Μέθοδος Μέγιστης Καθόδου με Προβολή

Θεολόγης Γεώργιος ΑΕΜ:10413
email: gtheolog@ece.auth.gr

Δεκέμβριος 2023

1 Εισαγωγή

Το αντικείμενο της παρούσας εργασίας είναι η υλοποίηση του αλγορίθμου της μέγιστης καθόδου με προβολή , η οποία θα αξιοποιηθεί για την εύρεση του ελαχίστου συνάρτησης δύο μεταβλητών $f : \mathbb{R}^2 \to \mathbb{R}$ εντός περιορισμένου συνόλου στο $\mathbb{R}^2$. Συγκεκριμένα θα γίνει εφαρμογή στην συνάρτηση:

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

Το πρόβλημα βελτιστοποίησης αφορά την εύρεση του ελαχίστου της συνάρτησης f στο σύνολο X:

$$-10 \leq x_1 \leq 5, -8 \leq x_2 \leq 12$$

2 Θέμα 1

Για αρχή θα ελαχιστοποιήσουμε την συνάρτηση f(x) **χωρίς περιορισμούς** χρησιμοποιώντας την μέθοδο μέγιστης καθόδου της προηγούμενης εργασίας. Πριν την εφαρμογή της μεθόδου θα πρέπει να ελέγξουμε άμα η συνάρτηση f(x) είναι γνησίως κυρτή έτσι ώστε να έχει μοναδικό ελάχιστο και η μέθοδος μέγιστης καθόδου να οδηγεί σε αυτό καθώς και τι σταθερό βήμα γ_k μπορούμε να χρησιμοποιήσουμε ώστε η μέθοδος να μπορεί να συγκλίνει.
Εύκολα διαπιστώνουμε ότι η συνάρτηση f(x) είναι γνήσια κυρτή καθώς η δεύτερη παραγωγός της (Εσσιανός πίνακας) είναι θετικά ορισμένη:

$$\nabla^2 f = \begin{bmatrix} \frac{2}{3} & 0 \\ 0 & 6 \end{bmatrix} > 0$$

αφού είναι συμμετρικός , με θετικές ιδιοτιμές οι οποίες είναι θετικές $\lambda_1 = \frac{2}{3}, \lambda_2 = 6$. Συνεπώς, η συνάρτηση f έχει μοναδικό σημείο ελαχίστου το οποίο είναι το $x^* = (0, 0)$
Επίσης για να μπορέσει να συγκλίνει ο αλγόριθμος θα πρέπει το γ_k να είναι τέτοιο ώστε να τηρείται το 1ο κριτήριο καλής λειτουργίας:

$$f(x_{k+1}) < f(x_k)$$

Σε κάθε επανάληψη k όπου $\nabla f(x_k) \neq 0$.
Το x_{k+1} είναι ίσο με:

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

$$x_{k+1} = \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} - \gamma_k \nabla \begin{bmatrix} \frac{2}{3} \cdot x_{1,k} \\ 6 \cdot x_{2,k} \end{bmatrix}$$

$$x_{k+1} = \begin{bmatrix} x_{1,k}(1 - \frac{2}{3}\gamma_k) \\ x_{2,k}(1 - 6\gamma_k) \end{bmatrix} \quad (a)$$

Άρα για να ισχύει το 1ο κριτήριο καλής λειτουργίας πρέπει:

$$f(x_{k+1}) < f(x_k) \Rightarrow \frac{1}{3}x_{1,k+1}^2 + 3x_{2,k+1}^2 < \frac{1}{3}x_{1,k}^2 + 3x_{2,k}^2$$

$$\frac{1}{3}x_{1,k}^2 \cdot (1-\frac{2}{3}\gamma_k^2)^2 + 3x_{2,k}^2 \cdot (1-6\gamma_k)^2 < \frac{1}{3}x_{1,k}^2 + 3x_{2,k}^2$$
$$\frac{1}{3}x_{1,k}^2 \cdot (-\frac{4}{3}\gamma_k + \frac{4}{9}\gamma_k^2) + 3x_{2,k}^2 \cdot (-12\gamma_k + 36\gamma_k^2) < 0$$
$$\frac{1}{3}x_{1,k}^2 \cdot (-\frac{4}{3} + \frac{4}{9}\gamma_k) + 3x_{2,k}^2 \cdot (-12 + 36\gamma_k) < 0 \qquad (1.1)$$

