Τεχνικές Βελτιστοποίησης

ΕΡΓΑΣΙΑ 3

ΝΙΚΟΛΑΟΣ ΙΣΤΑΤΙΑΔΗΣ

9175 nikoista@ece.auth.gr

Μέθοδος μέγιστης καθόδου με προβολή

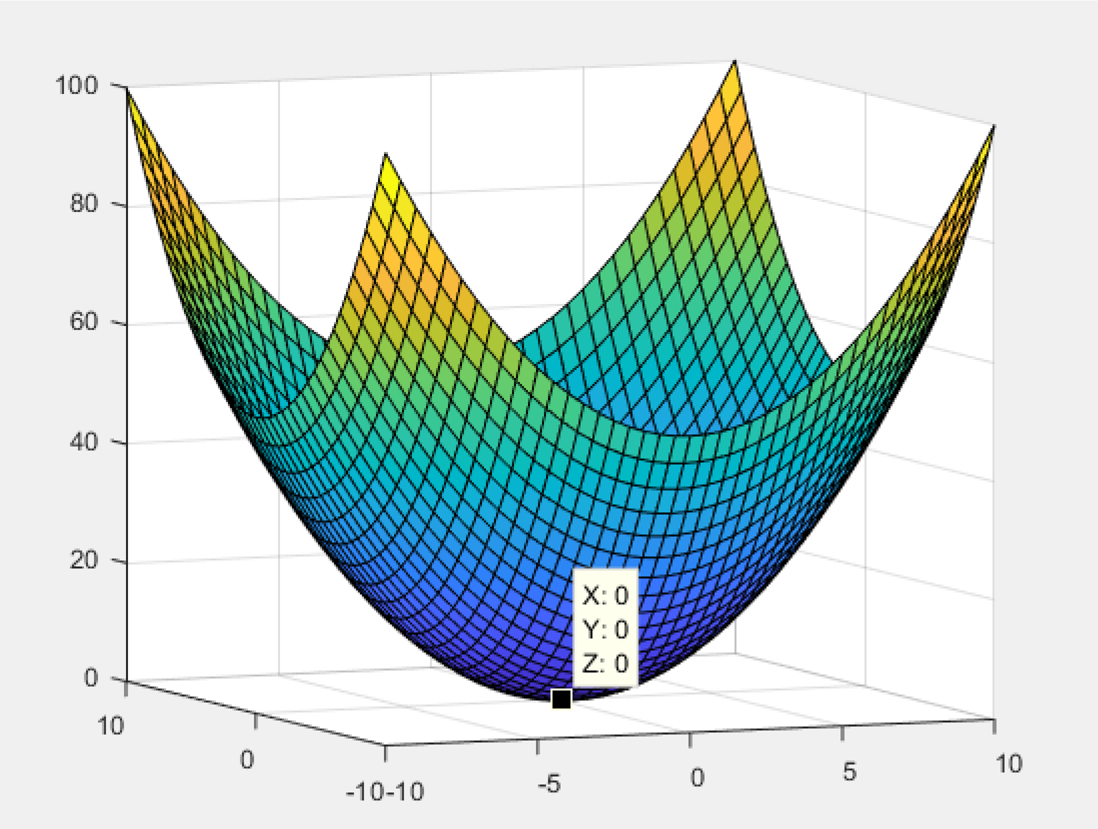
Θεωρούμε την τετραγωνική συνάρτηση:

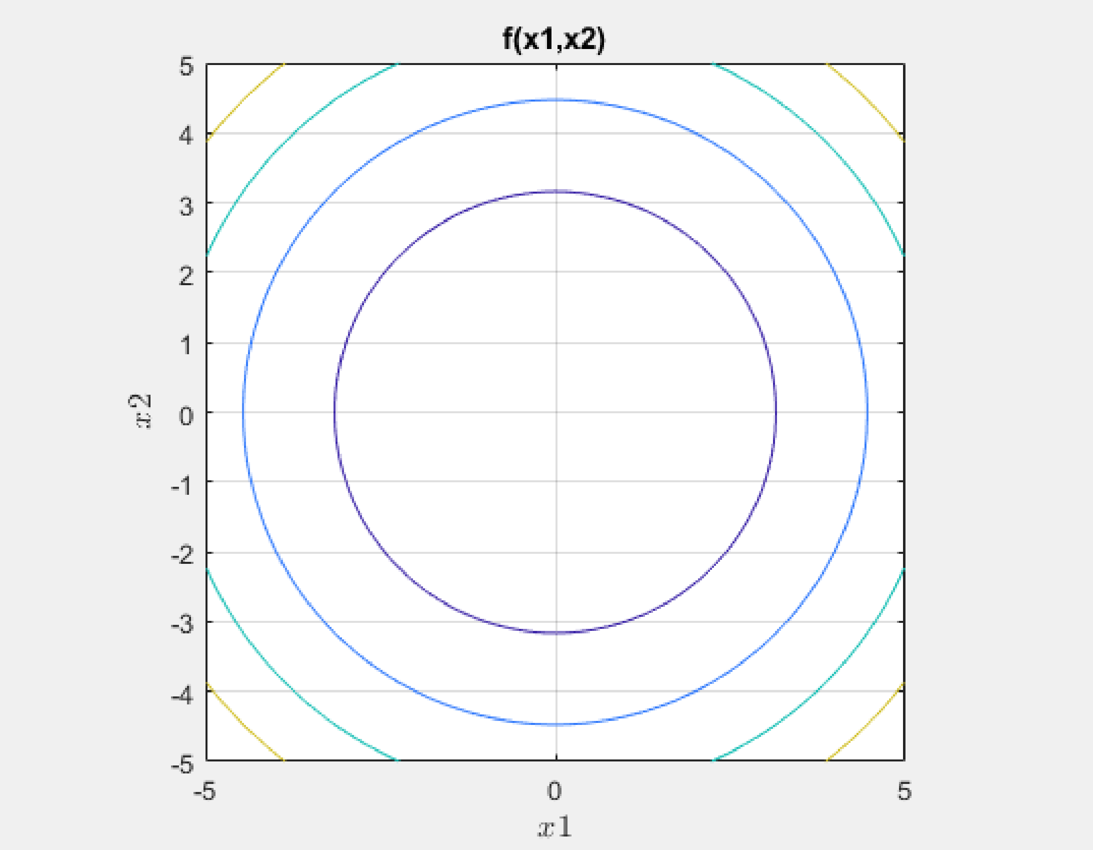
$$f: \mathrm{R}^2 \rightarrow \mathrm{R}\,,\; f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$$

Σχεδίαση Συνάρτησης με Matlab

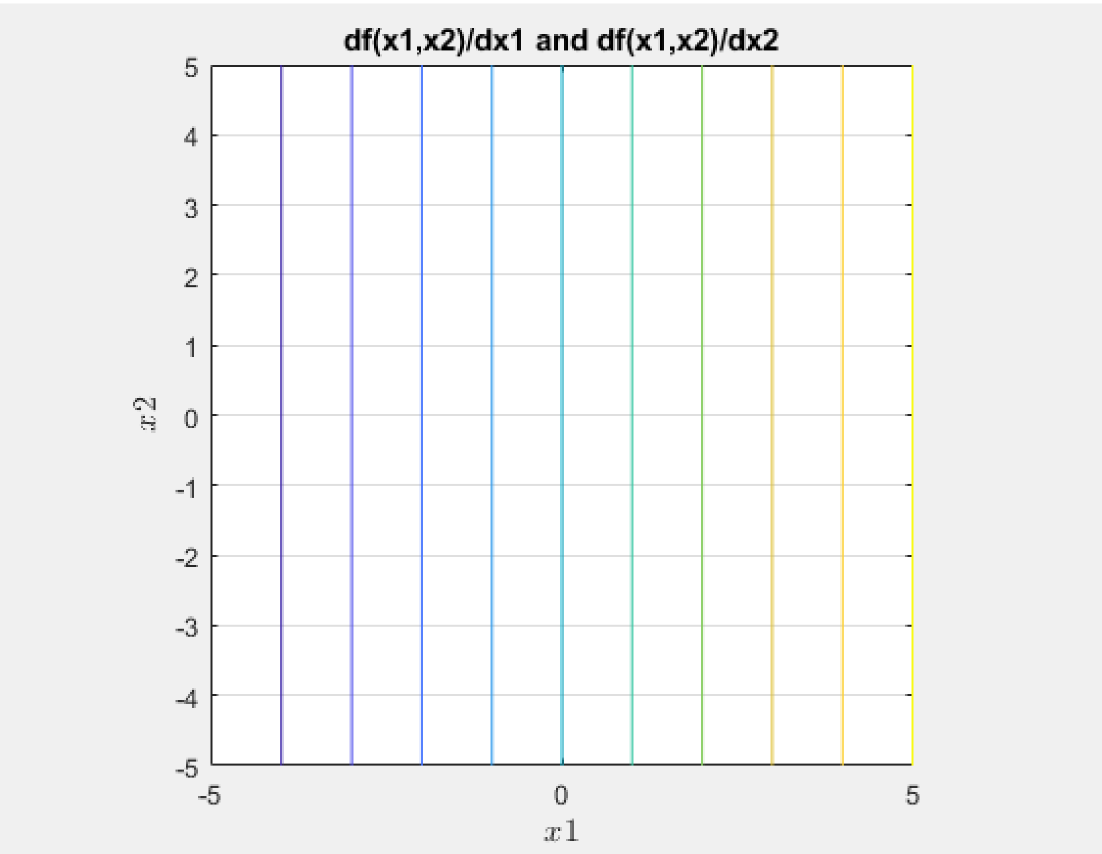
Εφαρμόζοντας την συνάρτηση f(x) με Matlab έχουμε τα εξής plot:

Είναι προφανές πως το ολικό ελάχιστο της f είναι το x*= (0,0)

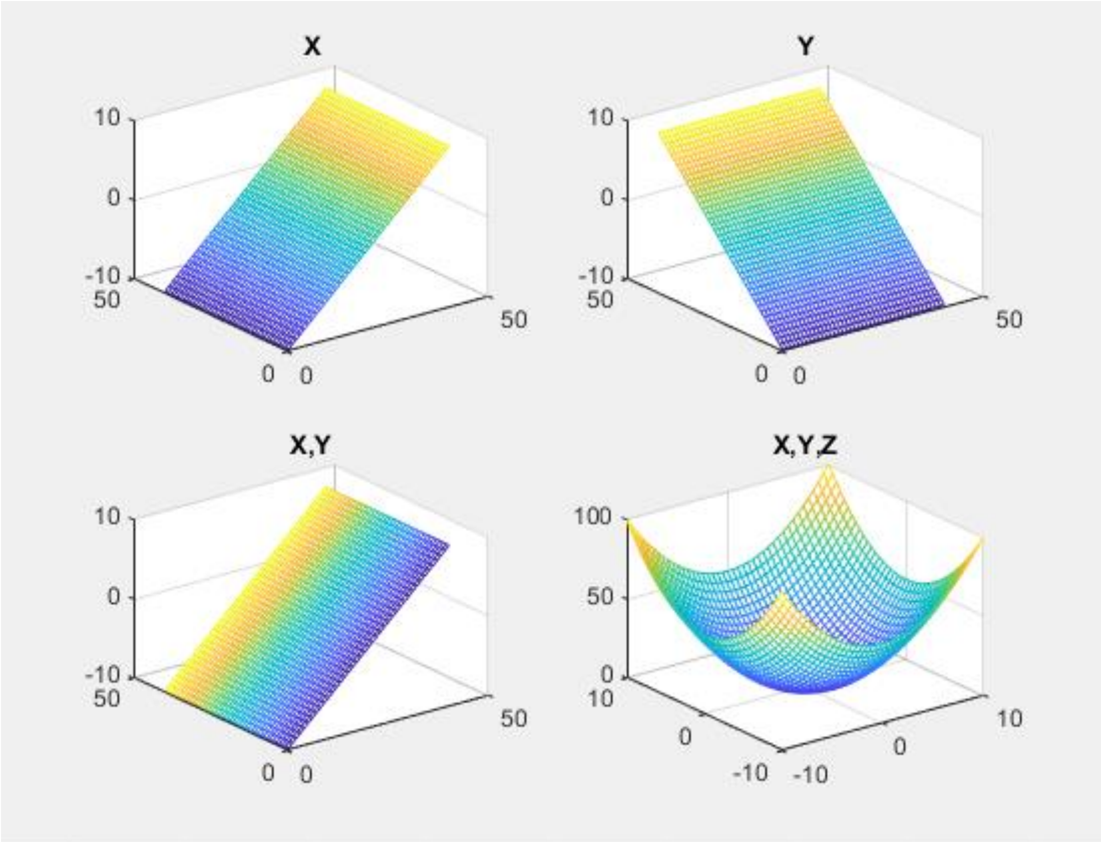




f(x1,x2)
5
4
3
2
1
0
-1
-2
-3
-4
-5
-5
0
5
x1
x2



df(x1,x2)/dx1 and df(x1,x2)/dx2
5
4
3
2
1
0
-1
-2
-3
-4
-5
-5
0
5
x1
x2



Θέμα 1

Να χρησιμοποιηθεί η μέθοδος μεγίστης καθόδου (προηγούμενη εργασία) με ακρίβεια ε = 0.01 και βήμα i) γκ = 0.1 , ii) γκ = 2 , iii) γκ = 2 , iv) γκ = 10 , και οποιοδήποτε σημείο εκκίνησης διάφορο του (0,0) . Τι παρατηρείτε; Να αποδειχθούν τα αποτελέσματα αυτά με μαθηματική αυστηρότητα.

