Τεχνικές Βελτιστοποίησης

Τμήμα ΗΜΜΥ ΑΠΘ 7ο Εξάμηνο

2η εργασία

Κίτσιος Κωνσταντίνος 9182

Σε αυτήν την δεύτερη εργασία του μαθήματος θα ασχοληθούμε με την ελαχιστοποίηση βαθμωτών συναρτήσεων πολλών μεταβλητών, συγκεκριμένα συναρτήσεων $f:R^2\to R$ αλλά οι ίδιοι αλγόριθμοι γενικεύονται εύκολα και για συναρτήσεις $f:R^n\to R$. Θα χρησιμοποιήσουμε διάφορους αλγόριθμους και θα συγκρίνουμε κάποια ποιοτικά χαρακτηριστικά τους όπως ο αριθμός των επαναλήψεων μέχρι την σύγκλιση και την ευαισθησία στις παραμέτρους του κάθε αλγόριθμου. Οι δύο αντικειμενικές συναρτήσεις(objective functions) για αυτόν τον σκοπό θα είναι οι

- $f(x, y) = x^3\exp(-x^2 - y^4)$
- $g(x, y) = x^4 + y^2 - 0.2\sin(2\pi x) - 0.3\cos(2\pi x)$

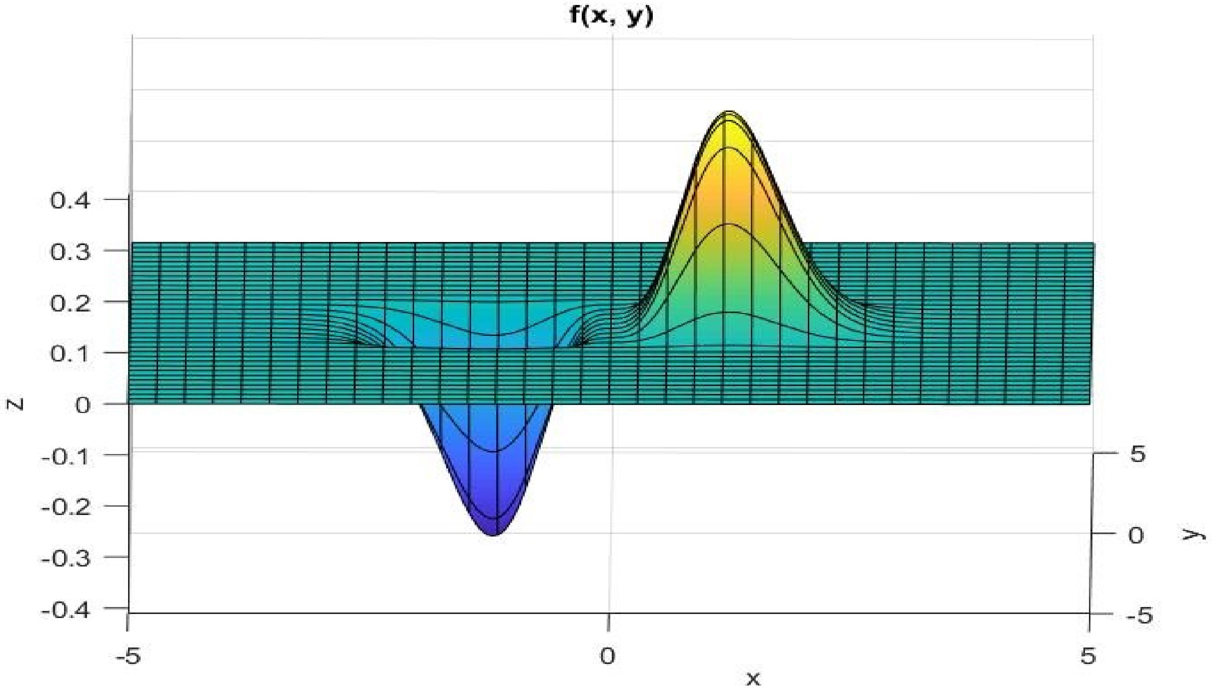
Για την ελαχιστοποίηση αυτών θα χρησιμοποιηθούν 5 διαφορετικοί αλγόριθμοι, οι 3 πρώτοι για την f και οι επόμενοι 2 για την g:

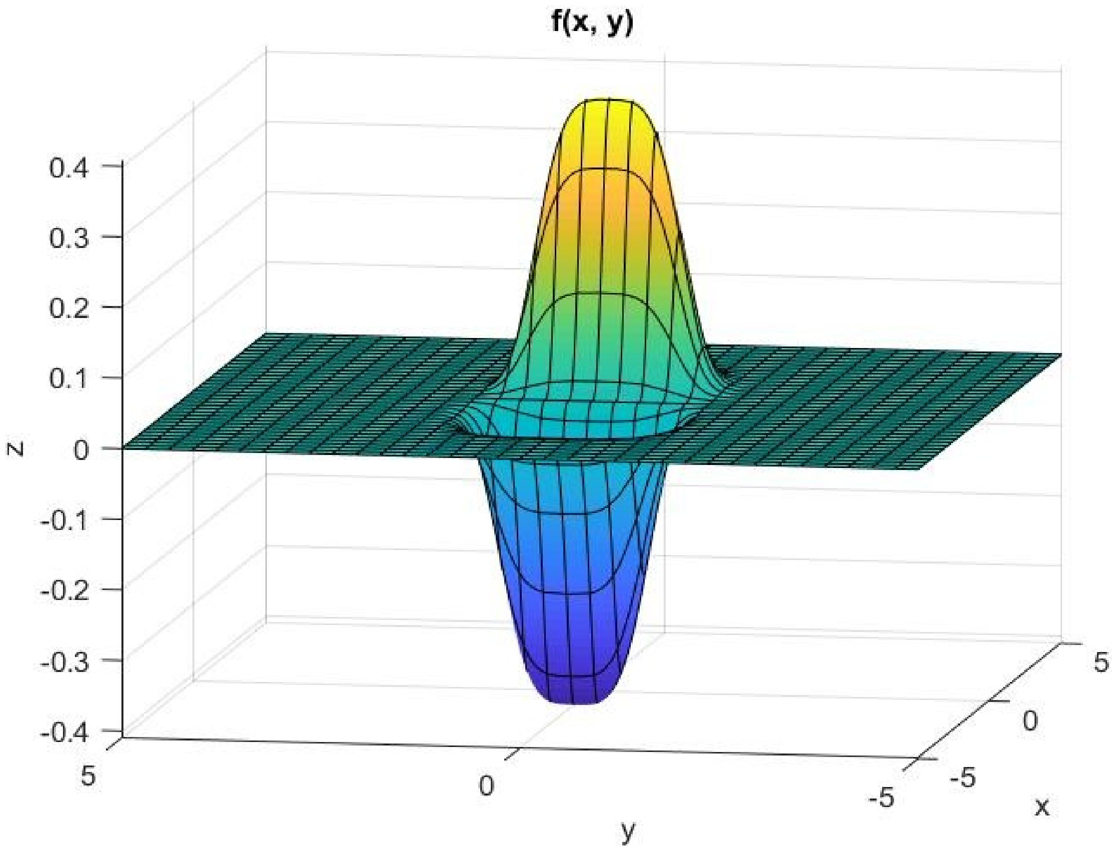
1. Μέγιστης καθόδου(Steepest descent)
2. Newton
3. Levenberg-Marquardt
4. Συζυγών κλίσεων
5. Σχεδόν Newton(Quasi-Newton)

