό ισχύει για γ_k= 0.1 και γ_k = 0.3, αλλά όχι για γ_k = 3 και γ_k = 5.

Ρυθμοί σύγκλισης των μεταβλητών:

Αναδρομικά προκύπτει $x_{1\ k} = \left(1 - \gamma_k * \frac{2}{3}\right)^k * x_{1\ 0}$ και $x_{2\ k} = (1 - \gamma_k * 6)^k * x_{1\ 0}$

Επιλέγω ως αρχικό σημείο το (5,-5).



i) γ_k= 0.1
Ο αλγόριθμος τερματίζει μετά από 119 επαναλήψεις, και επιστρέφει ως αποτέλεσμα το σημείο (0.0015, -0.0000).
$f(0.0015, 0) = \ 7.5 * 10^{-07}$, ένας πολύ μικρός αριθμός, ικανοποιητικά κοντά στην πραγματική τιμή ελαχίστου 0.
Η σύγκλιση είναι σχετικά αργή, διότι το βήμα γ_k είναι μικρό.

$x_{1\,k} = 5 * \left(1 - 0.1 * \frac{2}{3}\right)^k = 5 * (0.9333)^k$ και

$x_{2\,k} = -5 * (1 - 0.1 * 6)^k = -5 * (0.4)^k$

Αφού $0.4 < 0.9333$, το x_2 συγκλίνει πιο γρήγορα στο 0 από το x_1

ii) γ_k= 0.3
Ο αλγόριθμος τερματίζει μετά από 48 επαναλήψεις, και επιστρέφει ως αποτέλεσμα το σημείο (0.0001394, 0.0001394).
$f(0.0001394, 0.0001394) = \ 6.4775 * 10^{-8}$, ένας πολύ μικρός αριθμός, ικανοποιητικά κοντά στην πραγματική τιμή ελαχίστου 0.
Παρατηρώ ότι με αυτό το (μεγαλύτερο) βήμα, ο αλγόριθμος συγκλίνει κάνοντας λιγότερες επαναλήψεις, και με ελαφρώς καλύτερη ακρίβεια απ' ότι στην πρώτη περίπτωση.

$x_{1\,k} = 5 * \left(1 - 0.3 * \frac{2}{3}\right)^k = 5 * (0.8)^k$ και

$x_{2\,k} = -5 * (1 - 0.3 * 6)^k = -5 * (-0.8)^k$

Τα x_1 , x_2 συγκλίνουν με τον ίδιο ρυθμό στο 0, με διαφορά ότι το x_2 αλλάζει πρόσημο σε κάθε επανάληψη.

iii) γ_k= 3
Ο αλγόριθμος γίνεται αριθμητικά ασταθής. Τερματίζει μετά από 252 επαναλήψεις, και επιστρέφει ως αποτέλεσμα (-5, NaN).

Αυτό συμβαίνει επειδή το βήμα είναι πολύ μεγάλο, και προκαλεί απόκλιση.

$x_{1\,k} = 5 * \left(1 - 3 * \frac{2}{3}\right)^k = 5 * (-1)^k$

$x_{2\,k} = -5 * (1 - 3 * 6)^k = -5 * (-17)^k$

Το x_1 μένει σταθερό κατ' απόλυτη τιμή, και αλλάζει πρόσημο σε κάθε επανάληψη, ενώ το x_2 αυξάνει ραγδαία σε τιμή (και αλλάζει πρόσημο σε κάθε επανάληψη) εώς ότου το Matlab δεν έχει πλέον τη δυνατότητα να την αναπαραστήσει. Στις δύο τελευταίες θέσεις του πίνακα result, τον οποίο επιστρέφει η μέθοδος, εμφανίζονται οι εκφράσεις Inf και NaN.

iv) γ_k= 5

Ο αλγόριθμος γίνεται αριθμητικά ασταθής, για τον ίδιο λόγο με την περίπτωση (iii). Τερματίζει μετά από 213 επαναλήψεις, και επιστρέφει ως αποτέλεσμα (0, NaN).

$$x_{1\,k} = 5 * \left(1 - 5 * \frac{2}{3}\right)^k = 5 * \left(-\frac{7}{3}\right)^k$$
$$x_{2\,k} = -5 * (1 - 5 * 6)^k = -5 * (-29)^k$$

Τα x_1 , x_2 αποκλίνουν, αλλάζοντας πρόσημο σε κάθε επανάληψη. Το x_2 αυξάνει κατ απόλυτη τιμή γρηγορότερα από το x_1. Όμοια με την περίπτωση (iii), στις δύο τελευταίες θέσεις του πίνακα result εμφανίζονται οι εκφράσεις Inf και NaN.