Θεωρητική Ανάλυση - Μέθοδο Μέγιστης Καθόδου

Έχουμε την συνάρτηση : $f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$ και έστω ότι γκ=γ και σημείο εκκίνησης xk οποιοδήποτε εκτός από το (0,0). Είναι προφανές ότι το ζητούμενο ελάχιστο είναι το (0,0). Άρα σύμφωνα με την μέθοδο της Μέγιστης Καθόδου έχουμε $\nabla f(x_1, x_2)$ = [x1 , x2] και για $x_{k+1} = x_k + \gamma_k d_k$ με dk = -Dk$\nabla f(x_k)$ ενώ Dk=I μοναδιαίος Πίνακας . Με πράξεις καταλήγουμε στα εξής :

$$x_{1k+1} = x_{1k} - \gamma_k x_{1k}$$

$$x_{2k+1} = x_{2k} - \gamma_k x_{2k}$$

Θέλουμε $\frac{|x_k+1|}{|x_k|} = |1 - \gamma_k|$ ώστε η ακολουθία των παραγόμενων σημείων να συγκλίνει στο $x^* = (0,0)$. Οπότε $0 < \gamma\kappa < 2$ και για κάθε άλλη επιλογή συντελεστή γκ η μέθοδος αποκλίνει.

Σύμφωνα με το θεώρημα 5.2.1 φανερώνει πως η πορεία προς το ελάχιστο είναι μια τεθλασμένη γραμμή με διαδοχικά ευθύγραμμα τμήματα που σχηματίζουν 90° γωνία . Αυτή η "ζικ-ζακ¨ συμπεριφορά καθιστά τη μέθοδο ως επί το πλείστον αργή και κατά συνέπεια αναποτελεσματική.

Η μόνη περίπτωση ο αλγόριθμος της μεγίστης καθόδου να μας οδηγήσει στο ελάχιστο x^*=(0,0) σε ένα βήμα είναι όταν το διάνυσμα που το συνδέει το χκ με το x^* είναι συγγραμικό με το $\nabla f(x_k)$. Τέλος για μικρό βήμα η μέθοδος έχει πολύ μικρή ταχύτητα σύγκλισης ενώ για μεγάλο βήμα η μέθοδος οδηγεί σε ταλαντώσεις ή ακόμα και να αποκλίνει όταν το βήμα γκ είναι αρκούντος μεγάλο.

Κριτήρια Καλής Λειτουργίας.

Όπως είδαμε στην θεωρητική ανάλυση στην μέθοδο Μέγιστης Καθόδου έχουμε δύο σοβαρά μειονεκτήματα:

1) Ο υπολογισμός του βήματος γκ

2) Η κίνησή της σε κάθετα βήματα , που καθιστούν τη σύγκλισή της αργή και κατά συνέπεια υπολογιστικά ανεπαρκή.

Συνεπώς για κάθε κ , απαιτούμε από τα γκ και dk να ικανοποιούν τα παρακάτω κριτήρια:

Κριτήριο 1 $f(x_k+1) < f(x_k)$, οποτεδήποτε $\nabla f(x_k) \neq 0$

Κριτήριο 2 $d_k^T \nabla f(x_k) < 0$

Κριτήριο 3 Υπάρχει β με 0 < β < 1 τέτοιο ώστε $d_k^T \nabla f(x_k+1) > b d_k^T \nabla f(x_k)$

Κριτήριο 4 Υπάρχει α,β με 0< α < β < 1 τέτοιο ώστε$f(x_{k+1}) \leq f(x_k) + a\gamma_k d_k^T \nabla f(x_k)$

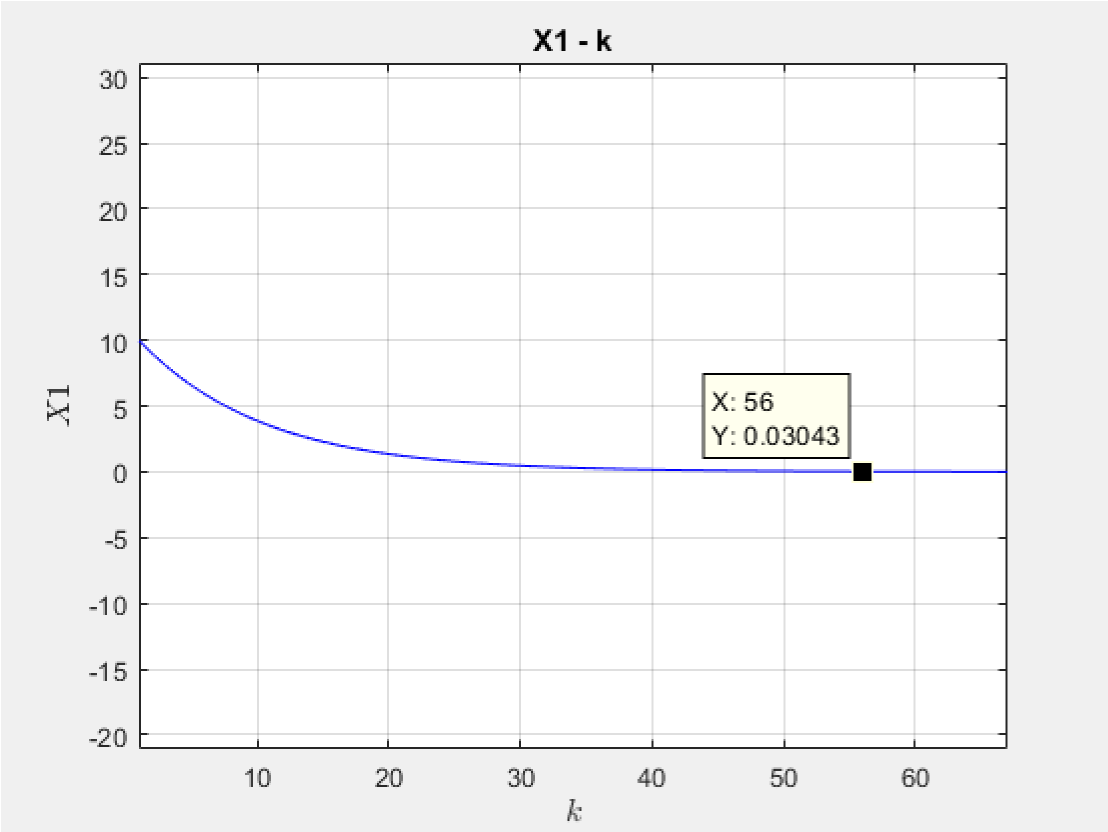
Matlab

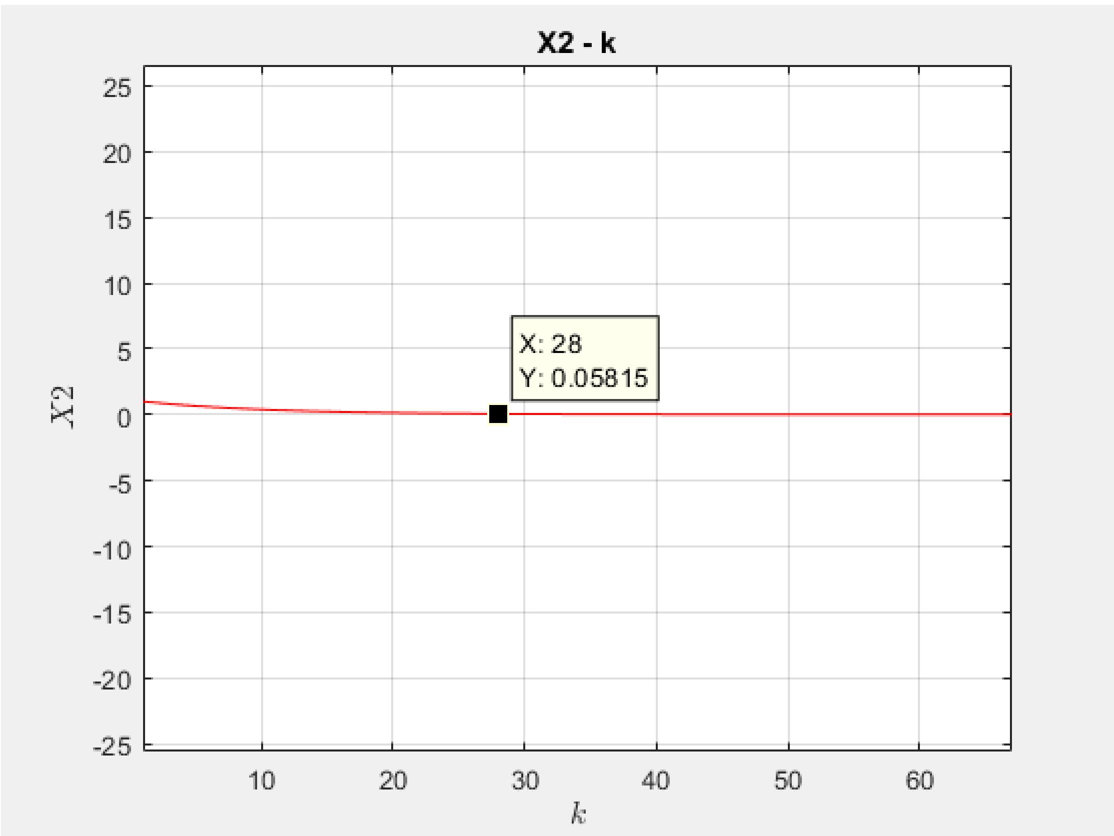
Στο αρχείο TEX_BEL_THEMA_1_STATHERA_GAMMA_DECENT.m είναι η επίλυση του 1 θέματος . Για τα γκ που αναφέρει η εκφώνηση και για αρχική συνθήκη το (10,1) που θεώρησα ότι είναι αντιπροσωπευτικό διότι περιέχει το x1 κοντά στο ελάχιστο και το x2 μακριά , έχουμε τα εξής plots και συμπεράσματα:

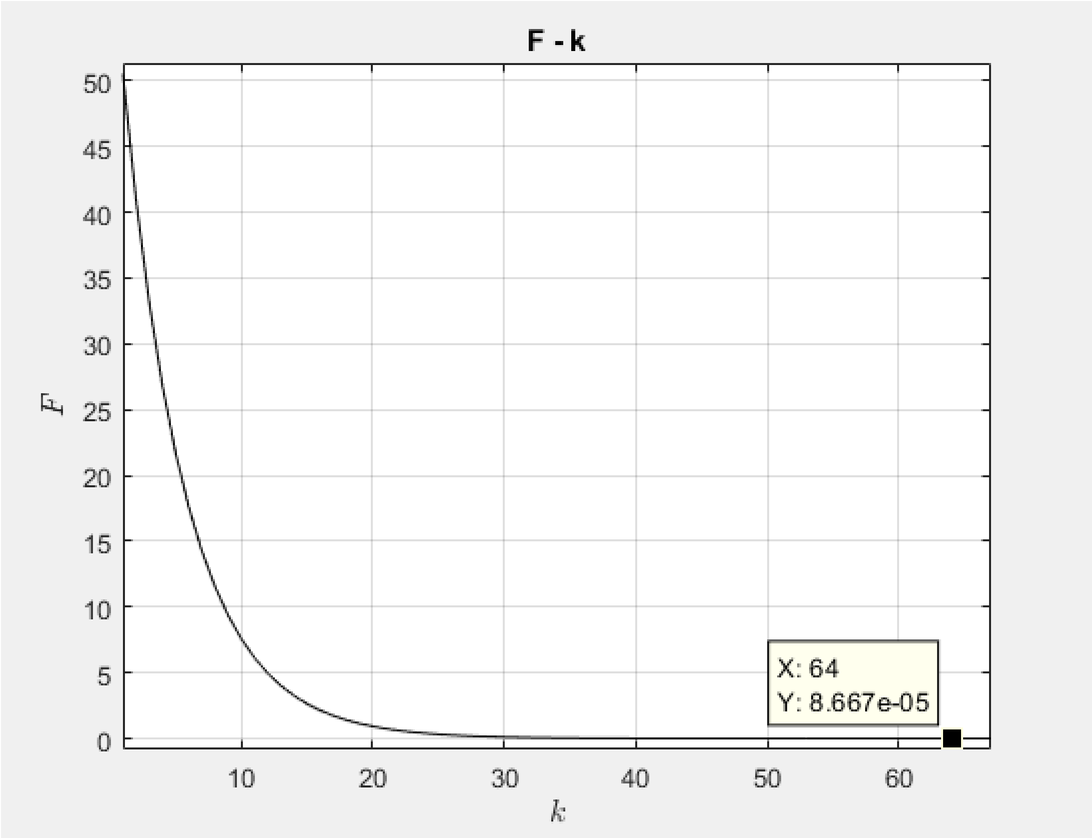
i) γκ = 0.1 με σημείο εκκίνησης (10,1) και ακρίβεια ε =0.01

Όπως βλέπουμε ισχύει η σύγκλιση της μεθόδου διότι $0 < \gamma k < 2$. Μέσα σε 67 επαναλήψεις έχουμε πολύ καλό διάστημα σύγκλισης [0.0095, 0.00095] και

fk= 0.0000460. Όσο αυξάνουμε την ακρίβεια τόσο οδηγούμαστε προς το (0,0)

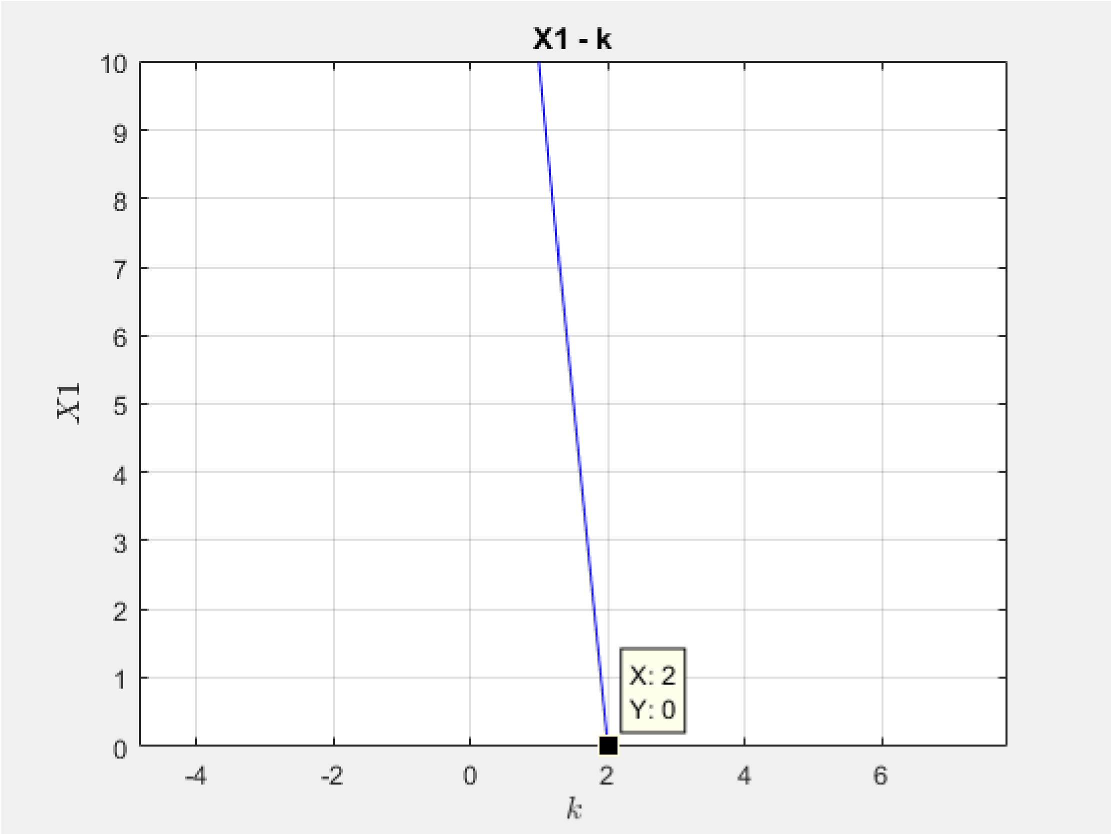

X1 - k
30
25
20
15
10
5
0
-5
-10
-15
-20
X1
X: 56
Y: 0.03043
10
20
30
40
50
60
k

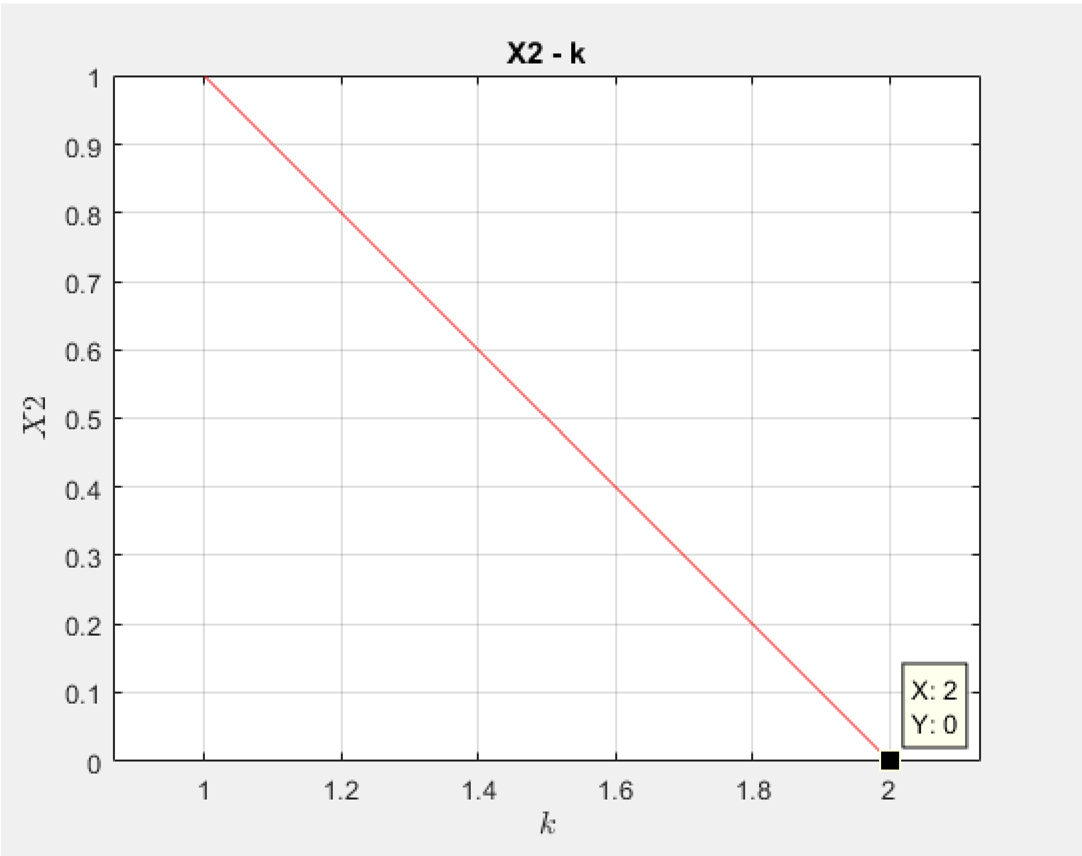

X2 - k
25
20
15
10
5
0
-5
-10
-15
-20
-25
X2
X: 28
Y: 0.05815
10
20
30
40
50
60
k

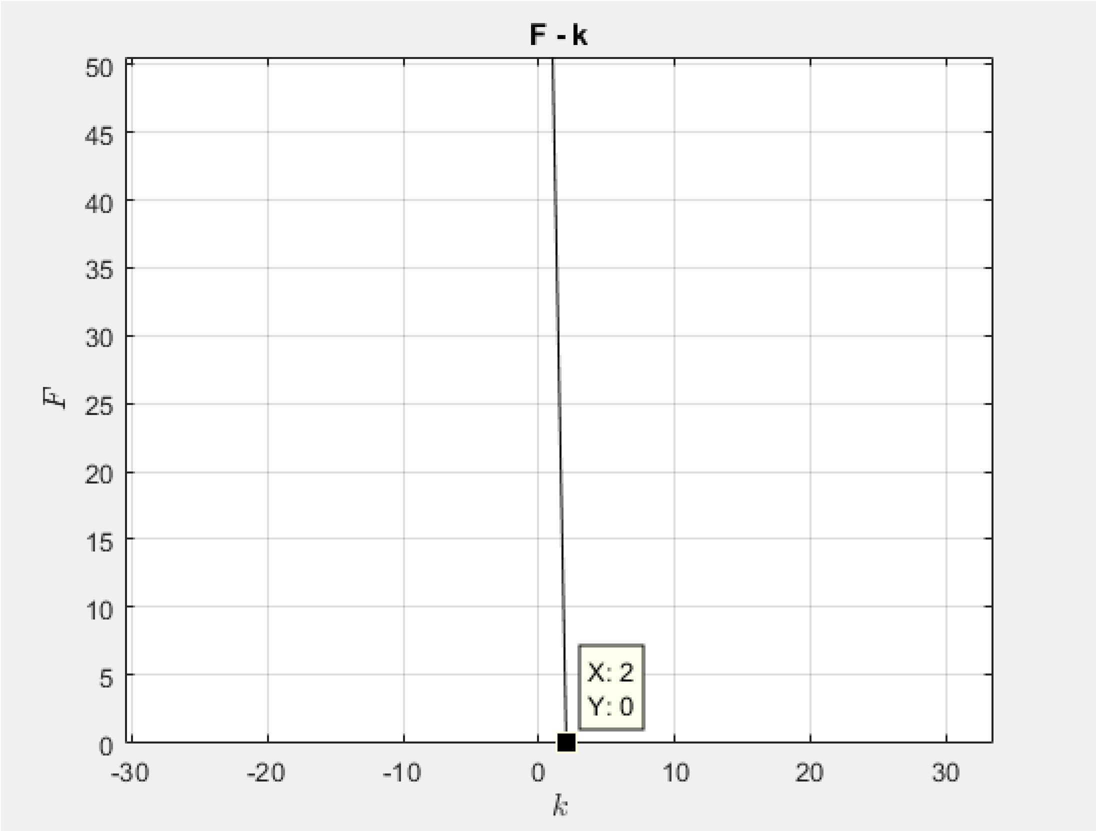


ii) γκ = 1 με σημείο εκκίνησης (10,1) και ακρίβεια ε =0.01

Όπως βλέπουμε ισχύει η σύγκλιση της μεθόδου διότι $0 < \gamma k < 2$. Μέσα σε 1 επανάληψη έχουμε το τέλειο διάστημα σύγκλισης [0, 0] και fk= 0. Τονίσαμε στην θεωριτική ανάλυση πως για να συγκλίνει ο αλγόριθμος σε ένα βήμα πρέπει το διάνυσμα που συνδέει το χκ με το x* να είναι συγγραμικό με το $\nabla f(x_k)$όπου ισχύει για γκ = 1 και αποτελεί την βέλτιστη λύση του προβλήματος εύρεσης ελαχίστου