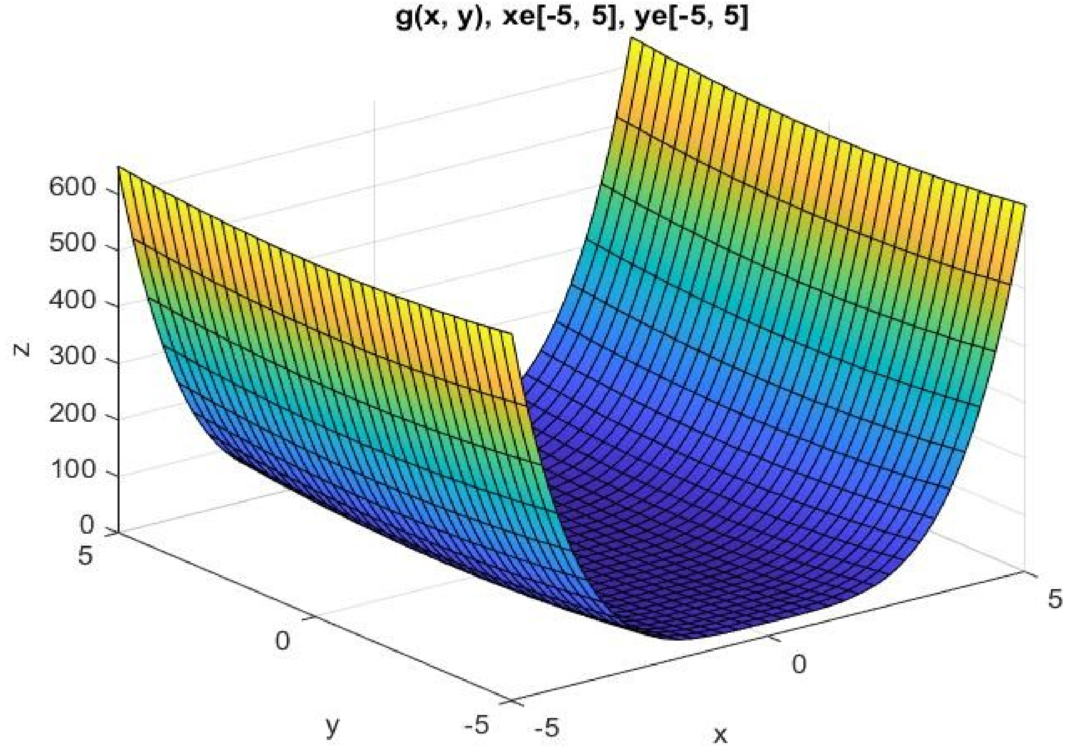
Όλοι οι αλγόριθμοι θα υλοποιηθούν σε Matlab.

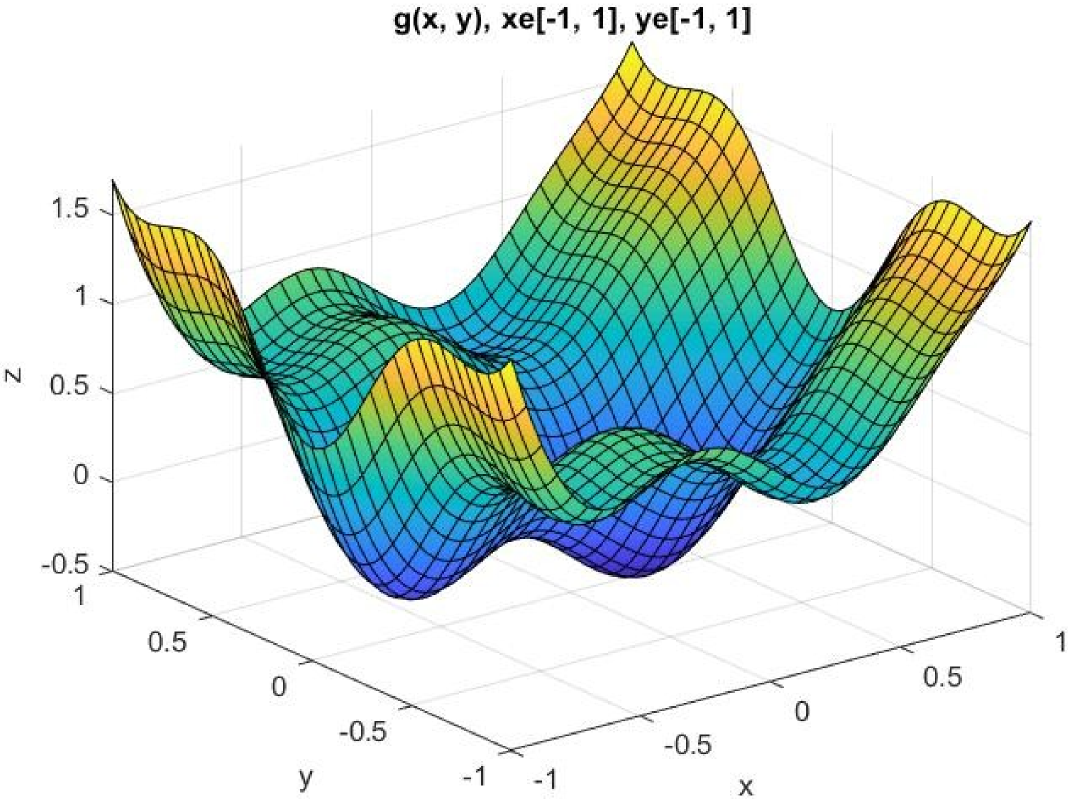
A) Γραφικές παραστάσεις των f, g.

Αρχικά θα σχεδιάσουμε τις γραφικές παραστάσεις των δύο συναρτήσεων για να πάρουμε μια γενική εικόνα της μορφής τους. Θα χρησιμοποιήσουμε την συνάρτηση του Matlab fsurf() στο διάστημα [-5, 5] για την f και στα διαστήματα [-5, 5] και [-1, 1] για την g.