Θεωρήστε τώρα τους περιορισμούς: $-10 \le x_1 \le 5$ και $-8 \le x_2 \le 12$.

<u>Θέμα 2</u>*:* Να χρησιμοποιηθεί η Μέθοδος Μέγιστης Καθόδου με Προβολή, με $s_k = 5$, $\gamma_k = 0.5$, σημείο εκκίνησης το (5, −5) και ακρίβεια ε = 0.01. Τι παρατηρείτε σε σχέση με το Θέμα 1;

Καταρχήν πραγματοποιώ μαθηματική ανάλυση, για να γνωρίζω πότε μπορώ να αναμένω σύγκλιση και πότε όχι.

Ο αλγόριθμος της μέγιστης καθόθου με προβολή είναι:

$$x_{k+1} = x_k - \gamma_k * (\overline{x_k} - x_k)$$

όπου $\overline{x_k} = \Pr_X\{x_k - s_k * \nabla f(x_k)\} = \Pr_X\{x_k - s_k * \left[\frac{2}{3}, 6\right] * x_k\} = \Pr_X\{\left(I - s_k * \left[\frac{2}{3}, 6\right]\right) * x_k\}$

$$\overline{x_{1k}} = -10, \qquad \alpha\nu\ x_{1k} * \left(1 - \frac{2}{3} * s_k\right) \le -10$$

$$\overline{x_{1k}} = x_{1k} * \left(1 - \frac{2}{3} * s_k\right),\ \alpha\nu - 10 < x_{1k} * \left(1 - \frac{2}{3} * s_k\right) < 5$$

$$\overline{x_{1k}} = 5, \qquad \alpha\nu\ x_{1k} * \left(1 - \frac{2}{3} * s_k\right) \ge 5$$

$$\overline{x_{2k}} = -8, \qquad \alpha\nu\ x_{2k} * (1 - 6 * s_k) \le -8$$

$$\overline{x_{2k}} = x_{2k} * (1 - 6 * s_k), \qquad \alpha\nu\ -8 < x_{1k} * (1 - 6 * s_k) < 12$$

$$\overline{x_{2k}} = 12, \qquad \alpha\nu\ x_{2k} * (1 - 6 * s_k) \ge 12$$

$$x_{1\ k+1} = x_{1\ k} + \gamma_{1k} * (x_{1k} * \left(1 - \frac{2}{3} * s_{1k}\right) - x_{1k})$$

$$x_{1\ k+1} = x_{1\ k} + \gamma_{1k} * (-\frac{2}{3} * s_{1k} * x_{1\ k})$$

$$x_{1\ k+1} = (1 - \gamma_{1k} * \frac{2}{3} * s_{1k}) * x_{1\ k}$$

Πρέπει:

$$\left|1 - \gamma_{1k} * \frac{2}{3} * s_{1k}\right| < 1$$

$$-1 < 1 - \gamma_{1k} * \frac{2}{3} * s_{1k} < 1$$

$$0 < \gamma_{1k} * \frac{2}{3} * s_{1k} < 2$$

$$0 < \gamma_{1k} * s_{1k} < 3$$

Όμοια:

$$x_{2\,k+1} = x_{2\,k} + \gamma_{2k} * (x_{2k} * (1 - 6 * s_{2k}) - x_{2k})$$
$$x_{2\,k+1} = x_{2\,k} + \gamma_{2k} * (-6 * s_{2k} * x_{2\,k})$$
$$x_{2\,k+1} = (1 - \gamma_{2k} * 6 * s_{2k}) * x_{2\,k}$$

Πρέπει:

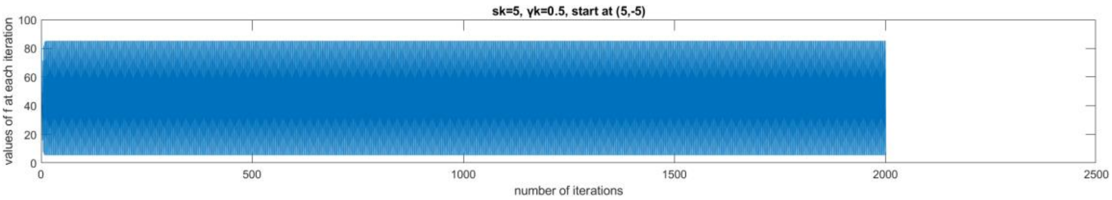
$$|1 - \gamma_{2k} * 6 * s_{2k}| < 1$$
$$-1 < 1 - \gamma_{2k} * 6 * s_{2k} < 1$$
$$0 < \gamma_{2k} * 6 * s_{2k} < 2$$
$$0 < \gamma_{2k} * s_{2k} < \frac{1}{3}$$

Ο αλγόριθμος ξεκινά από το εφικτό σημείο (5,-5), με $\gamma_k = 0.5$ και $s_k = 5$.

$$\gamma_k * s_k = 0.5 * 5 = 2.5$$

Επομένως η σύγκλιση είναι αναμενόμενη για τη x_1, αλλά όχι για τη x_2.