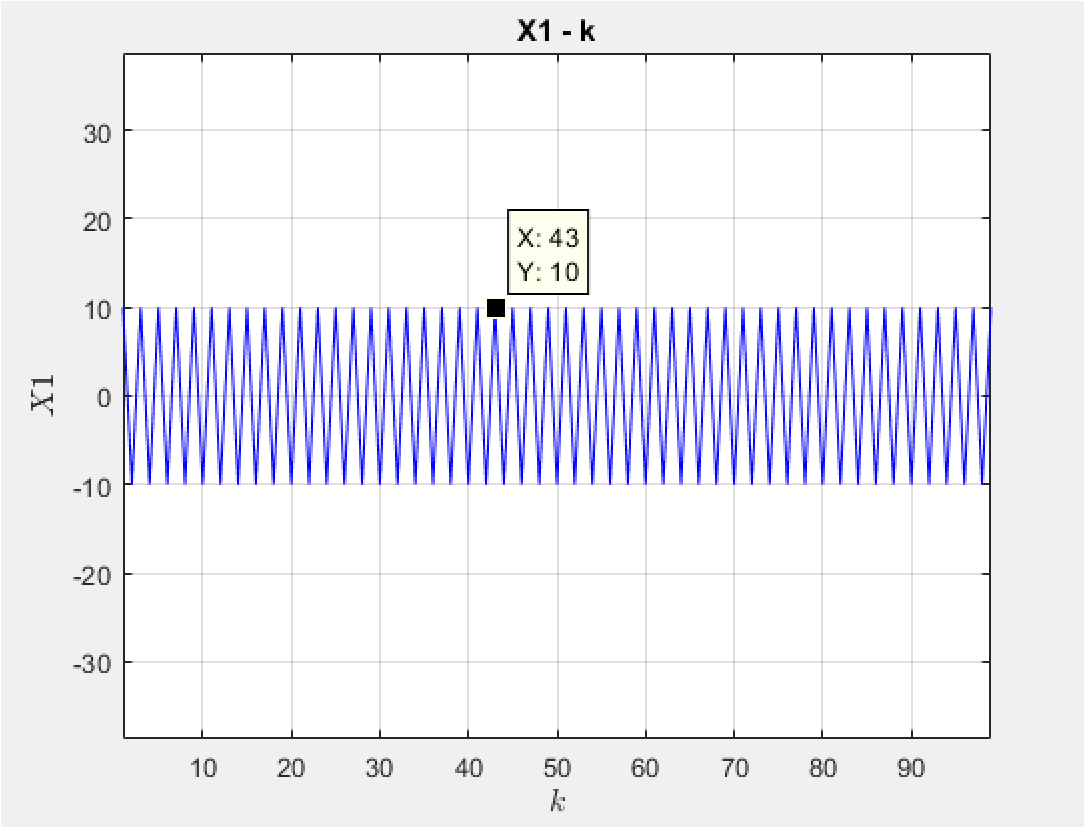

X1 - k
10
9
8
7
6
5
4
3
2
1
0
$X1$
-4
-2
0
2
4
6
k
X: 2
Y: 0

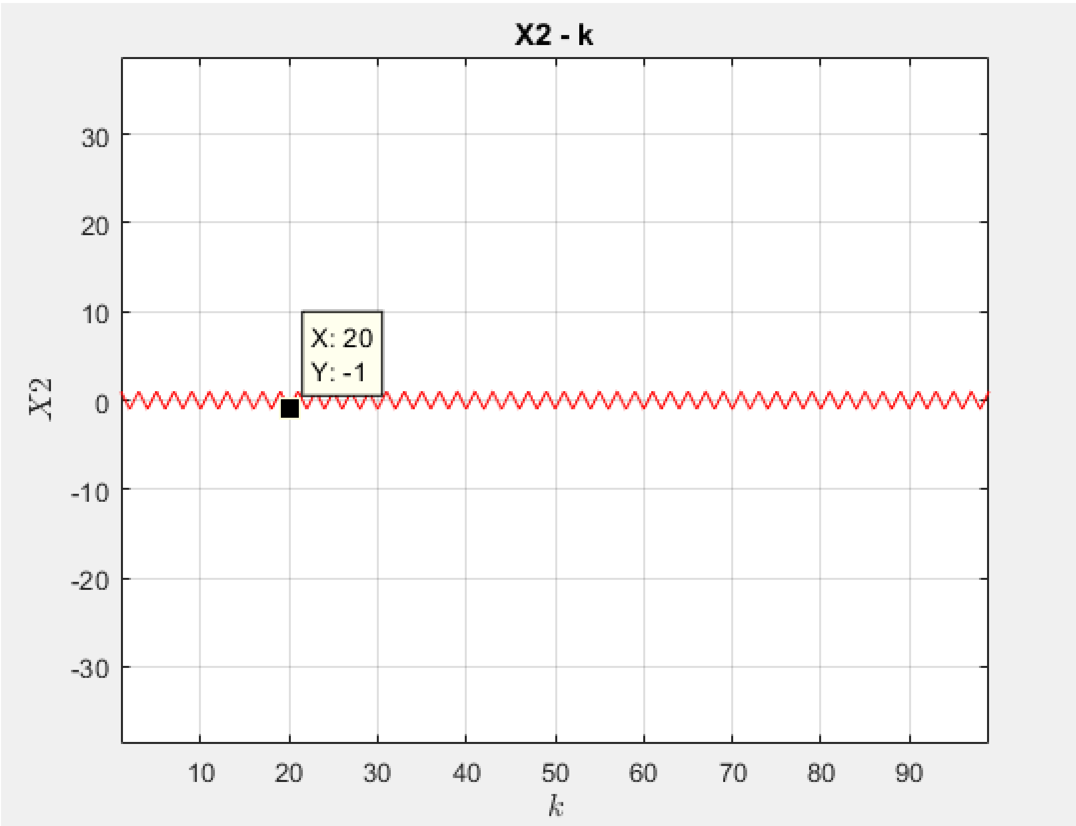

X2 - k
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
$X2$
1
1.2
1.4
1.6
1.8
2
k
X: 2
Y: 0

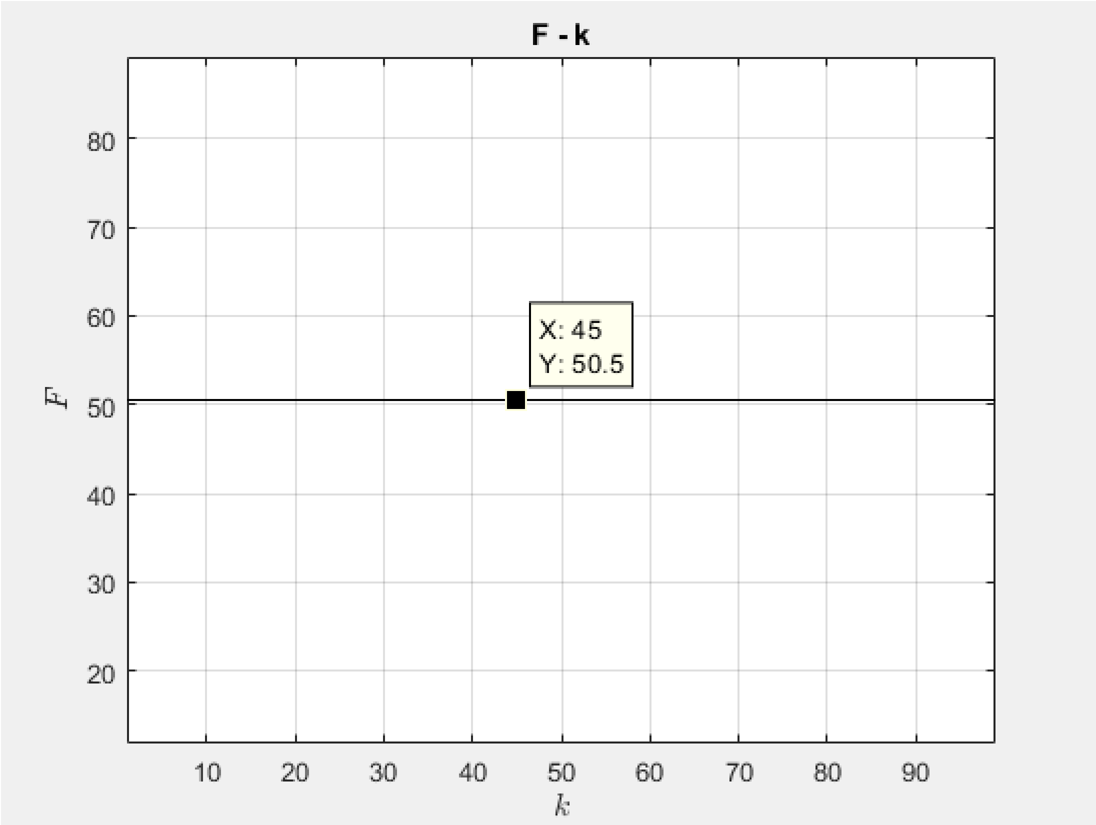


iii) γκ = 2 με σημείο εκκίνησης (10,1) και ακρίβεια ε =0.01

Όπως βλέπουμε δεν ισχύει η σύγκλιση της μεθόδου διότι $0 < \gamma k < 2$. Για μεγάλο βήμα η μέθοδος οδηγεί σε ταλαντώσεις όπως βλέπουμε στην περίπτωσή μας . Το κριτήριο 4 δεν πληρείται και ο αλγόριθμος ταλαντώνεται ανάμεσα στο σημείο εκκίνησης και το αντίθετο του ((10,0) και (-10,-1)) . Δεν έχουμε εύρεση ελαχίστου για γκ =2 .