Από τις γραφικές παραστάσεις συμπεραίνουμε ότι η f έχει ένα ολικό μέγιστο και ένα ολικό ελάχιστο ενώ η g αν και σε μεγάλη κλίμακα φαίνεται να έχει ένα ολικό ελάχιστο μόνο στο 0, εαν δούμε σε πιο μικρή κλίμακα όπως στην δεύτερη εικόνα φαίνεται ότι έχει πολλά τοπικά μέγιστα και ελάχιστα. Αυτό είναι λογικό λόγω της παρουσίας των ημιτόνων και συνημιτόνων στην έκφραση της. Τα σημεία των ολικών ακρότατων της f φαίνεται να είναι περίπου στο (-1.2, 0) για το ελάχιστο και (1.2, 0) για το μέγιστο. Επίσης η f έχει ένα σημείο καμπής στο (0, 0) όπου grad(f)=0 αλλά και Hess(f)=0 οπότε δεν αποτελεί ακρότατο.

B) Αλγόριθμος Μέγιστης Καθόδου(Steepest Descent)

Ο αλγόριθμος αυτός ψάχνει το ακρότατο στην κατεύθυνση του διανύσματος κλίσης, δηλαδή για τα διαδοχικά διανύσματα ισχύει

$x_{k+1} = x_k - \gamma_k$ grad(f(x_k)). Η συνθήκη τερματισμού που χρησιμοποιήσαμε είναι η νόρμα του διανύσματος κατεύθυνσης να είναι κάτω από μια τιμή ε=0.01. Για τους διάφορους συνδιασμούς αρχικού σημείου x_0 και τιμών γ_κ πήραμε τα παρακάτων αποτελέσματα.

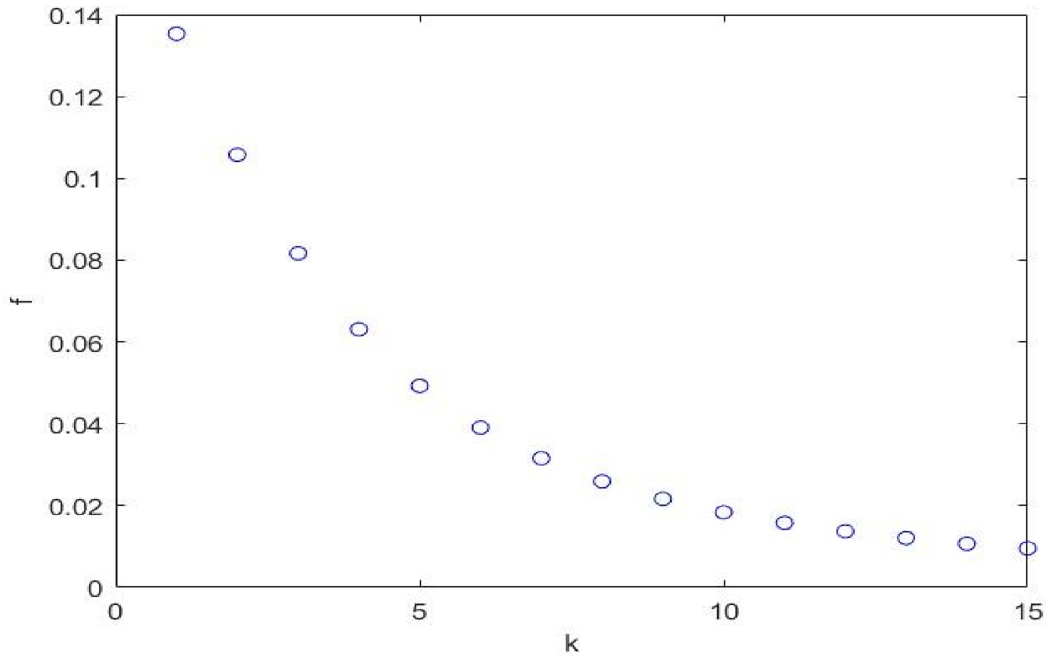
- **x_0 = (0, 0), γ_k minimizing f($x_k + \gamma_k$*d_k).**

Εδώ έχουμε σύγκλιση του αλγορίθμου στο σημειο x_0=(0, 0) από το πρώτο κίολας βήμα, δηλαδή δεν γίνεται κανένα update στην τιμή του x_k. Αυτό συμβαίνει γιατί όπως αναφέραμε και παραπάνω στο x_0 ισχύει grad(f)=0 οπότε ικανοποιείται η συνθήκη τερματισμού. Αυτό δίνει φυσικά λάθος αποτέλεσμα γιατί το x_0 δεν αποτελεί σημείο ελαχίστου αλλά σημείο καμπής της f γιατί Hess(f)=0 στο σημείο αυτό

- **x_0 = (1, 1), γ_k minimizing f($x_k + \gamma_k$*d_k).**

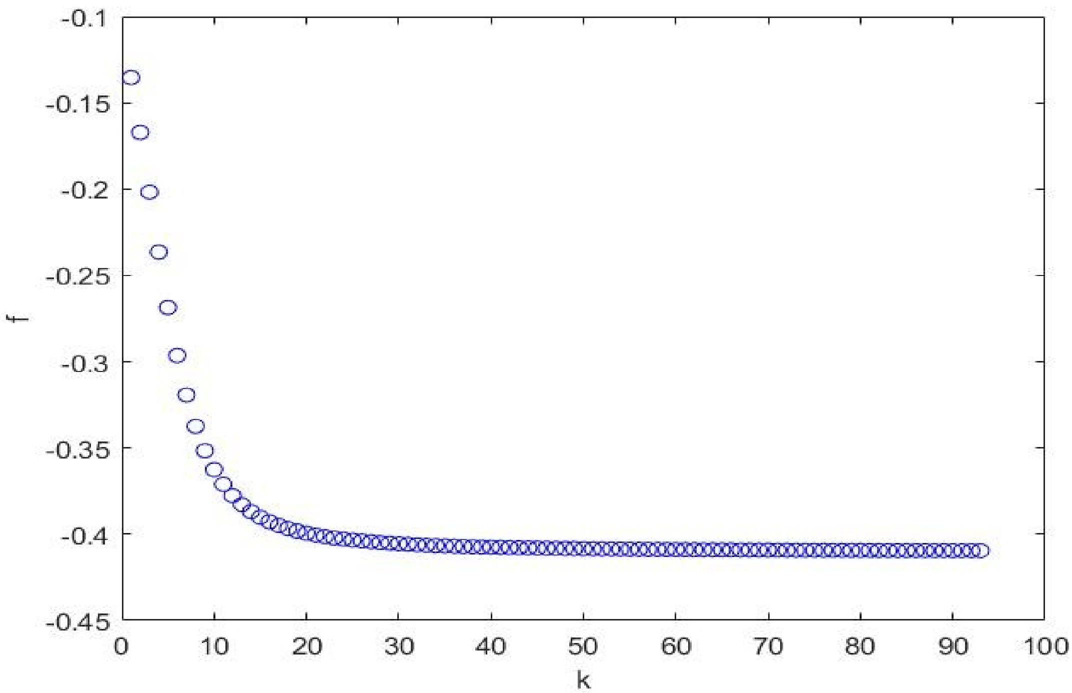
Ο αλγόριθμος συγκίνει μετά από κ=14 επαναλήψεις στο σημείο