Αφού $\gamma_k > 0$. Προφανώς για να ισχύει η ανίσωση (1.1) για κάθε x_k πρέπει $-\frac{4}{3} + \frac{4}{9}\gamma_k < 0$ και $-12 + 36\gamma_k < 0$ δηλαδή $\gamma_k < 3$ και $\gamma_k < \frac{1}{3}$. Τελικά για να συγκλίνει ο αλγόριθμος σίγουρα από οποιοδήποτε σημείο έναρξης αρκεί να επιλέξουμε σταθερό βήμα:

$$\gamma_k < \frac{1}{3}$$

Καθώς αυτές οι τιμές του γ_k ικανοποιούν το πρώτο κριτήριο λειτουργίας για κάθε σημείο x_k τότε για την ακολουθία τιμών $f(x_k)$ ισχύει το κριτήριο λόγου :

$$lim_{k\to\infty}\left|\frac{f(x_{k+1})}{f(x_k)}\right| < 1$$

και συνεπώς η ακολουθία $f(x_k)$ συγκλίνει στο μηδέν για άπειρες επαναλήψεις και συνεπώς η ακολουθία σημείων x_k συγκλίνει επίσης στο μηδέν. Επίσης, λύνοντας προς την (1.1) προς γ_k έχουμε:

$$\gamma_k < \frac{\frac{4}{9}\cdot x_{1,k}^2 + 36\cdot x_{2,k}^2}{\frac{4}{27}\cdot x_{1,k}^2 + 108\cdot x_{2,k}^2} \qquad (1.2)$$

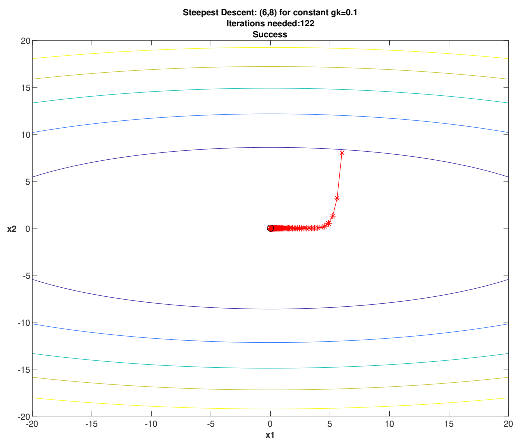
Αυτή η σχέση μας αφήνει να διερευνήσουμε άμα υπάρχουν πιθανά σημεία έναρξης του αλγορίθμου για τα οποία μπορούμε να πάρουμε βήμα τέτοιο ώστε να μην υπακούει στην $\gamma_k < \frac{1}{3}$. Εύκολα βλέπουμε ότι άμα ξεκινήσουμε από τον άξονα του x_1 δηλαδή από σημείο $(x_0, 0)$ ο αλγόριθμος παραμένει πάντα στον άξονα καθώς $\nabla f(x_k) = \begin{bmatrix}\frac{2}{3}x_{1,k}\\ 0\end{bmatrix}$ και συνεπώς από την (1.2) μπορούμε να έχουμε βήμα $\gamma_k < 3$ άμα ξεκινήσει ο αλγόριθμος την αναζήτηση στον άξονα x_1 .

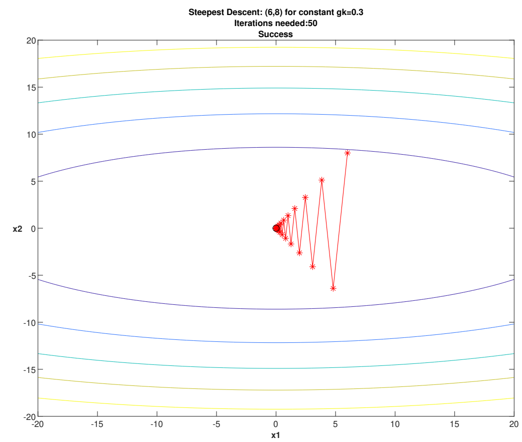
Μάλιστα από την σχέση (α) έχουμε :

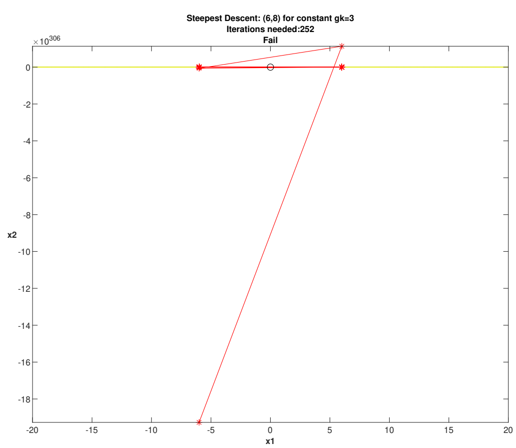
$$x_{k+1} = \begin{bmatrix}x_{1,k}(1-\frac{2}{3}\gamma_k)\\ x_{2,k}(1-6\gamma_k)\end{bmatrix} \quad (a)$$