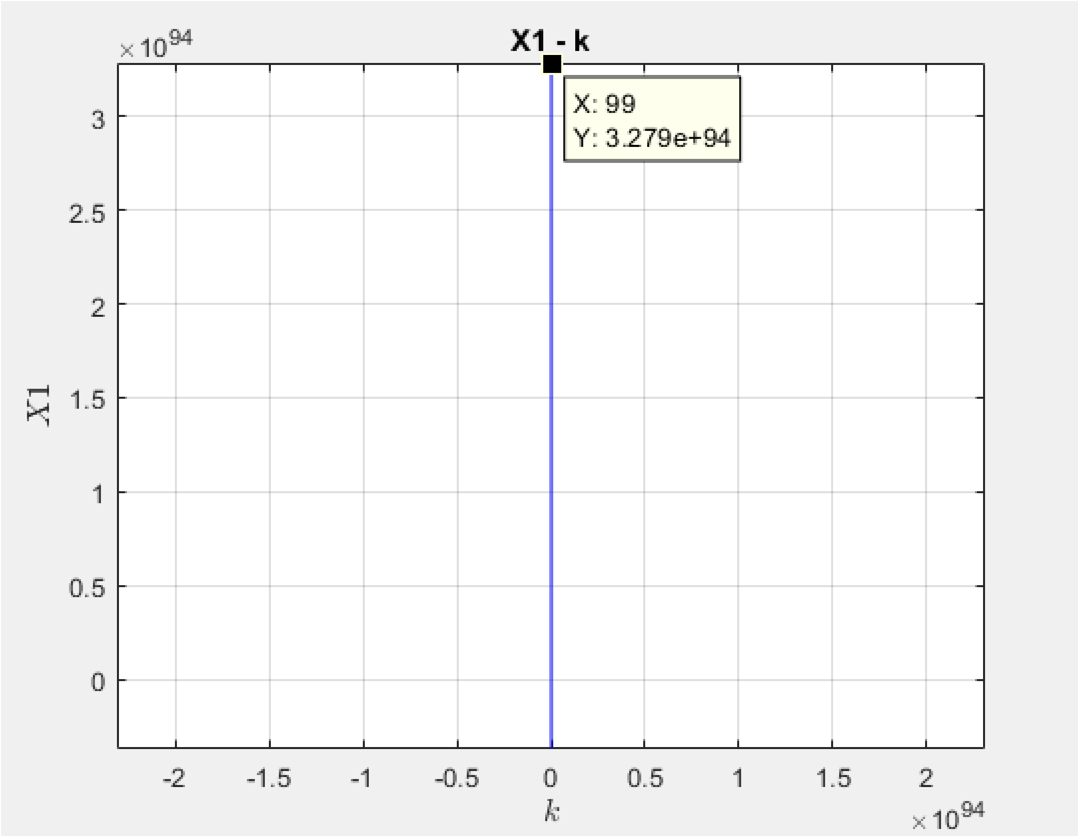

X1 - k
X: 43
Y: 10
30
20
10
0
-10
-20
-30
X1
10
20
30
40
50
60
70
80
90
k

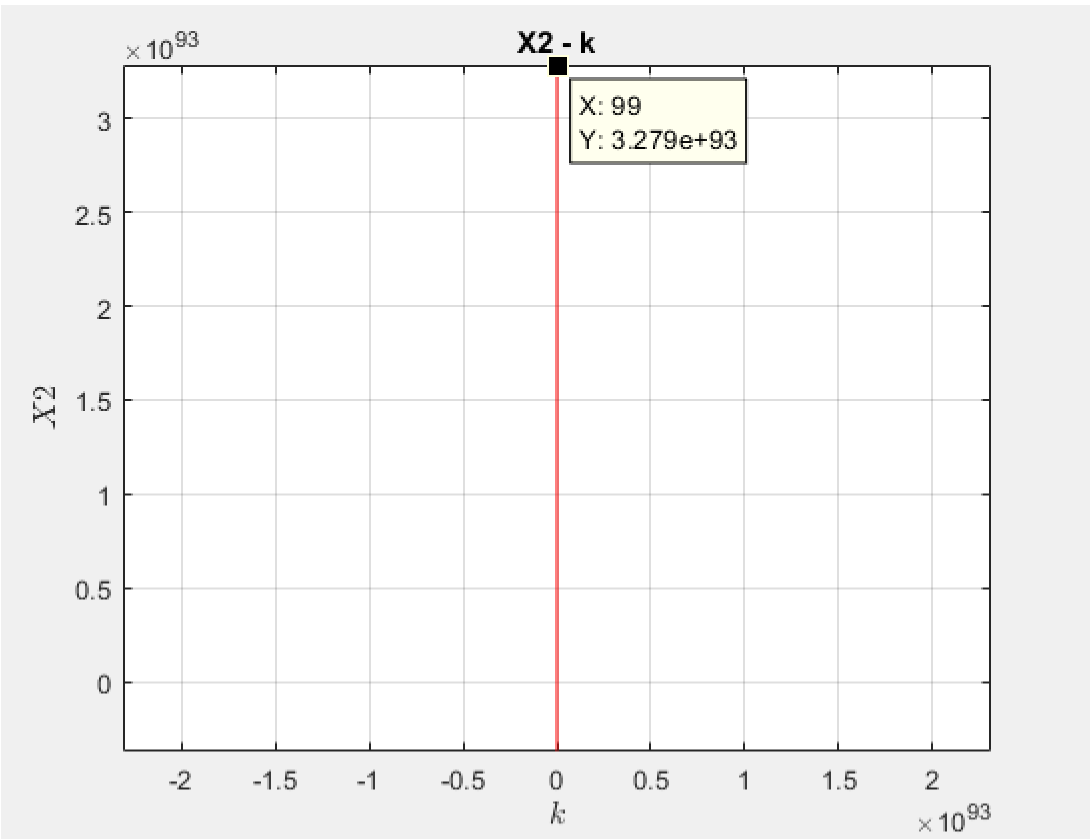

X2 - k
X: 20
Y: -1
30
20
10
0
-10
-20
-30
X2
10
20
30
40
50
60
70
80
90
k

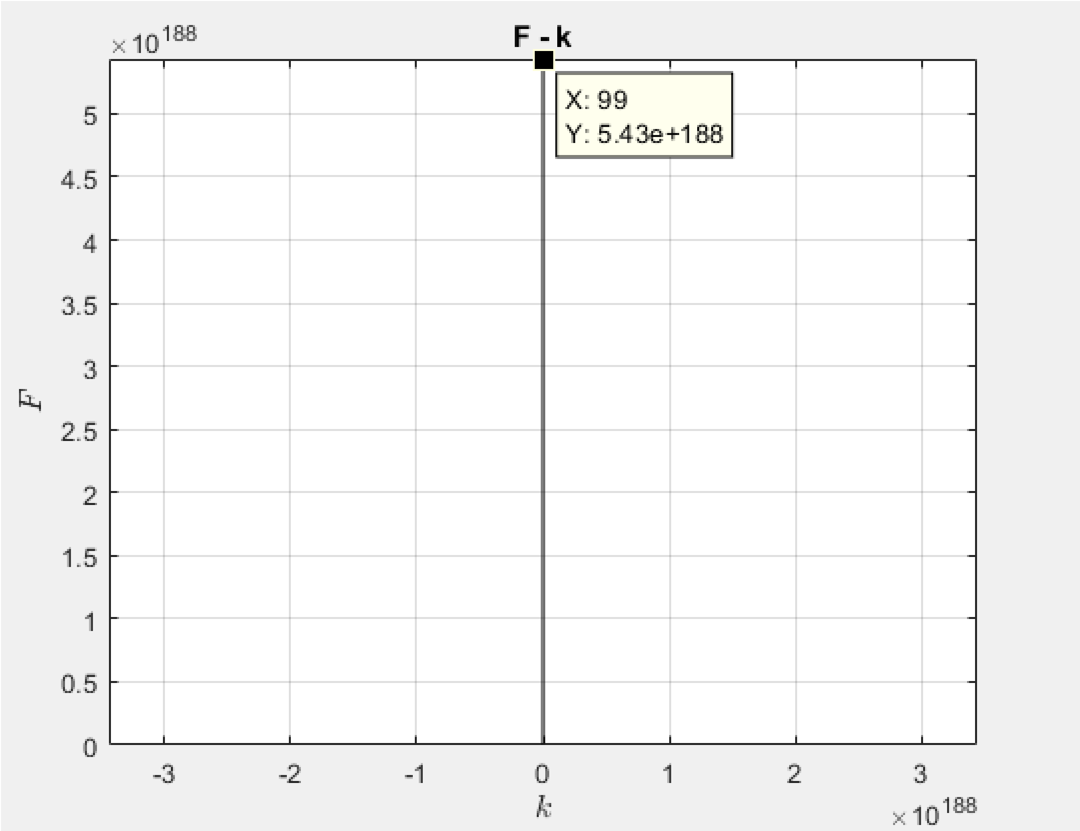


iv) γκ = 10 με σημείο εκκίνησης (10,1) και ακρίβεια ε =0.01

Όπως βλέπουμε δεν ισχύει η σύγκλιση της μεθόδου διότι $0 < \gamma k < 2$. Για τόσο μεγάλο βήμα η μέθοδος αποκλίνει όπως βλέπουμε στην περίπτωσή μας . Το κριτήριο 1,4 δεν πληρούνται και ο αλγόριθμος οδηγείται προς το άπειρο . Δεν έχουμε εύρεση ελαχίστου για γκ =10 .


X1 - k
×10^94
X: 99
Y: 3.279e+94
3
2.5
2
1.5
1
0.5
0
X1
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
k
×10^94


X2 - k
×10^93
X: 99
Y: 3.279e+93
3
2.5
2
1.5
1
0.5
0
X2
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
k
×10^93



Θεωρείστε τώρα τους περιορισμούς:

$$-20 \leq X_1 \leq 10 \ , \ -12 \leq X_2 \leq 15$$

Θεωρητική Ανάλυση -Μέθοδο Μέγιστης Καθόδου με Προβολή

Η μέθοδος μέγιστης καθόδου με προβολή είναι αλγόριθμος εφικτών σημείων της μορφής :

$$x_{k+1} = x_k + \gamma_k \left(\bar{x}_k - x_k\right) , \gamma_k \in (0,1]$$

$$\bar{x}_k = P_r x\{x_k - s_k \nabla f(x_k)\}$$

Η μέθοδος τερματίζει αν και μόνο αν καταλήξει σε στάσιμο σημείο ,οπότε και

$x^* = Prx\{x^* - s_k \nabla f(x_k)\}$, $\forall s > 0$

Οι περιορισμοί μας εκφράζονται από φράγματα στις μεταβλητές της f(x) με

$X = \{x \in \mathrm{R}^n : a_i \leq x_i \leq bi$, i = 1,2,…,n}

Τότε η i-συντεταγμένη του διανύσματος προβολής θα είναι :

$$[\mathrm{PrX}\{\mathrm{x}\}]\mathrm{i} == \begin{cases} \alpha i, xi \leq ai \\ \beta_i, xi \geq bi \\ \times i, ai \leq xi \leq \beta i \end{cases}$$