(0.92, 1.37). Προφανώς ούτε αυτό το σημείο αποτελεί ελάχιστο της συνάρτησης και αυτό συμβαίνει γιατί το σημείο εκκίνησης είναι κοντά στο ολικό μέγιστο άρα ο αλγόριθμος τείνει να βρεί ελάχιστο σε εκείνη την γειτονεία, το οποίο μάλιστα και το πετυχαίνει όπως βλέπουμε στην γραφική παράσταση των τιμών της f σε κάθε επανάληψη:



- **x_0 = (-1, -1), γ_k minimizing f(x_k + γ_k*d_k)**.

Ο αλγόριθμος συγκλίνει σε k=92 επαναλήψεις στο σημείο (-1.22, -0.18) το οποίο και είναι όντως πάρα πολύ κοντά στο ολικό ελάχιστο όπως επιθυμούσαμε. Η διαφορά με τις άλλες δυο αρχικές τιμές x_0 εδώ είναι οτι ξεκινάμε ήδη σχετικά κοντά στο ελάχιστο και για αυτό και ο αλγόριθμος λειτουργεί ορθά. Η αντίστοιχη γραφική παράσταση φαίνεται παρακάτω.



<u>*Συμπέρασματα:*</u> Ο αλγόριθμος της μέγιστης καθόδου επηρεάζεται πολύ από το σημείο εκκίνησης x_0.

Στην συνέχεια θα εξετάσουμε τον ίδιο αλγόριθμο με τα ίδια αρχικά σημεία αλλά αυτήν την φορά η παράμετρος γ δεν θα επιλέγεται έτσι ώστε να ελαχιστοποιείται η παράσταση $f(x_k + \gamma_k * d_k)$ αλλά θα είναι σταθερή και ίση με 0.01 σε όλα τα βήματα.

- **x_0 = (0, 0), γ_k=0.01=constant**

Όμοια με την πρώτη περίπτωση εδώ έχουμε λανθασμένη σύγκλιση στο αρχικό σημείο (0, 0) σε 1 βήμα λόγω του μηδενικού διανύσματος κλίσης στο σημείο αυτό.

- **x_0 = (1, 1), γ_k=0.01=constant**

Εδώ έχουμε σύγκλιση του αλγορίθμου σε κ=889 βήματα, περίπου 10πλάσια από την πρώτη περίπτωση ενώ το σημείο σύγκλισης είναι σχεδόν το ίδιο, δηλαδή το (0.90, 1.58) το οποίο όπως αναφέραμε και παραπάνω είναι λανθασμένο αποτέλεσμα.

- **x_0 = (-1, -1), γ_k=0.01=constant**

Εδώ βλέπουμε σύγκλιση του αλγορίθμου σε κ=922 βήματα, περίπου 10πλάσια από την περίπτωση του μεταβαλλόμενου γ_κ ενώ το σημείο σύγκλισης είναι σχεδόν το ίδιο, δηλαδή το (-1.22, -0.18).

Επαναλαμβάνουμε τις παραπάνω μετρήσεις αλλα με το γ_κ να επιλέγεται σύμφωνα με τον κανόνα Armijo και βλέπουμε ότι τα αποτελέσματα είναι ακριβώς τα ίδια με την περίπτωση του σταθερού γ, δηλαδή σε κάθε επανάληψη ο κανόνας Armijo δίνει $\gamma_\kappa=\gamma_0$=0.01.

<u>*Συμπεράσματα*</u>: Η μεταβολή της παραμέτρου γ_κ μπορεί να μην επηρεάζει το σημείο σύγκλισης αλλά επηρεάζει σημαντικά την ταχύτητα του αλγορίθμου, δίνοντας επιτάχυνση μέχρι και x10.

Γ) Αλγόριθμος Newton

Ο αλγόριθμος Newton στηρίζεται στην Hessian(f) για τον υπολογισμό τον διανυσμάτων x_k και αυτό έχει νόημα μόνο εάν αυτή είναι θετικά ορισμένη στα x_k. Για την δοσμένη συνάρτηση f(x, y) και τα συγκεκριμένα αρχικά σημεία ο αλγόριθμος δεν μπορεί να λειτουργήσει καθώς η Hessian(f) δεν είναι θετικά ορισμένη στα σημεία αυτά. Αυτό μπορούμε εύκολα να το επαληθεύσουμε με έναν έλεγχο στις ιδιοτιμές του πίνακα οι οποίες πρέπει να είναι όλες θετικές, κάτι που δεν ισχύει.

Δ) Αλγόριθμος Levenberg-Marquardt

Ο αλγόριθμος αυτός είναι ουσιαστικά ένας συνδιασμός της μεθόδου Newton και Μέγιστης Καθόδου και τα x_k υπολογίζονται στην κατεύθυνση του διανύσματος $\mathbf{d_{k+1}=[Hess(f(x_k, y_k) + \mu_k I]^{-1}*grad(f(x_k, y_k))}$. Έτσι για μεγάλο μ_κ έρχεται πολύ κοντά στον αλγόριθμο μέγιστης καθόδου ενώ για μικρό μ_κ στον αλγόριθμο Newton. Η χρησιμοτητα του μ_κ είναι τα κάνει τον πίνακα θετικά ορισμένο σε περίπτωση που δεν είναι, και μια τιμή που το εγγυάται αυτό σε κάθε επανάληψη είναι η μέγιστη ιδιοτιμή του πίνακα και αυτήν θα χρησιμοποιήσουμε στην υλοποιησή μας. Για τα 3 δοσμένα αρχικά σημεία πήραμε τα ακόλουθα αποτελέσματα.

- **x_0 = (0, 0), γ_k=0.01=constant**

Ξανά έχουμε σύγκλιση σε 1 βήμα στο σημείο (0, 0) για τους λόγους που αναλύθηκαν παραπάνω.

- **x_0 = (1, 1), γ_k=0.01=constant**

Σε αυτήν την περίπτωση ο αλγόριθμος συγκλίνει σε κ=894 επαναλήψεις στο σημείο (0.98, 1.58), δηλαδή δίνει περίπου ίδια αποτελέσματα με τον αλγόριθμο μέγιστης καθόδου ενώ το σημείο σύγκλισης είναι φυσικά λανθασμένο πάλι όπως εξηγήσαμε παραπάνω.