Πράγματι:



Η μεταβλητή x_1 τείνει στο 0 με την πάροδο της κάθε επανάληψης, ενώ η μεταβλητή x_2 από την επανάληψη 19 και μετά ταλαντώνεται μεταξύ των τιμών 5.3333 και -1.3333. Ο αλγόριθμος δε συγκλίνει, και για να αποφύγω τον εγκλωβισμό σε ατέρμονο βρόγχο θέτω ανώτατο όριο επαναλήψεων ίσο με 2000.

Σε αντίθεση με το Θέμα 1, η ακατάλληλη επιλογή βημάτων γ_k, s_k δεν οδηγεί τη μέθοδο σε αριθμητική αστάθεια. Αυτό οφείλεται στην ύπαρξη περιορισμών για τις μεταβλητές, και συγκεκριμένα στον περιορισμό $-8 \le x_2 \le 12$.

Για $x_{2k} = 5.3333$, $x_{2k} * (1 - 6 * s_k) = 5.3333 * (1 - 6 * 5) = -154.6657 < -8$, άρα $\overline{x_{2k}} = -8$.

Για $x_{2k} = -1.3333$, $x_{2k} * (1 - 6 * s_k) = -1.3333 * (1 - 6 * 5) = 38{,}6657 > 12$, άρα $\overline{x_{2k}} = 12$.

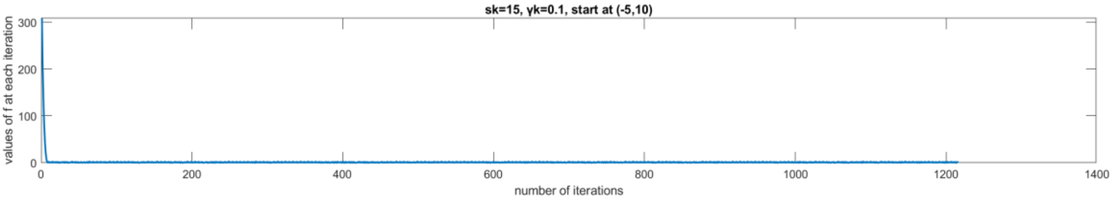
Επομένως ο αλγόριθμος δεν μπορεί να αποκλίνει.

Θέμα 3: Να χρησιμοποιηθεί η Μέθοδος Μέγιστης Καθόδου με Προβολή, με $s_k = 15,\ \gamma_k = 0.1$, σημείο εκκίνησης το (-5, 10) και ακρίβεια ε = 0.01. Τι παρατηρείτε σε σχέση με τα Θέματα 1 και 2; Προτείνετε έναν απλό πρακτικό τρόπο ώστε η μέθοδος να συγκλίνει στο ελάχιστο.

Ο αλγόριθμος ξεκινά από το εφικτό σημείο (-5,10), με $\gamma_k = 0.1$ και $s_k = 15$.

$$\gamma_k * s_k = 0.1 * 15 = 1.5$$

Επομένως η σύγκλιση είναι αναμενόμενη για τη x_1, αλλά όχι για τη x_2.



Σε αντίθεση με το Θέμα 2, αλγόριθμος συγκλίνει, αλλά μετά από υπέρογκο αριθμό επαναλήψεων (1216 επαναλήψεις). Η μεταβλητή x_1 φτάνει στο 0 στην 9^{η} επανάληψη, ενώ η μεταβλητή x_2 πραγματοποιεί μια μεγάλη σε μήκος ταλάντωση, μέχρι που τελικά καταλήγει στην τιμή 0.000916.