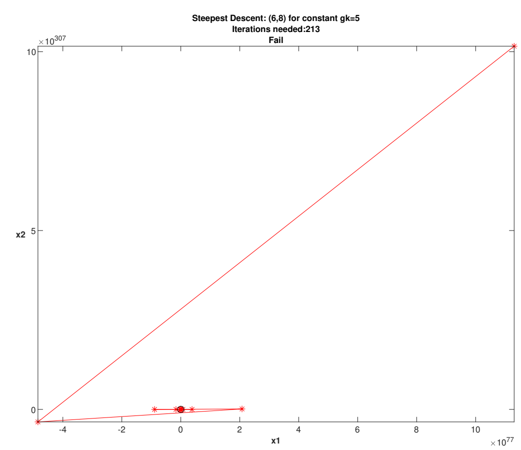
Για να συγκλίνουν αυτές οι ακολουθίες κατά αντιστοιχία με πάνω μέσω του κριτήριου λόγου πρέπει να ισχύει: $|1-\frac{2}{3}\gamma_k| < 1$ άρα $\gamma_k < 3$ για συγκλιση χ1 και $|1-6\gamma_k| < 1$ άρα $\gamma_k < \frac{1}{3}$ για την σύγκλιση του χ2 στο 0 . Άρα για έχουμε σίγουρα ολική σύγκλιση στο μηδέν (0,0) αν και μόνο άν $\gamma_k < \frac{1}{3}$, αφου παίρνουμε το ελάχιστο για να συγκλίνουν και οι δυο.

Έχοντας πραγματοποιήσει την μαθηματική ανάλυση και διερεύνηση τώρα μπορούμε να εφαρμόσουμε την μέθοδο και να παρατηρήσουμε άμα επαληθεύονται τα συμπεράσματα που εξάγαμε προηγουμένως.
Επιλέγουμε ως σημείο εκκίνησης του αλγορίθμου το σημείο $x_0 = (6, 8)$ και εφαρμόζουμε την μέθοδο μέγιστης καθόδου στην f(x) για σταθερά βήματα $i)\gamma_k = 0.1, ii)\gamma_k = 0.3, iii)\gamma_k = 3, iv)\gamma_k = 5$. Τα διαγράμματα της ακολουθίας σημείων x_k που σχηματίζει ο αλγόριθμος στο x_k είναι τα ακόλουθα:


Steepest Descent: (6,8) for constant gk=0.1
Iterations needed:122
Success
x1
x2


Steepest Descent: (6,8) for constant gk=0.3
Iterations needed:50
Success
x1
x2


Steepest Descent: (6,8) for constant gk=3
Iterations needed:252
Fail
×10^306
x1
x2


Steepest Descent: (6,8) for constant gk=5
Iterations needed:213
Fail
×10^307
×10^77
x1
x2

Παρατηρούμε ότι σύγκλιση του αλγορίθμου έχουμε μονάχα για τα $\gamma_k = 0.1$ και $\gamma_k = 0.3$ καθώς αυτά ικανοποιούν την επιθυμητή συνθήκη $\gamma_k < \frac{1}{3}$. Αντίθετα τα $\gamma_k = 3$ και $\gamma_k = 5$ που δεν την ικανοποιούν παρατηρούμε ότι αποκλίνουν και όπως άλλωστε περιμέναμε αυξάνουν διαδοχικά την τιμή της $f(x_k)$ και θεωρητικά την οδηγούν στον απειρισμό αν και πρακτικά ο κώδικας στο matlab μπόρεσε να υπολογίσει εώς τιμές συντεταγμένων της τάξης του 10^{307} και μετά καθώς αρχίζει να υπολογίζει τιμές ως NaN (Not a Number) , η λογική συνθήκη τερματισμού της while κύριας επανάληψης παίρνει τιμή το λογικό μηδέν και συνεπώς τερματίζει ο αλγόριθμος.
Για τις δυο πρώτες περιπτώσεις που έχουμε σύγκλιση στο ελάχιστο παρατηρούμε ότι για βήμα $\gamma_k = 0.3$ η σύγκλιση γίνεται σε λιγότερες επαναλήψεις καθώς το βήμα είναι μεγαλύτερο και πάμε πιο γρήγορα στην γειτονία του ελαχίστου και επίσης για τον ίδιο λόγο είναι αρκετά πιο εμφανής η τροχία τεθλασμένων ευθειών (ζικ-ζακ) την οποία ακολουθεί η ακολουθία των σημείων x_k.