- **x_0 = (-1, -1), γ_k=0.01=constant**

Εδώ έχουμε σύγκλιση στο σημείο (-1.21, -0.39) σε κ=486 επαναλήψεις αλλά με σταθερά τερματισμού ε=0.1 αντί για ε=0.01 που χρησιμοποιείται στις υπόλοιπες περιπτώσεις γιατί εδώ δεν είχαμε σύγκλιση για ε=0.01.

- **x_0 = (0, 0), γ_k minimizing $f(x_k + \gamma_k*d_k)$.**

Εδώ έχουμε σύγκλιση στο σημείο (0, 0) για τους προαναφερθέντες λόγους

- **x_0 = (1, 1), γ_k minimizing $f(x_k + \gamma_k*d_k)$.**

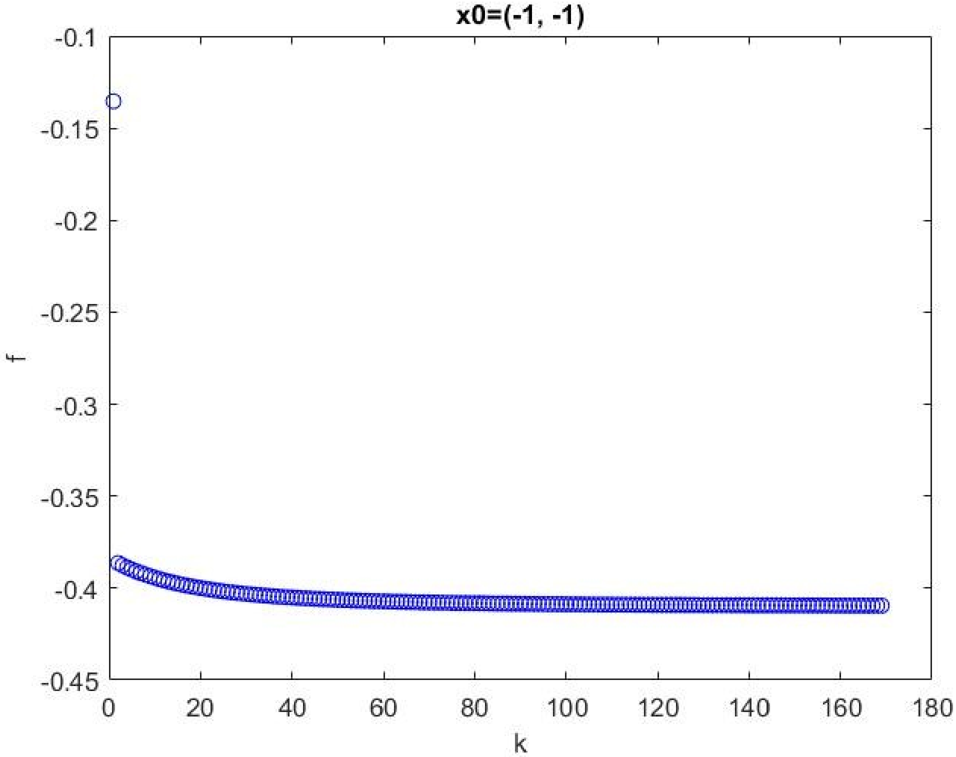
Έχουμε ξανά σύγκλιση στο σημείο (1.03, 1.58) μετά από κ=89 επαναλήψεις, παρόμοια δηλαδή με την μέθοδο μέγιστης καθόδου.

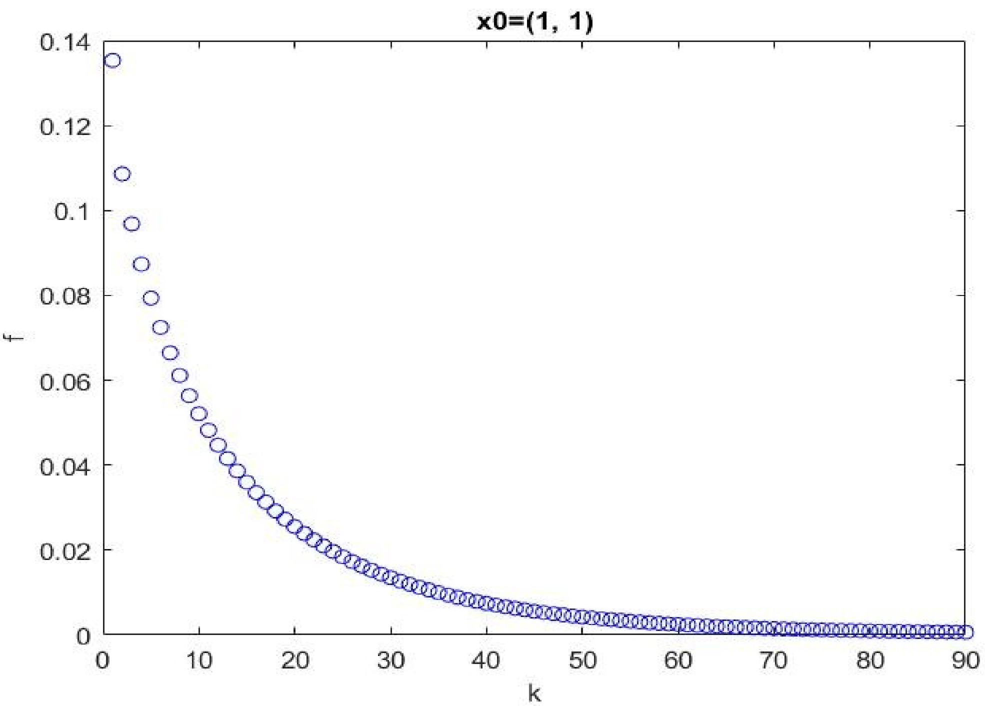
- **x_0 = (-1, -1), γ_k minimizing $f(x_k + \gamma_k*d_k)$.**

Εδώ έχουμε σύγκλιση στο σημείο (-1.22, -0.18) σε κ=164 επαναλήψεις,

δηλαδή περίπου 1.5 φορές περισσότερο από τον αλγόριθμο μέγιστης καθόδου.

Επαναλαμβάντας τις τελευταίες μετρήσεις με τον κανόνα του Armijo για το γ παίρνουμε ακριβώς τις ίδες μετρήσεις. Παρακάτω φαίνονται οι γραφικές παραστάσεις της f συναρτήσει του κ για αρχικά σημεία (-1, -1) και (1, 1) αντίστοιχα, με το γ να προκύπτει απο την ελαχιστοποίηση της παράστασης $\mathbf{f(x_k + \gamma_k{*}d_k)}$.