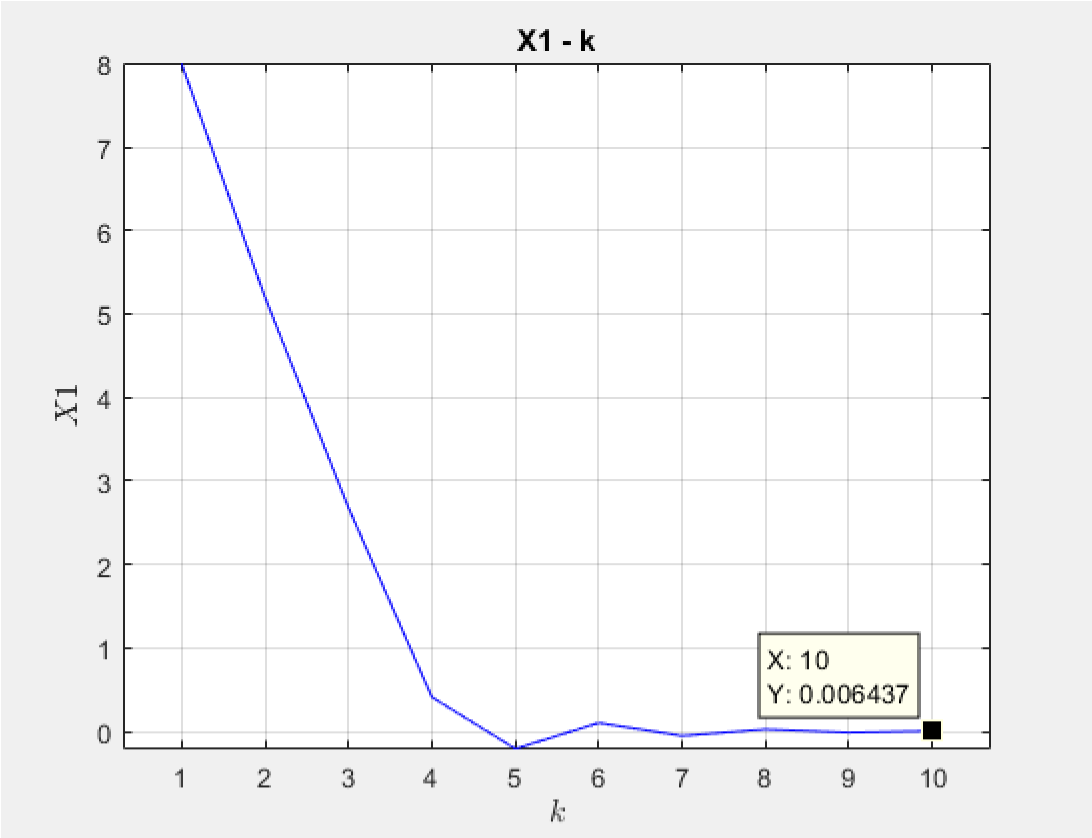
Μέσω της μεθόδου προβολής εξασφαλίζουμε την αναζήτηση ελαχίστου χρησιμοποιώντας την έννοια της προβολής για την αποφυγή της εξόδου από το X με την προϋπόθεση ότι το σημείο εκκίνησης είναι εφικτό δηλαδή x0 ανήκει στο X ,όπου ισχύει. Βρίσκω την προβολή στο $\bar{x}_i$ στο xi και λέω ότι είναι το επόμενο σημείο , μέχρις ότου να βρω το x*. Σε κάθε βήμα ελέγχω αν είμαι εντός του X , ακόμα και στο σύνορο. Αν όχι βρίσκω το πιο κοντινό σημείο του xi το $\bar{x}_i$ και συνεχίζω από εκεί. Έτσι οδηγούμαστε μέσα σε κάποιο αριθμό επαναλήψεων σε σύγκλιση εφόσον πληρούνται τα κριτήρια καλής λειτουργίας.

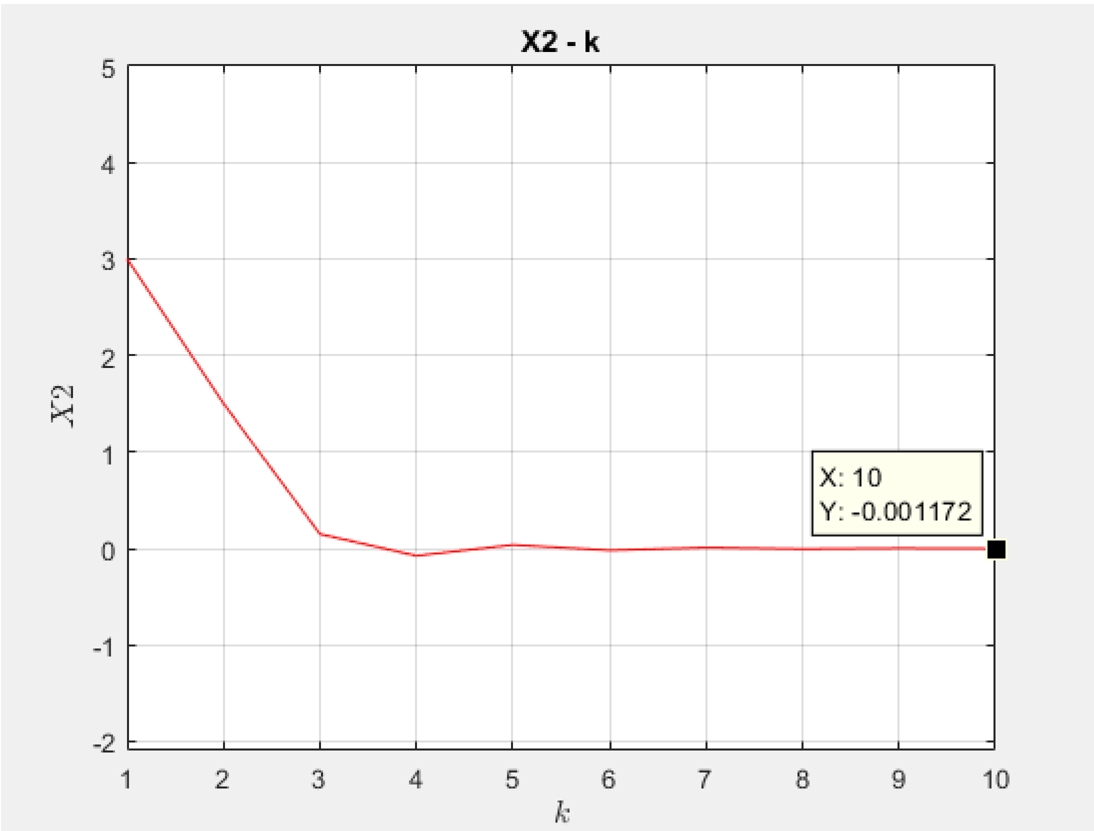
Θέμα 2

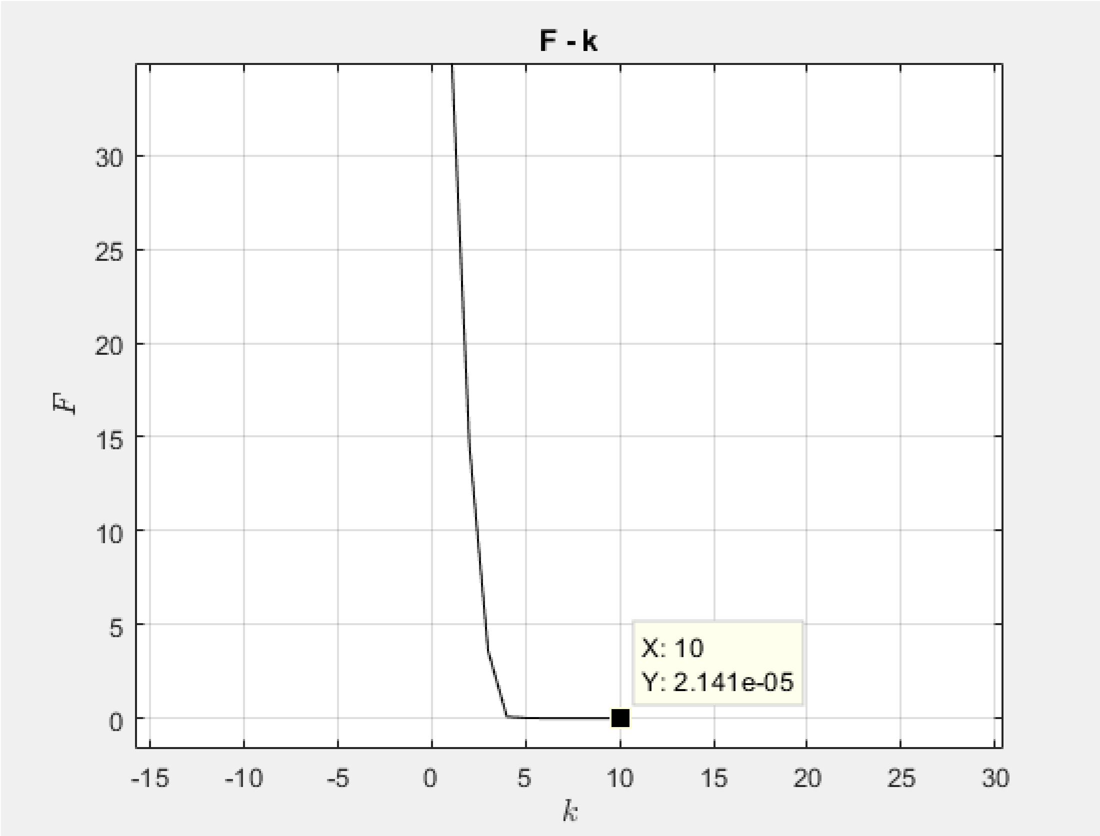
Να χρησιμοποιηθεί η μέθοδος μέγιστης καθόδου με προβολή, θεωρώντας $s_\kappa=15$, $\gamma_\kappa = 0.1$, σημείο εκκίνησης το (8,3) και ακρίβεια $\varepsilon = 0.01$. Τι παρατηρείτε σε σχέση με το (Θέμα 1, i) και το (Θέμα 1, iv); Είναι αναμενόμενο αυτό;

Matlab

Στο αρχείο TEX_BEL_THEMA_2_PERIORISMOI_DECENT.m είναι η επίλυση του 2 θέματος . Για τα $\gamma_\kappa = 0.1$, $s_k= 15$ για σημείο εκκίνησης το (8,3) και για ακρίβεια το $\varepsilon = 0.01$ έχουμε τα εξής plots και συμπεράσματα:


X1 - k
8
7
6
5
4
3
2
1
0
X1
1
2
3
4
5
6
7
8
9
10
k
X: 10
Y: 0.006437


X2 - k
5
4
3
2
1
0
-1
-2
X2
1
2
3
4
5
6
7
8
9
10
k
X: 10
Y: -0.001172



Είναι φανερό πως ο αλγόριθμος συγκλίνει σε πολύ λιγότερες επαναλήψεις σε σχέση με το ερώτημα i) του Θέματος 1 , καθώς σε 10 μόνο επαναλήψεις έχουμε φτάσει σε ικανοποιητικό αποτέλεσμα. Το διάστημα σύγκλισης είναι

[0.006437, -0.001171] πολύ κοντά στο 0 και fk = 0.0000214 . Επίσης όλα τα κριτήρια καλής λειτουργία πληρούνται όπως και στο ερώτημα i)του Θέματος 1. Το ερώτημα iv) του Θέματος 1 αποκλίνει από το ελάχιστο x* ενώ σε αυτή την περίπτωση περιορίζω τον αλγόριθμο να είναι μέσα στο X. Επίσης στο ερώτημα iv) του Θέματος 1 δεν πληρούνται όλα τα κριτήρια καλής λειτουργίας σε αντίθεση με την περίπτωση αυτή.