3 Θέμα 2

Έχοντας εξάγει τα συμπεράσματα μας για το πρόβλημα ελαχιστοποιήσης της f(x) χωρίς περιορισμούς , τώρα μπορούμε να αντιμετωπίσουμε το αρχικό πρόβλημα , στο οποίο έχουμε περιορισμούς στο σύνολο που μπορεί να βρίσκονται τα σημεία x_k που υπολογίζει ο αλγόριθμος. Άρα το πρόβλημα είναι το εξής:

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

Όπου το x ανήκει στο σύνολο:

$$-10 \leq x_1 \leq 5, -8 \leq x_2 \leq 12$$

Θα χρησιμοποιήσουμε σε αυτό και στα επόμενα θέματα την μέθοδο μεγίστης καθόδου με προβολή στο σύνολο $X : \{-10 \leq x_1 \leq 5, -8 \leq x_2 \leq 12\}$. Η περιοχή X που ορίζει ο περιορισμός είναι το εσωτερικό μαζί και με το σύνορο ορθογωνίου με άκρα τα (-10,-8),(-10,12),(5,-8),(-5,12) στο $\mathbb{R}^2$,του οποίου το σύνορο θα απεικονίζεται στα plots που ακολουθούν.
Η μέθοδος αυτή δημιουργεί μια εφικτή κατεύθυνση $\bar{x}_k - x_k$ με τα $\bar{x}_k, x_k$ ανήκουν στο X και η κατεύθυνση αυτή είναι της μορφής $-c_k\nabla f(x_k), c_k > 0$. Το επόμενο σημείο της ακολουθίας επιλέγεται το $x_{k+1} = x_k + \gamma_k(\bar{x}_k - x_k), \gamma_k \in (0, 1]$, συνεπώς είναι ένα σημείο πάνω στο ευθύγραμμο τμήμα με άκρα τα σημεία $\bar{x}_k, x_k$. Για να ισχύουν τα παραπάνω πρέπει το $\bar{x}_k$ να επιλέγεται ως :

$$\bar{x}_k = Pr_X\{x_k - s_k\nabla f(x_k)\}, \ \ s_k > 0$$

Άρα η ποσότητα εφικτής κατεύθυνσης $ed_k = x_k - s_k\nabla f(x_k)$ είναι:

$$\begin{bmatrix} ed_{1,k} \\ ed_{2,k} \end{bmatrix} = \begin{bmatrix} x_{1,k}(1 - \frac{2}{3}s_k) \\ x_{2,k}(1 - 6s_k) \end{bmatrix}$$

Άρα ισχύει ότι :

$$\bar{x}_{1,k} = \begin{cases} -10 & \text{if } x_{1,k}(1 - \frac{2}{3}s_k) \leq -10 \\ 5 & \text{if } x_{1,k}(1 - \frac{2}{3}s_k) \geq 5 \\ x_{1,k}(1 - \frac{2}{3}s_k) & \text{otherwise} \end{cases}$$

$$\bar{x}_{2,k} = \begin{cases} -8 & \text{if } x_{2,k}(1 - 6s_k) \leq -8 \\ 12 & \text{if } x_{2,k}(1 - 6s_k) \geq 12 \\ x_{2,k}(1 - 6s_k) & \text{otherwise} \end{cases}$$

Άμα θεωρήσουμε ότι τελικά αυτό το $\bar{x}_k$ **ανήκει για κατάλληλες επιλογές τιμών παραμέτρων στο X όταν βρισκόμαστε κοντά στο σημείο ελαχίστου** τότε η ακολουθία των σημείων είναι η:

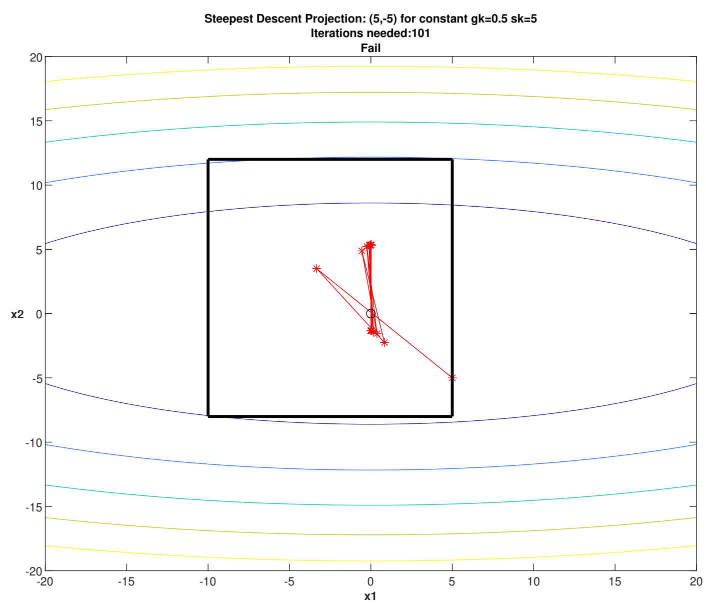
$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} x_{1,k}(1 - \frac{2}{3}s_k\gamma_k) \\ x_{2,k}(1 - 6s_k\gamma_k) \end{bmatrix}$$