E) Αλγόριθμος συζυγών κλίσεων

Ο αλγόριθμος αυτός είναι μια ειδική περίπτωση της οικογένειας αλγορίθμων **συζυγών κατευθύνσεων** στον οποίο για την παραγωγή των Q-συζυγών διανυσμάτων χρησιμοποιούνται τα γραμμικά ανεξάρτητα διανύσματα z_k**=-grad(f(x_k)).** Εαν η συνάρτηση f είναι τετραγωνική τότε ο αλγόριθμος συγκλίνει το πολύ σε n βήματα όπου n η διάσταση των διανυσμάτων χ, ενώ για συναρτήσεις μη τετραγωνικές δεν ισχύει το ίδιο.

Επιλέγοντας το γ_κ έτσι ώστε να ελαχιστοποιείται η παράσταση

$f(x_k + \gamma_k*d_k)$ λαμβάνουμε τα παρακάτων αποτέλεσματα με την μέθοδο των συζυγών κλίσεων για την ελαχιστοποίηση της συνάρτησης g(x, y):

- **x_0=(0, 0)**

Σε αυτήν την περίπτωση έχουμε σύγκλιση του αλγορίθμου στο σημείο (0.23, 0) σε κ=2 βήματα.

- **x_0=(1, 1)**

Σε αυτήν την περίπτωση έχουμε σύγκλιση του αλγορίθμου στο σημείο (-0.61, 0) σε κ=5 βήματα.

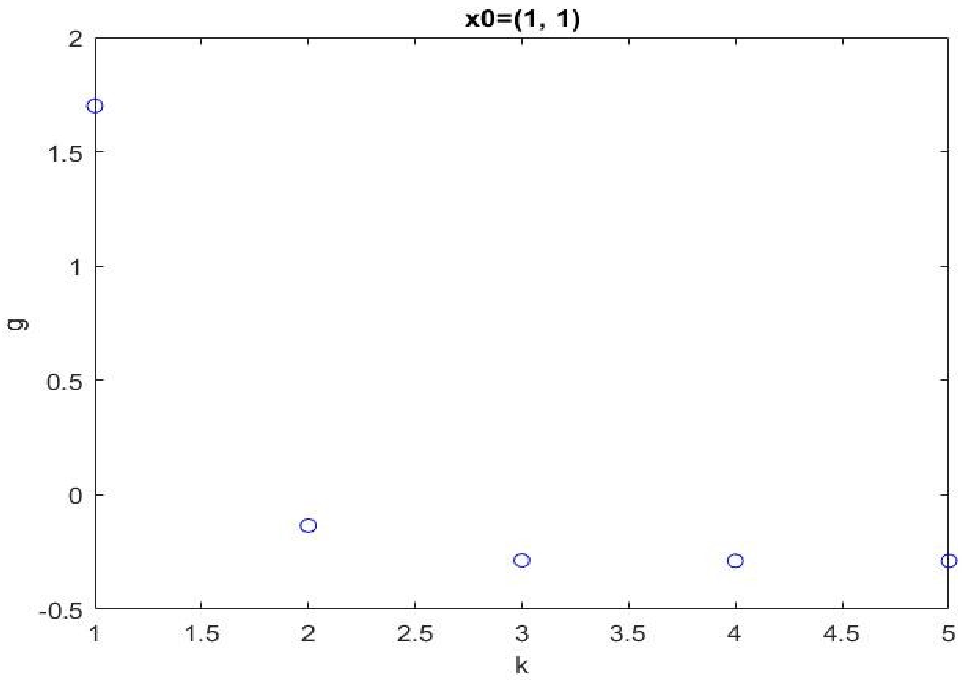
- x_0**=(-0.6, -0.6)**

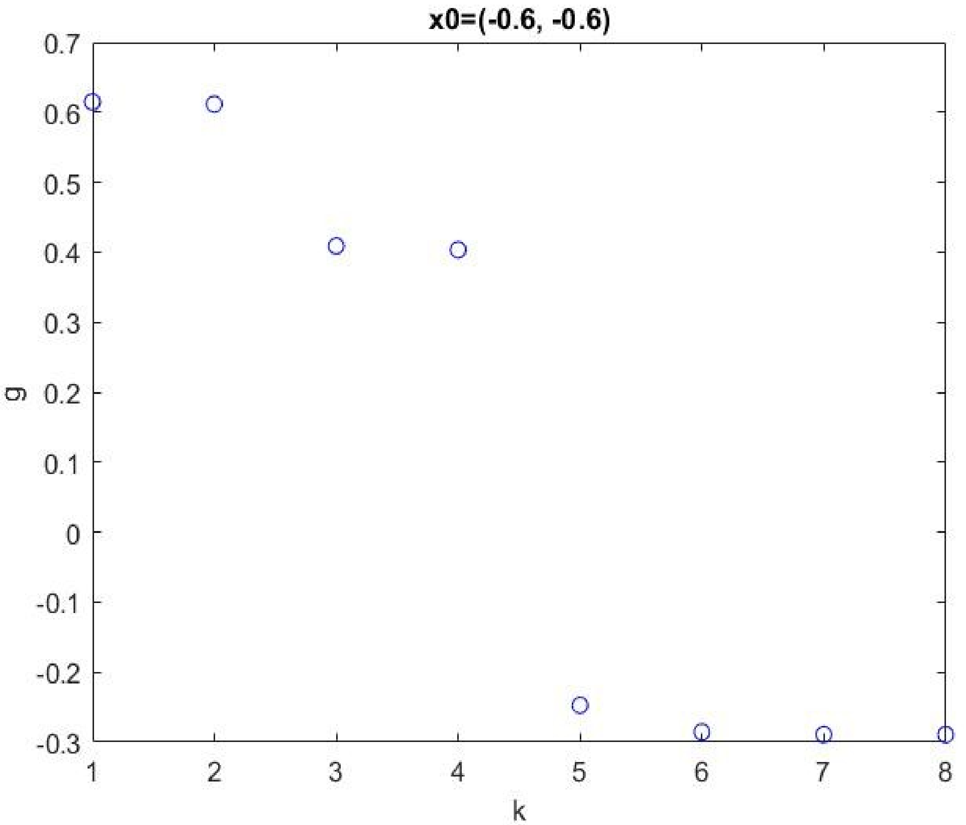
Σε αυτήν την περίπτωση έχουμε σύγκλιση του αλγορίθμου στο σημείο (-0.61, 0) σε κ=9 βήματα.