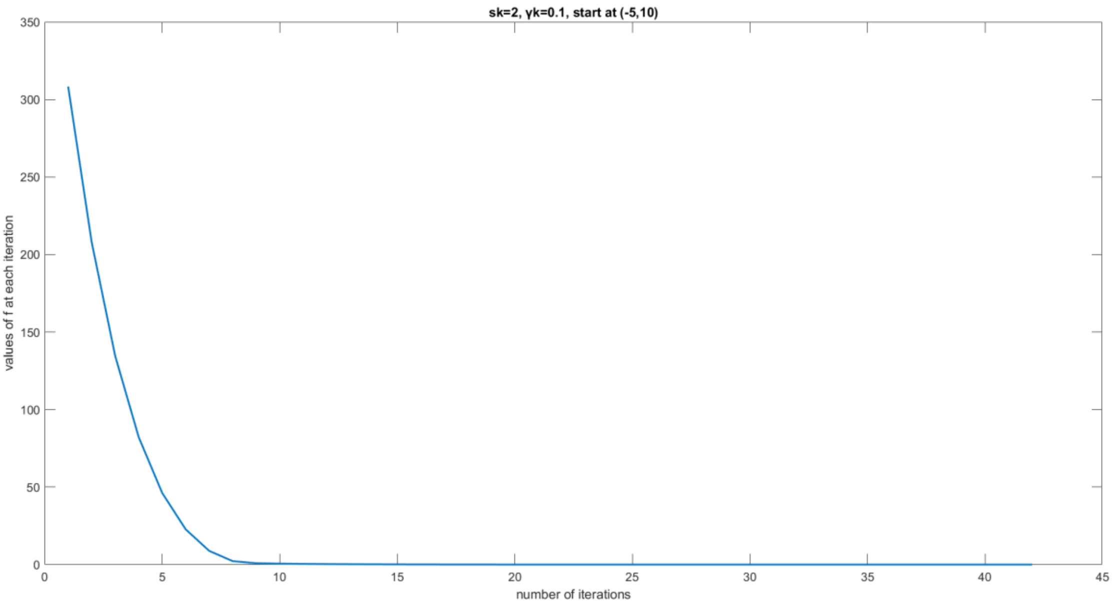
$$f(0, 0.000916) = \ \ 2.5172 * 10^{-6}$$

Η ακρίβεια του αποτελέσματος είναι καλή, αλλά η ταχύτητα σύγκλισης όχι.

Ένας τρόπος να μειωθούν οι επαναλήψεις είναι η επιλογή μικρότερου βήματος s_{2k}, έτσι ώστε να ισχύει

$$0 < \gamma_{2k} * s_{2k} < \frac{1}{3}$$

Για παράδειγμα, με $s_k = 2$, ο αλγόριθμος πραγματοποιεί 42 επαναλήψεις και επιστρέφει ως αποτέλεσμα το σημείο (-0.0142, 0).



Θέμα 4: Να χρησιμοποιηθεί η Μέθοδος Μέγιστης Καθόδου με Προβολή, με $s_k = 0.1$, $\gamma_k = 0.2$, σημείο εκκίνησης το (8, −10) και ακρίβεια ε = 0.01. Σε αυτή την περίπτωση, έχουμε εκ των προτέρων κάποια πληροφορία σχετικά με την σύγκλιση του αλγορίθμου; Να γίνει η εκτέλεση του αλγορίθμου. Τι παρατηρείτε;

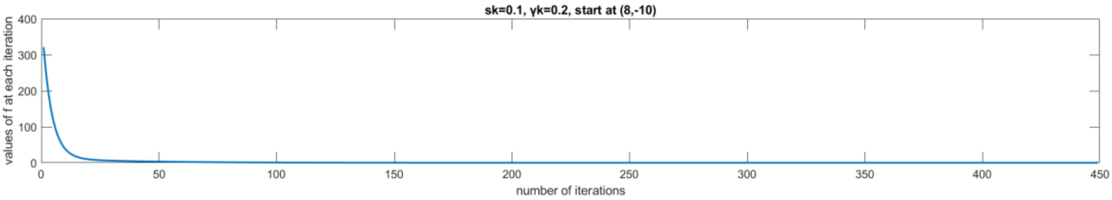
Επειδή

$$\gamma_k * s_k = 0.1 * 0.2 = 0.02$$

η σύγκλιση είναι αναμενόμενη και για x_1 και για x_2.

Βέβαια, επειδή οι τιμές των γ_k , s_k είναι αρκετά μικρότερες από όσο είναι απαραίτητο για να ισχύουν οι επιθυμητές ανισότητες, είναι αναμενόμενο ο αλγόριθμος να πραγματοποιήσει αρκετές επαναλήψεις.

Παρατηρώ επίσης ότι το αρχικό σημείο (8,-10) δεν είναι εφικτό.



16 επαναλήψεις αργότερα, ο αλγόριθμος έχει εισέλθει στο εφικτό σύνολο.

Πραγματοποιεί 449 επαναλήψεις, και επιστρέφει μια αρκούντως ακριβή απάντηση (0.0149,0).