Η προηγούμενη υπόθεση είναι λογική καθώς για να συγκλίνει ο αλγόριθμος πρέπει σιγά σιγά να προσεγγίσει το σημείο ελαχίστου 0 και στη γειτονία του το $x_k - s_k\nabla f(x_k)$ δεν θέλουμε να ξεφεύγει εκτός της περιοχής X αφού κάτι τέτοιο θα τοποθετούσε το $\bar{x}_k$ στο σύνορο και για τιμή του γκ που δεν είναι πολύ κοντά στο μηδέν , θα οδηγούταν το x_{k+1} εξαιτίας της παρεμβολής πιο μακρία απο το ελάχιστο από ότι είναι το x_k. Με άλλα λόγια , θέλουμε όταν η μέθοδος μεγίστης καθόδου με προβολή φτάσει αρκετά κοντά στο σημείο ελαχίστου , να εκφυλίζεται εσωτερικά σε μέθοδο μέγιστης καθόδου χωρίς προβολή. Κατά συνέπεια για την σύγκλιση των $x_{1,k}, x_{2,k}$ στο μηδέν πρέπει σε πλήρη αναλογία με την ανάλυση του θέματος 1 να ισχύει:

$$\gamma_k s_k < \frac{1}{3}$$

Αυτή δεν είναι μια συνθήκη που πρέπει να πληρεί αναγκαστικά ο αλγόριθμος για να συγκλίνει. Μπορεί , λόγω κατάλληλων πολύ ειδικών επιλογών το $\bar{x}_k$ να προκύψει σε τέτοιο σημείο που ανάλογα με το x_k που η παρεμβολή τους (γκ) , μπορεί να τύχει πάνω στο ελάχιστο, ενώ τα ίδια απείχαν αρκετά από αυτό . Κάτι τέτοιο θα δούμε στο θέμα 3. Στη γενική περίπτωση , θέλουμε όταν το x_k πλησιάζει στο ελάχιστο να διακοπεί η έξοδος της ποσότητας $x_k - s_k \nabla f(x_k)$ από το X και να συγκλίνουμε στο σημείο ελαχίστου σαν να χρησιμοποιούμε την μέθοδο μέγιστης καθόδου χωρίς προβολή.

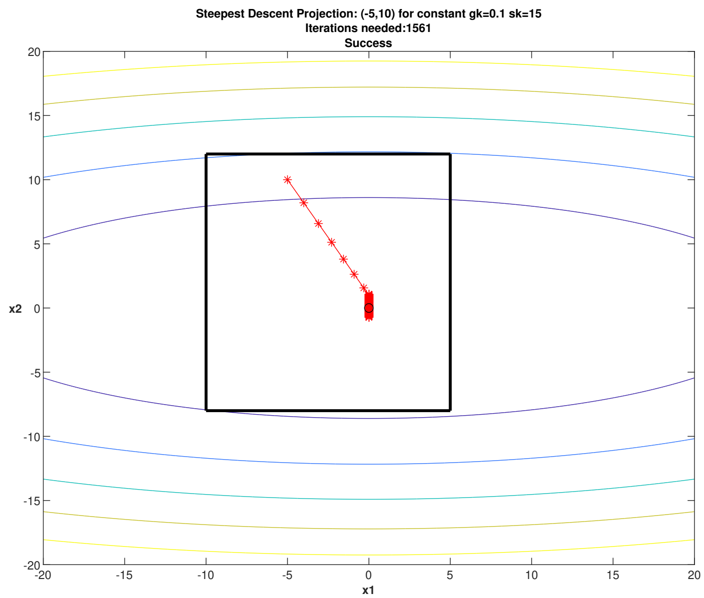
Έχοντας πραγματοποιήσει την ανάλυση την εφαρμογή της μεθόδου της μέγιστης καθόδου με προβολή είμαστε έτοιμη να εφαρμόσουμε την μέθοδο στα δεδομένα του θέματος 2.
Ακολουθεί το διάγραμμα της ακολουθίας των σημείων x_k για εφαρμογή της μεθόδου στην f, με αρχικό σημείο το (5,-5) (το σημείο έναρξης ανήκει στο X) και $\gamma_k = 0.5$ και $s_k = 5$:



Αν και φαίνεται ότι ο αλγόριθμος έχει κάνει μόνο 101 επαναλήψεις στη πραγματικότητα ο αλγόριθμος έχει εγκλωβιστεί μεταξύ δυο σημείων και για αυτό τον τερματίσαμε τεχνητά με μια if. Τα δυο σημεία αυτά είναι τα [0,-1.333] και [0,5.333]. Πράγματι αν πούμε ότι $x_{2,k} = 5.333$, τότε έχουμε $x_{2,k} - sk \cdot 6x_{2,k} = -154.657 < -8$. Άρα: $\bar{x}_{2,k} = -8$ και για γκ=0.5: $x_{2,k+1} = \frac{x_{2,k}+\bar{x}_{2,k}}{2} = -1.3333$. Αντίστοιχα για $x_{2,k} = -1.333$, τότε έχουμε $x_{2,k} - sk \cdot 6x_{2,k} = 38.657 > 12$. Άρα: $\bar{x}_{2,k} = 12$ και για γκ=0.5: $x_{2,k+1} = \frac{x_{2,k}+\bar{x}_{2,k}}{2} = 5.3333$. Βλέπουμε ότι εδώ το γινόμενο $\gamma_k s_k = 2.5 > \frac{1}{3}$. Επομένως ακόμα και να φτάνουμε κοντά στο μηδέν το $\bar{x}$ οδηγείται πάνω στο σύνορο του X και εδώ αφού είναι το γκ=0.5 είναι δύσκολο η παρεμβολή να οδηγήσει στο 0. Εκτός αμα είμαστε τυχεροί ο αλγοριθμός δεν θα τερματίσει ποτέ και στη συγκεκριμένη περίπτωση κατέληξε να κάνει ταλάντωση μεταξύ των δυο σημείων. Άρα επαληθεύεται ότι αμά δεν πάρουμε τις παραμέτρους κατάλληλα Το $x_{1,k}$ ήδη έχει συγκλίνει αρκετά κοντά στο 0.

4 Θέμα 3

Στο παρόν θέμα θα εφαρμόσουμε την μέθοδο ξεκινώντας από το σημείο Ακολουθεί το διάγραμμα της ακολουθίας των σημείων x_k για εφαρμογή της μεθόδου στην f, με αρχικό σημείο το (-5,10) και $\gamma_k = 0.1$ και $s_k = 15$:



Παρατηρούμε ότι η μέθοδος σύγκλινε αφού έκανε πάρα πολλές επαναλήψεις (1261 επαναλήψεις). Έδω το γινόμενο των παραμέτρων είναι:

$$\gamma_k s_k = 1.5 > \frac{1}{3}$$

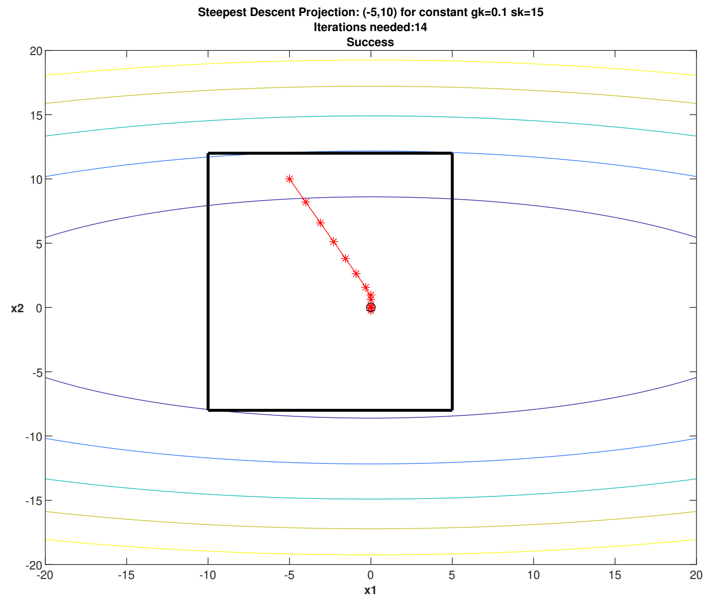
Συνεπώς, δεν ισχύει η ζητούμενη συνθήκη και δεν θα μπορεί να συγκλίνει ομαλά η μέθοδος όπως προαναφέρθηκε στην ανάλυση στο θέμα 2. Παρατηρούμε, εδώ ότι το $x_{1,k}$ έχει καταφέρει να συγκλίνει αφού $\gamma_k s_k < 3$ και για αυτό όταν το αυτό είναι αρκετά μικρό ώστε ο όρος $ed_{1,k}$ να ανήκει εντός του X (και συνεπώς να μην γίνει προβολή για την χ1 συνιστώσα) , τότε γιατί γκsk=1.5 η τιμή του $x_{1,k}$ γίνεται ίση με το μηδέν , όπως μπορούμε να δούμε στις προηγούμενες σχέσεις . Ωστόσο, το $x_{2,k}$ κινείται γύρω από το μηδέν καθώς λόγω του μεγάλου sk=15, η $\bar{x}_{2,k}$ στέλνεται μακριά από το μηδέν και άμα τύχει να φτάσει ξανά κοντά στο μηδέν το $x_{2,k}$ λόγω της παρεμβολής , πάλι τότε θα ξεφύγει εξαιτίας του $\gamma_k s_k > \frac{1}{3}$. Θεωρητικά μόνο αν είμαστε τυχεροί θα μπορέσει κάποιο από τα $x_{2,k}$ κάπου κοντά στο μηδέν ώστε ικανοποιείται η επιθυμητή ακρίβεια και ο αλγόριθμος να τερματιστεί.