Συμπεράσματα: Βλέπουμε πως ο αλγόριθμος συζυγών κλίσεων συγκλίνει πάρα πολύ γρήγορα σε σχέση με τους υπόλοιπους αλγόριθμους που εξετάσαμε, το οποίο επιβεβαιώνεται και θεωρητικά. Επίσης, τα σημεία σύγκλισης είναι διαφορετικά για διαφορετικά σημεία εκκίνησης και αυτί συμβαίνει λόγω των πολλών τοπικών ακροτάτων της συνάρτησης g(x, y) όπως φαίνεται και στην γραφική της παράσταση στην αρχή της αναφοράς. Αυτό καθιστά πολύ δύσκολο τον υπολογισμό του ολικού ελαχίστου καθώς ο αλγόριθμος εγκλωβίζεται γρήγορα σε κάποιο τοπικό ακρότατο.

Παρακάτω φαίνονται οι γραφικές παραστάσεις της g συναρτήσει του κ για

x_0=(1, 1) και **x_0=(-0.6, -0.6)**


x0=(1, 1)
g
k
2
1.5
1
0.5
0
-0.5
1
1.5
2
2.5
3
3.5
4
4.5
5


x0=(-0.6, -0.6)
g
k
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
1
2
3
4
5
6
7
8

ΣΤ) Αλγόριθμος σχεδόν Newton(Quasi-Newton)

Ο αλγόριθμος αυτός χρησιμοποιεί μια προσσέγγιση του Εσσιανού πίνακα σε κάθε σημείο ώστε να αποφευχθούν οι n*(n+1)/2 πράξεις σε κάθε επανάληψη για τον ακριβή υπολογισμό του, προσπαθόντας ταυτόχρονα να εκμεταλευτεί την ταχύτητα σύγκλισης της μεθόδου Newton. Η προσσέγγιση αυτήν γίνεται μέσα από μακροσκελείς τύπους οι οποίοι δνε θα αναπτυχθούν εδώ για λόγους απλότητας. Θα δώσουμε τα αποτελέσματα αυτού του αλγορίθμου πάνω στην αντικειμενική συνάρτηση g(x, y) για τα 3 σημεία εκκίνσης της εκφώνησης, με τις παραμετρους γ_κ να ελαχιστοποιει την παράσταση $\mathbf{f(x_k + \gamma_k{*}d_k)}$ και ξ_κ=1. Ο αλγόριθμος σταματάει εαν ξεπεράσουμε το n, όπου n η διάσταση των διανυσμάτων x, στην περιπτωσή μας δηλαδή εκτελεί το πολύ 3 επαναλήψεις.

- $\mathbf{x_0=(0, 0)}$

Έχουμε σύγκλιση στο σημείο (0.26, 0.32) μετά από κ=3 επαναλήψεις

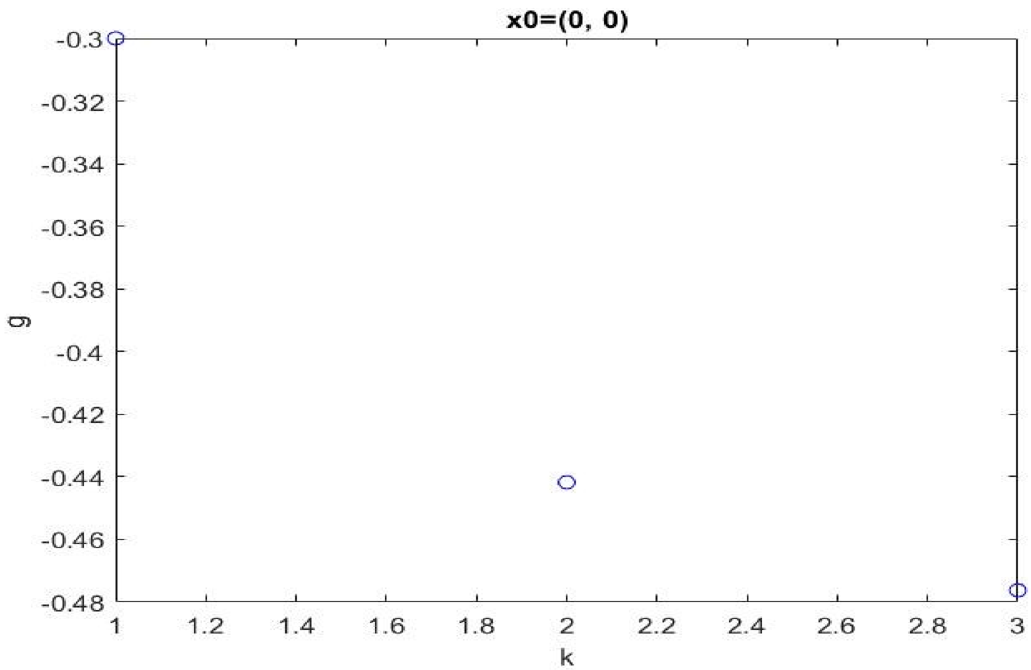
- $\mathbf{x_0=(1, 1)}$

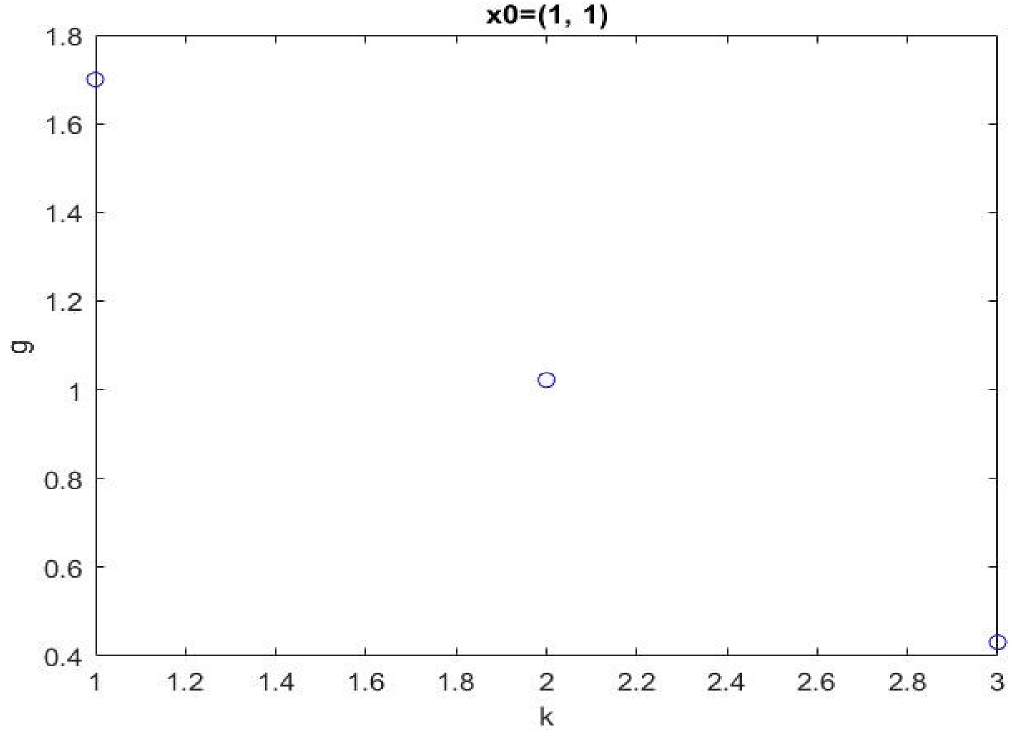
Έχουμε σύγκλιση στο σημείο (0.07, 0.04) μετά από κ=3 επαναλήψεις