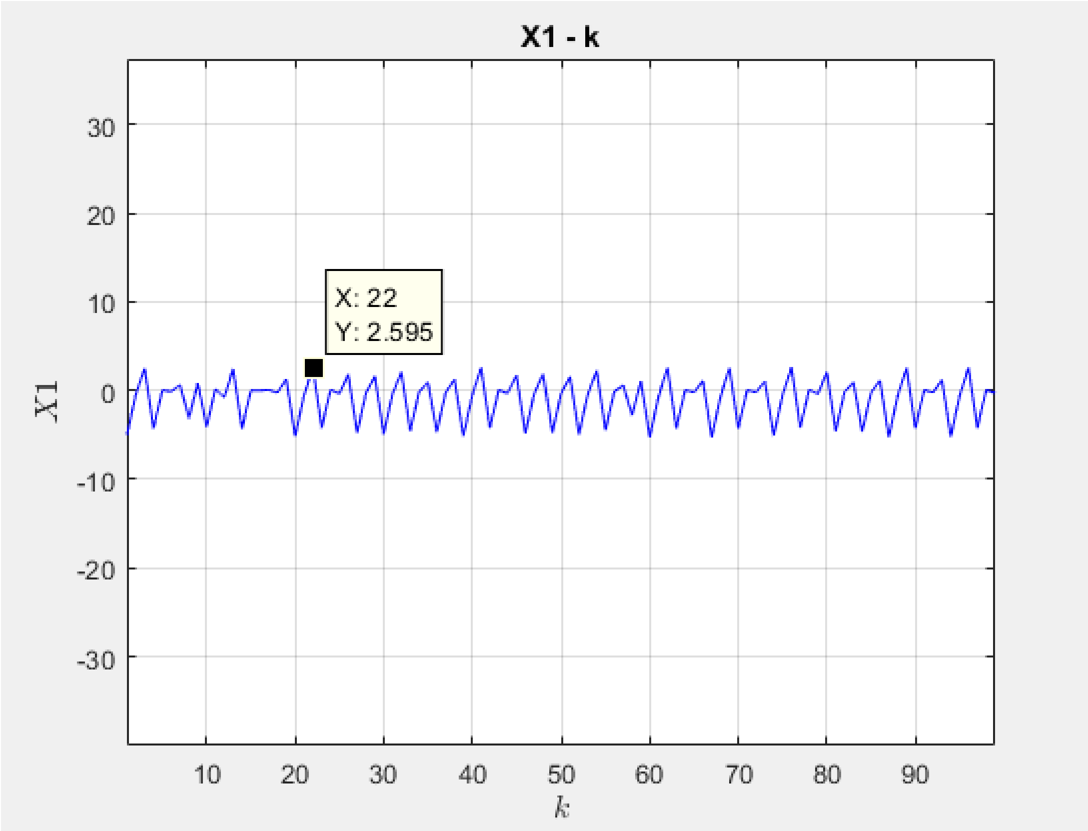
Θέμα 3

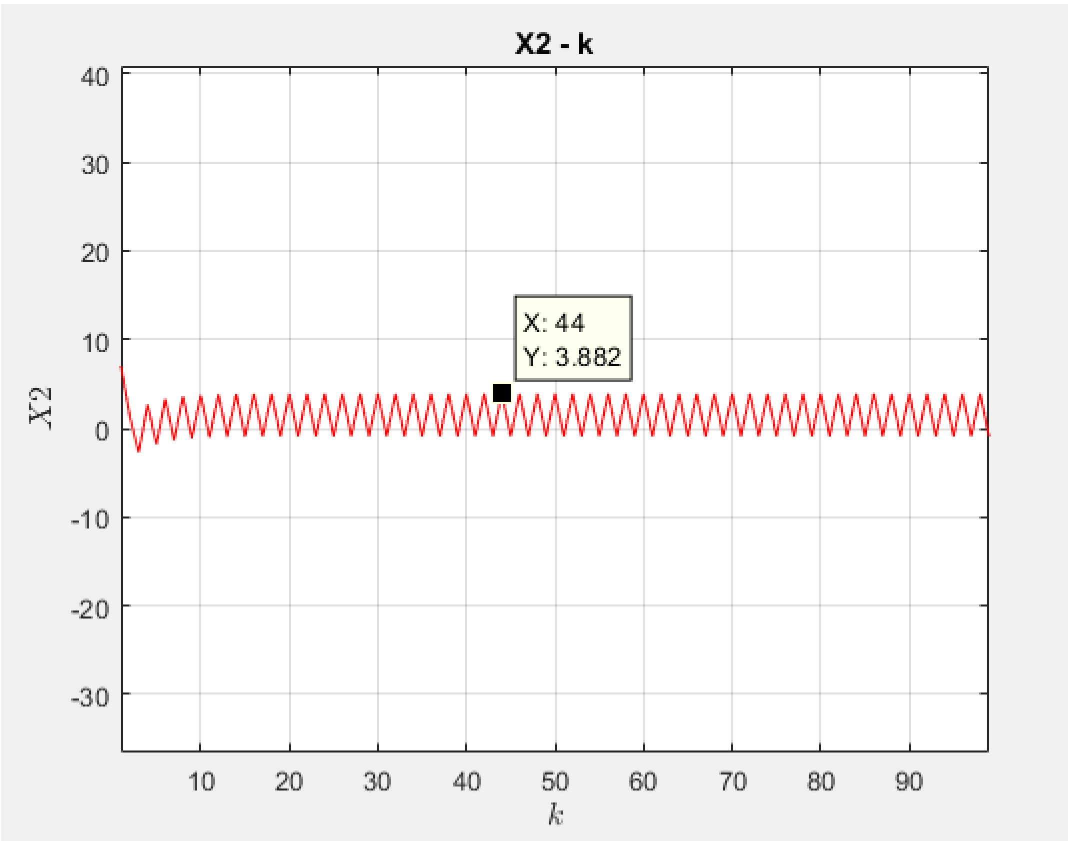
Να χρησιμοποιηθεί η μέθοδος μέγιστης καθόδου με προβολή, θεωρώντας

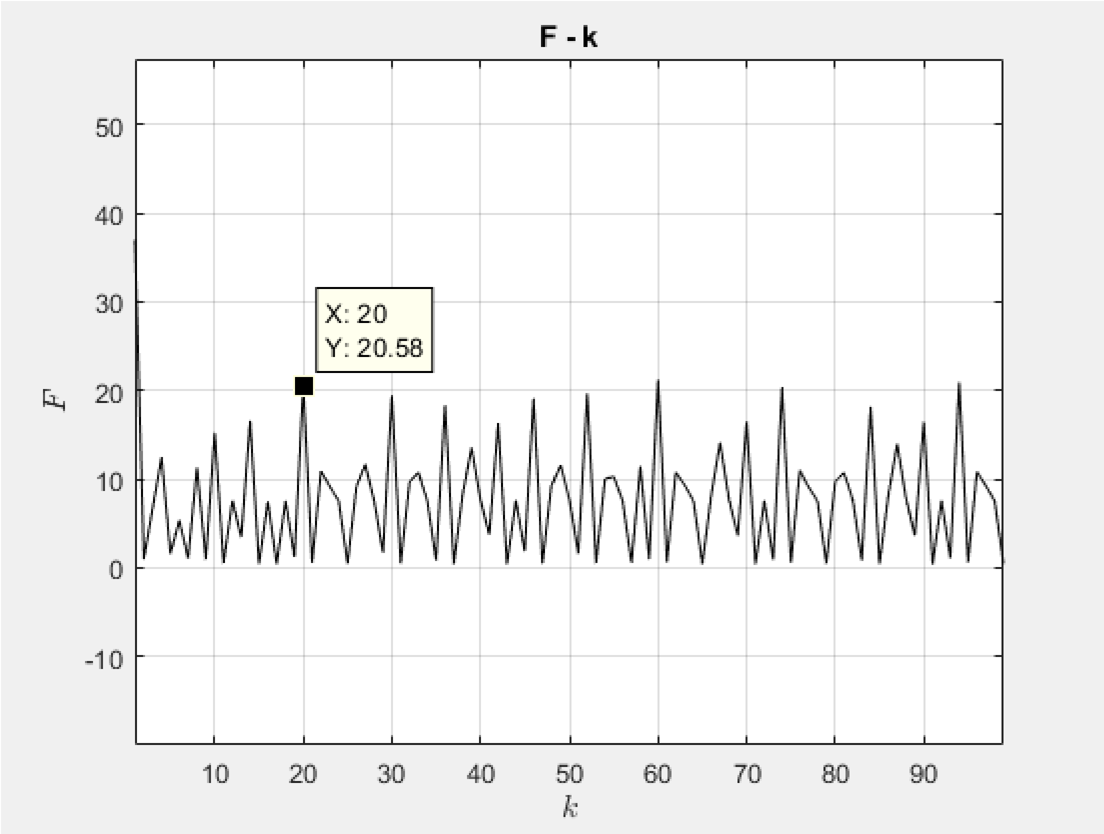
$s_κ=20$, $γ_κ = 0.3$, σημείο εκκίνησης το (-5,7) και ακρίβεια $ε = 0.02$. Τι παρατηρείτε σε σχέση με το (Θέμα 1, i) και το (Θέμα 1, iv); Είναι αναμενόμενο αυτό; Προτείνετε έναν απλό πρακτικό τρόπο ώστε η μέθοδος να συγκλίνει στο ελάχιστο.

Matlab

Στο αρχείο TEX_BEL_THEMA_3_PERIORISMOI_DECENT.m είναι η επίλυση του 3 θέματος . Για τα $γ_κ = 0.3$, $s_k= 20$ για σημείο εκκίνησης το (-5,7) και για ακρίβεια το $ε = 0.02$ έχουμε τα εξής plots και συμπεράσματα:

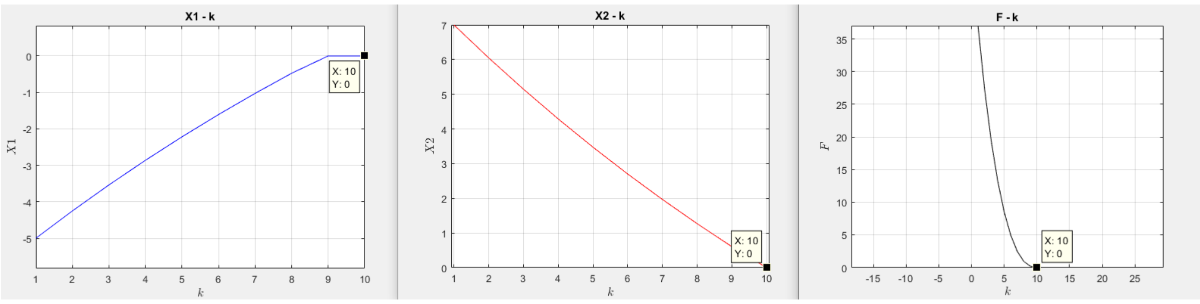

X1 - k
X: 22
Y: 2.595
X1
k
30
20
10
0
-10
-20
-30
10
20
30
40
50
60
70
80
90


X2 - k
X: 44
Y: 3.882
X2
k
40
30
20
10
0
-10
-20
-30
10
20
30
40
50
60
70
80
90

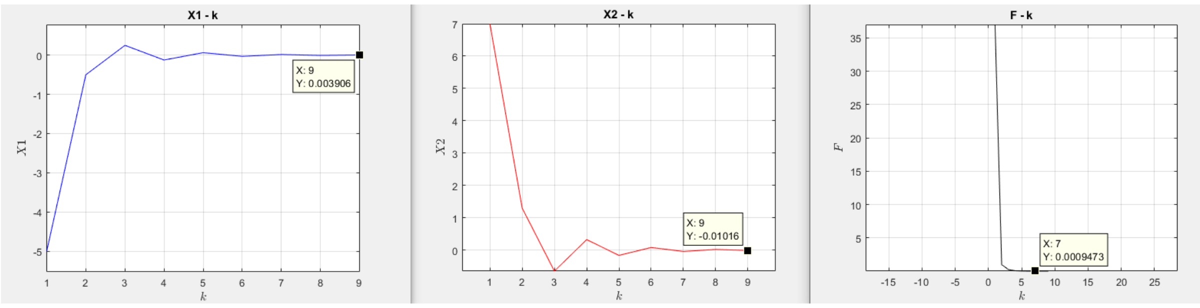


Είναι φανερό πως ο αλγόριθμος ταλαντώνεται γύρω από το [-5.241 , 2.537] για το x1 και [-0.8824,3.882] για το x2 αποκλίνοντας έτσι από το ελάχιστο x* καθώς δεν πληρούνται όλα τα κριτήρια καλής λειτουργία σε σχέση με το ερώτημα i) του Θέματος 1. Συγκεκριμένα δεν πληρείται το κριτήριο 1 δηλαδή το κριτήριο επαναληπτικής καθόδου που μας οδηγεί σε ολοένα και βελτιωμένες τιμές της f προς την ελαχιστοποίησή της. Στο ερώτημα iv) του Θέματος 1 ο αλγόριθμος δεν ικανοποιεί το κριτήριο 1,4 ενώ αποκλίνει προς κάποιο το άπειρο. Ένας απλός και πρακτικός τρόπος να συγκλίνει η μέθοδος προς το άπειρο είναι να επιλέγονται γκ και sk τέτοια ώστε να ισχύουν τα κριτήρια καλής λειτουργίας , οπότε πχ για γκ =0.05 και sκ = 20 ή για γκ = 0.3 και sk =**5** έχουμε σύγκλιση στο ελάχιστο x*=(0 , 0)