Ένας απλός τρόπος για να εξαναγκάσουμε την σύγκλιση της μεθόδου στο ελάχιστο είναι να αλλάξουμε εσωτερικά στην μέθοδο την παράμετρο του sk ώστε να εγγυηθούμε ότι όταν φτάσουμε ικανοποιητικά κοντά στο σημείο ελαχίστου η μέθοδος να εκφυλίζεται στην μέθοδο μέγιστης καθόδου χωρίς προβολή . Για να γίνει αυτό πρέπει να παραμείνει η ποσότητα $ed_k = x_k - s_k \nabla f(x_k)$ εντός του X ,και για αυτό πρέπει οι συνιστώσες της να τηρούν τους περιορισμούς:

$$-10 \leq ed_{1,k} \leq 5, -8 \leq ed_{2,k} \leq 12$$

Αυτό μπορούμε να το επιβάλλουμε οτί θα ισχύει για τον πρώτο όρο άμα τα $ed_{1,k} = x_{1,k}(1 - \frac{2}{3}s_k), x_{1,k}$ είναι ομόσημα και άρα ο συντελεστής $(1 - \frac{2}{3}s_k)$ είναι θετικός και μικρότερος του 1 δηλαδή άμα $s_k < 1.5$. Αυτό θα ισχύει γιατί το $x_{1,k}$ ικανοποιεί τον πρώτο περιορισμό και συνεπώς θα εγγυηθούμε ότι σίγουρα θα τον ικανοποιεί και ο $ed_{1,k}$ θα τον ικανοποιεί άν διατηρήσει το ίδιο πρόσημο και άπλα χαμηλώσει η απόλυτη τιμή του. Άρα κατά πλήρη αντίστοιχια έχουμε ότι για τον περιορισμό του $x_{k,2}$ θέλουμε να ισχύει $0 < (1 - 6s_k) < 1$ και άρα $s_k < \frac{1}{6}$. Άρα για να ισχύουν και τα δύο ταυτόχρονα παίρνοντας την τομή έχουμε $sk < \frac{1}{6}$. Αυτό είναι μονάχα ικανή αλλά **όχι** και αναγκαία συνθήκη για να ανήκει το ed_k στο εσωτερικό του X . Άρα έχοντας αυτή την αρκετά αυστηρή και περιοριστική συνθήκη αλλά και εξασφαλίζοντας ότι έχουμε και το $\gamma_k s_k < \frac{1}{3}$, μπορούμε να εγγυηγθουμε ότι το μέθοδος θα συγκλίνει και μάλιστα η λειτουργίας της θα είναι πανομοιότυπη με την απλή μέθοδο μέγιστης καθόδου χωρίς προβολή με βήμα γκsk αφού η προβολή δεν θα αξιοποιείται.

Έτσι άμα βάλουμε έναν περιορισμό της φύσης $|\nabla f| < 2$ στον οποίο θα είμαστε αρκετά κοντά στο ελάχιστο σημείο , τότε μπορούμε να αλλάξουμε τα s_k και γ_k τέτοια που να ικανοποιούν τις παραπάνω συνθήκες. Για παράδειγμα $\gamma_k = 1$ και $s_k = \frac{1}{7}$ για το οποία ισχύει ότι $\gamma_k s_k < \frac{1}{3}$ και $s_k < \frac{1}{6}$. Το γ_k το μεγαλώνουμε στην μέγιστη τιμή του ώστε να εκμεταλλευτούμε ένα μεγαλύτερο βήμα και συνεπώς έναν πιο γρήγορο ρυθμό σύγκλισης. Πράγματι , έχουμε:

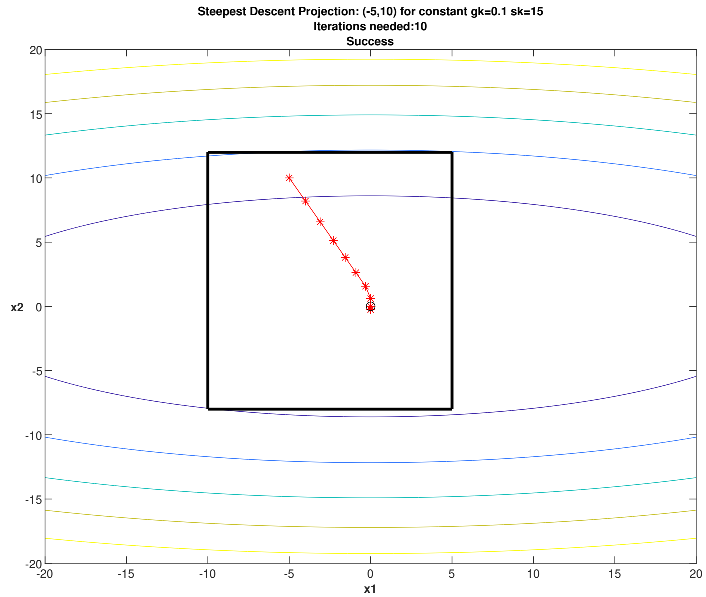


Από τον πίνακα της ακολουθίας xk βλέπουμε ότι έχουμε συγκλίνει αρκετά κοντά σρο μηδέν άν όχι πάνω σε αυτό.