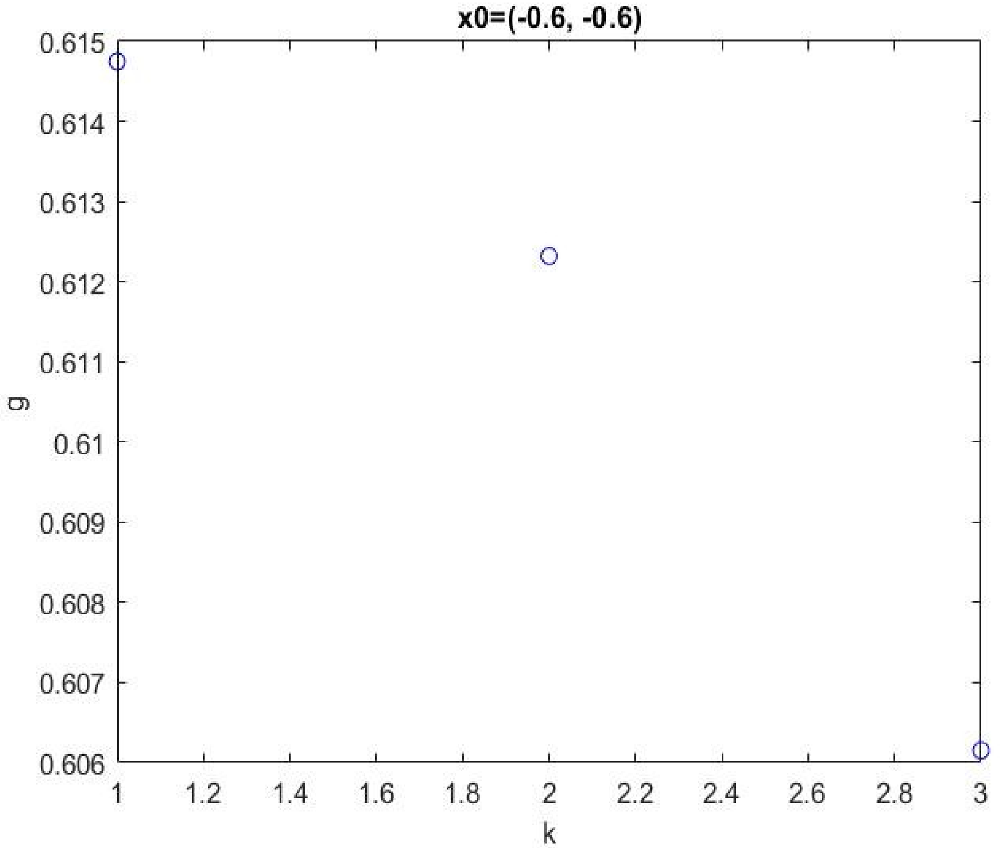
- $\mathbf{x_0=(-0.6, -0.6)}$

Έχουμε σύγκλιση στο σημείο (-0.48, -0.42) μετά από κ=3 επαναλήψεις

Οι γραφικές παραστάσεις της g συναρτήσει του κ φαίνονται παρακάτω για τις 3 περιπτώσεις:







Συμπεράσματα: Βλέπουμε ότι για 3 διαφορετικά σημεία εκκίνσης ο αλγόριθμος δίνει 3 διαφορετικές λύσεις, από τις οποίες βέλτιστη φαίνεται να είναι η πρώτη για x_0=(0, 0). Υπάρχει δηλαδή το πρόβλημα εγκλωβισμού σε τοπικά ακρότατα τα οποία υπάρχουν λόγω της μορφής της g(x, y). Η ταχύτητα σύγκλισης της μεθόδου σχεδόν Newton είναι στην χειρότερη περίπτωση σταθερή και ίση με την διάσταση του χώρου που εργαζόμαστε, το οποίο είναι και το κύριο πλεονέκτημά της.

Ζ) Τελικά συμπεράσματα

Μέσα από αυτήν την εργασία υλοποιήσαμε 5 διαφορετικούς αλγορίθμους ελαχιστοποίησης συναρτήσεων πολλών μεταβλητών και βγάλαμε χρήσιμα συμπεράσματα για κάθε έναν από αυτούς σχετικά με την ταχύτητα σύγκλισης και την ευαισθησία στις παραμέτρους.

Συγκεκριμένα, όλοι οι αλγόριθμοι που εξετάσαμε φαίνεται να είναι πολύ **ευαίσθητοι ως προς την επιλογή του αρχικού σημείο x_0.** Για διαφορετικά σημεία εκκίνησης λάβαμε διαφορετικά αποτέλεσματα το οποίο φανερώνει μια αστάθεια των επαναληπτικών αλγορίθμων. Μια λύση σε αυτό το πρόβλημα θα ήταν να εκτελούσαμε τον αλγόριθμο πολλές φορές με διαφορετικά σημεία εκκίνσης και να κρατούσαμε το αποτέλεσμα με την μικρότερη τιμή.

Επιπλέον, είδαμε ότι η τιμή του $γ_κ$ επηρεάζει σημαντικά την ταχύτητα σύγκλισης των αλγορίθμων και η σωστή επιλογή του προσφέρει επιτάχυνση πάνω απο 10 φορές. Αντιθέτως, το σημείο σύγκλισης δεν επηρεάζεται ιδιαίτερα εκτός από περιπτώσεις που η τιμή του $γ_κ$ είναι πολύ μεγάλη οπότε ο αλγόριθμος αποκλίνει.

Ακόμη είδαμε την αδυναμία των επαναληπτικών αλγορίθμων να αποφύγουν τα τοπικά ελάχιστα καθώς και τα σημεία με μηδενικό διάνυσμα κλίσης που δεν είναι ελάχιστα.

Τέλος από θέμα ταχύτητας σύγκλισης αλλά και ελαχιστοποίησης υπολογισμών φαίνεται πως οι τροποποιημένες μέθοδοι κλίσης και η μέθοδος σχεδόν Newton υπερτερούν των υπολοίπων, ενώ η μέθοδος Newton μπορεί να δίνει αποτέλεσμα σε μια επανάληψη για τετραγωνικές συναρτήσεις αλλά είναι ασταθής καθώς απαιτεί θετικά ορισμένο Εσσιανό πίνακα, κάτι που δεν συμβαίνει συχνά.