γκ =0.05 και sκ = 20



γκ = 0.3 και sk =5



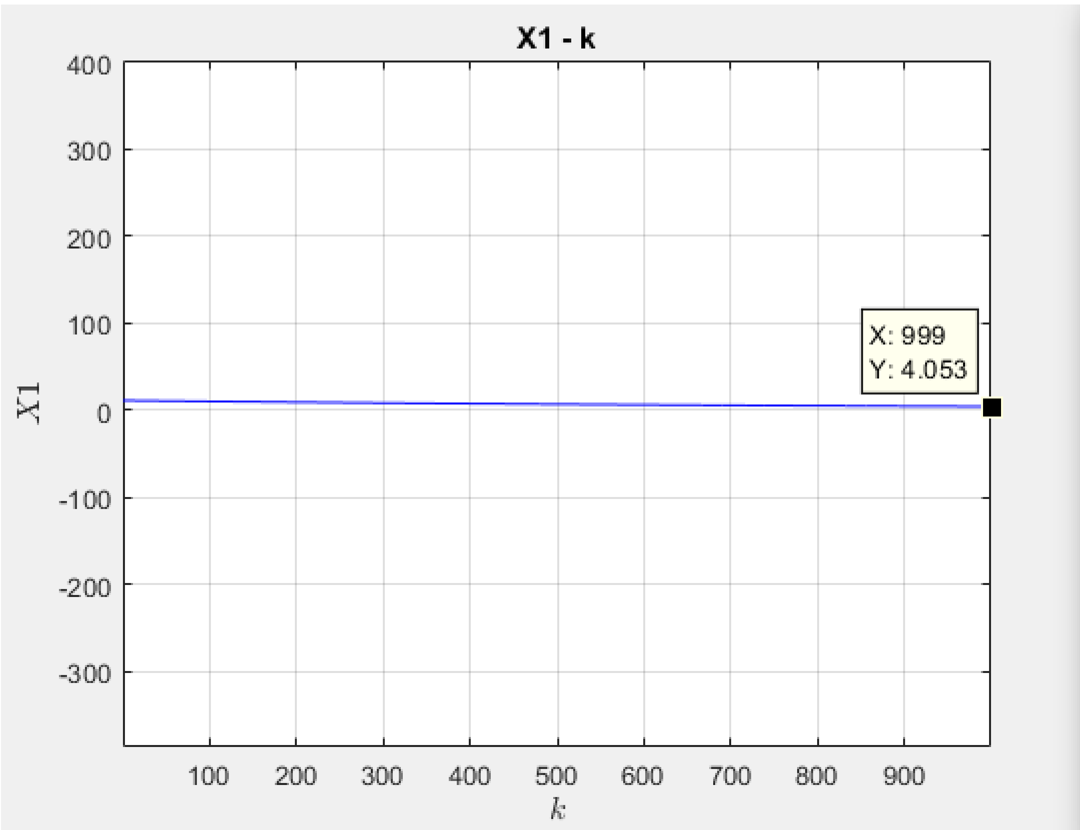
Θέμα 4

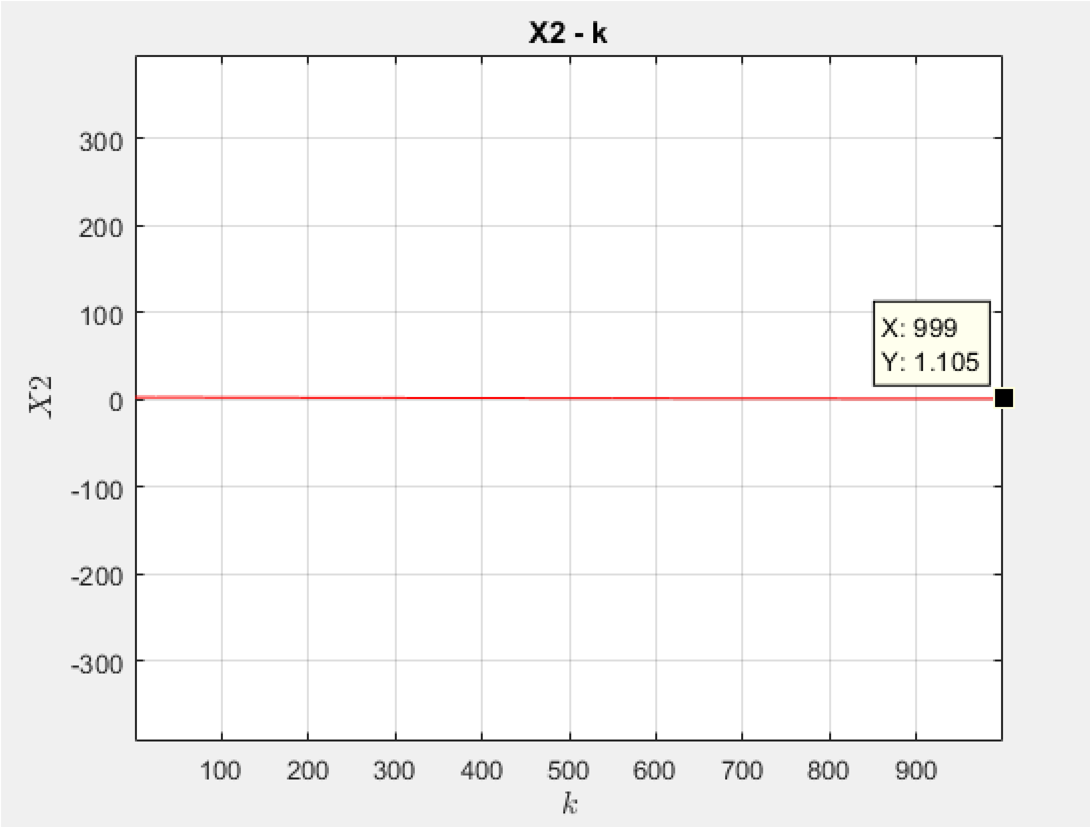
Θεωρούμε τη μέθοδο μέγιστης καθόδου με προβολή με sk=0.1 και γk = 0.01, σημείο εκκίνησης το (11,3) και ακρίβεια ε = 0.01. Σε αυτή την περίπτωση, έχουμε εκ των προτέρων (δηλ. πριν την εκτέλεση του αλγορίθμου) κάποια πληροφορία σχετικά με την ευστάθεια και τη σύγκλιση του αλγορίθμου; Να γίνει η εκτέλεση του αλγορίθμου. Τι παρατηρείτε;

Matlab

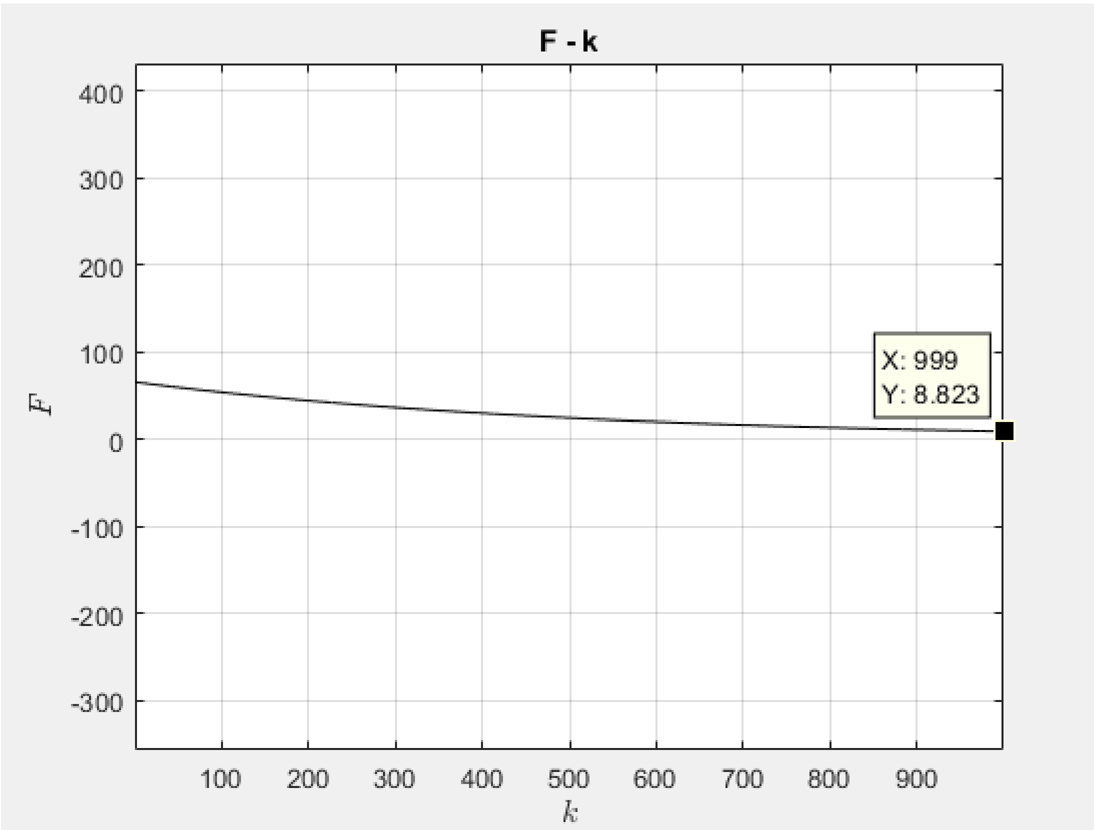
Στο αρχείο TEX_BEL_THEMA_4_PERIORISMOI_DECENT.m είναι η επίλυση του 4θέματος . Για τα γκ = 0.01 , sk= 0.1 για σημείο εκκίνησης το (11,3) και για ακρίβεια το ε = 0.01 έχουμε τα εξής plots και συμπεράσματα:

Αρχικά το γκ έχει πολύ μικρή τιμή , κοντά στο 0 , τότε είναι μάλλον απίθανο τα xk να κινηθούν γρήγορα προς το ελάχιστο x* , οδηγώντας τη μέθοδο σε μεγάλο αριθμό επαναλήψεων. Το κριτήριο 3 διασφαλίζει ότι το γκ δεν θα είναι πολύ μικρό. Ωστόσο σε αυτή την περίπτωση δεν πληρείται.





X2 - k
300
200
100
0
-100
-200
-300
X2
X: 999
Y: 1.105
100
200
300
400
500
600
700
800
900
k



F - k
400
300
200
100
0
-100
-200
-300
F
X: 999
Y: 8.823
100
200
300
400
500
600
700
800
900
k

Αφού εκτελέσουμε τον αλγόριθμο βλέπουμε πως η τιμή του x1 και του x2 συγκλίνει προς το 0 τόσο αργά που μέσα σε 1000 επαναλήψεις η τιμές τους έχουν μειωθεί στο ½ .Οπότε η σύγκλιση προς το 0 είναι πάρα πολύ χρονοβόρα διότι το βήμα γκ και sk είναι πολύ μικρά.

Συμπεράσματα

Στην παρούσα εργασία αποτυπώσαμε την λειτουργία της μεθόδου μέγιστης καθόδου (χωρίς περιορισμούς) και της μεθόδου μέγιστης καθόδου με προβολή (με περιορισμούς) και υπογραμμίσαμε τα κριτήρια καλής λειτουργίας που σύμφωνα με αυτά η αλγόριθμοι συγκλίνουν προς το ελάχιστο. Τέλος πρέπει να τονίσουμε πως η μέθοδος της μέγιστης καθόδου χωρίς περιορισμούς είναι μια αργή μέθοδος και υπολογιστικά αναποτελεσματική. Από την άλλη μεριά η ταχύτητα σύγκλισης της μεθόδου μεγίστης καθόδου με περιορισμούς δεν είναι μεγάλη όταν το X είναι ένα πολύεδρο, σχηματίζεται δηλαδή από γραμμικούς μετασχηματισμούς. Η κατάσταση βελτιώνεται όταν το X προσδιορίζεται από μη-γραμμικούς περιορισμούς ή όταν αυξάνει το πλήθος γραμμικών περιορισμών,ώστε να οδηγούμαστε σε πολύεδρα με μεγαλύτερο αριθμό εδρών.