Εναλλακτικά, μπορούμε να αξιοποιήσουμε την λογική που είδαμε στο μάθημα των ασκήσεων και αφού όπως είδαμε παραπάνω το sk=1.5 είναι τέτοιο που μηδενίζει την x1k όταν δεν αξιοποιείται η προβολή, μπορούμε να αλλάξουμε το sk σε τέτοια τιμή που εξαναγκάζει όπως το x1k τον μηδενισμό του x2k όταν δεν είμαι τόσο κοντά που δεν θα έχουμε προβολή. Άρα

$$x2_{k+1} = 0 \rightarrow x2_k(1 - 6\gamma k s_k) = 0 \rightarrow sk = \frac{10}{6}$$

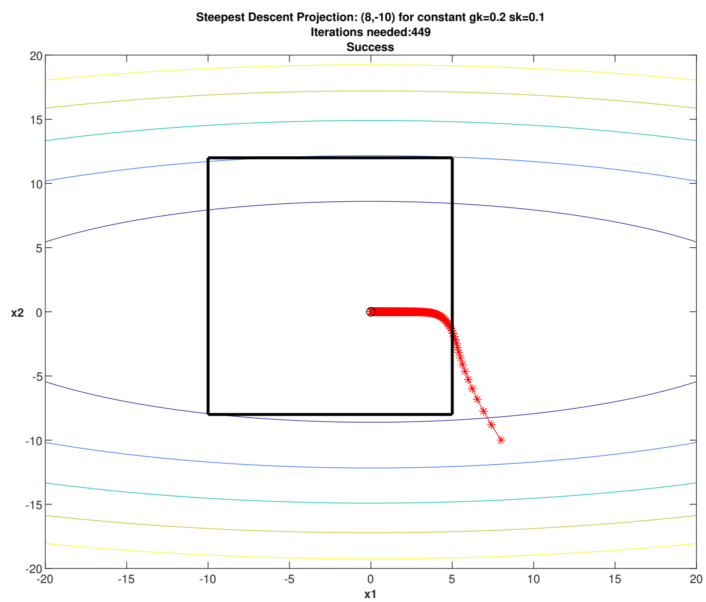
Πράγματι άμα κάνουμε αυτή την αλλαγή όταν μηδενιστεί το x1k έχουμε σύγκλιση όπως βλέπουμε παρακάτω:



Αυτές οι αλλαγές περιέχονται στον κώδικα της steepest_descent_projection.m σε σχόλιο. Για να της υλοποιήσουμε θα πρέπει η εκάστοτε αλλαγή να βγεί από σχόλιο χειροκίνητα.

5 Θέμα 4

Τέλος , εφαρμόζουμε την μέθοδο για σημείο έναρξης το [8 ,-10] και $\gamma_k = 0.2, s_k = 0.1$. Αυτό το σημείο έναρξης δεν ανήκει στο σύνολο X άρα κανονικά δεν θα έπρεπε να εκτελεστεί η μέθοδος. Ώστοσο, η άσκηση μας ζητάει να εκτελέσουμε τον αλγόριθμο όπως και να έχει και να δούμε τι θα πάρουμε. Καταρχάς οφείλουμε να πούμε ότι καθώς η συνάρτηση είναι κυρτή όσο είμαστε εξώ από το σύνολο περιορισμών το χκ θα κινείται προς το εσωτερικό του (καθώς το grad) διάνυσμα δείχνει προς αυτό και τελικά όταν μπούμε στο εσωτερικό του από την ανάλυση που κάναμε στο θέμα 3 ισχύει $s_k = 0.1 < \frac{1}{6}$ και $\gamma_k s_k = 0.02 < \frac{1}{3}$ Συνεπώς , για τους λόγους που είδαμε στο θέμα 3 η μέθοδος εκφυλίζεται σε αυτή της μέγιστης καθόδου με βήμα skγκ=0.02 το όποιο είναι κατάλληλο συμφώνα με το θέμα 1 ώστε η μέθοδος μέγιστης καθόδου χωρίς προβολή να συγκλίνει. Κατα συνέπεια έχουμε πράγματι σύγκλιση:


Steepest Descent Projection: (8,-10) for constant gk=0.2 sk=0.1
Iterations needed:449
Success
x1
x2
-20
-15
-10
-5
0
5
10
15
20

Άμα επιλέγξουμε να ξεκινήσουμε από την προβολή του (8,-10) στο X δηλαδή από το (5,-8) έχουμε πάλι σύγκλιση για τους ίδιους λόγους που αναφέρθηκαν προηγουμένως.