Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

6/12/2022

2η Εργαστηριακή Άσκηση. Ελαχιστοποίηση συνάρτησης δύο μεταβλητών με χρήση παραγώγων.

Μάθημα: Τεχνικές Βελτιστοποίησης
Καθηγητής: Γεώργιος Ροβιθάκης
Όνομα: Τσιμπλιαρίδης Νικόλαος
ΑΕΜ: 9652
E-mail: tenikola@ece.auth.gr

Χειμερινό εξάμηνο

2022-2023

Πίνακας περιεχομένων

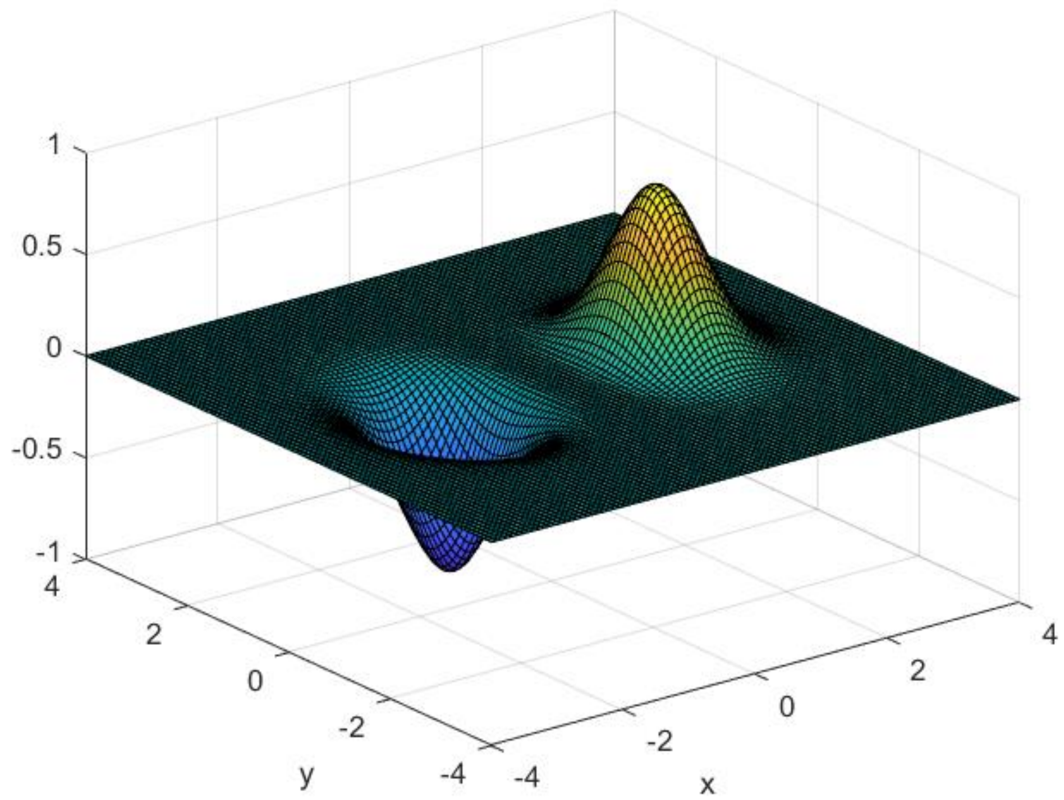
1η Εργαστηριακή Άσκηση

Ερώτημα 1ο: Σχεδίαση της αντικειμενικής συνάρτησης

Η αντικειμενική συνάρτηση που θα μελετήσουμε είναι η:

$$f\,(\,x, y) \;= x^5 \cdot e^{-x^2 - y^2}$$

Αρχικά θα την σχεδιάσουμε για να πάρουμε μία γενική εικόνα για την μορφή της:



Αξίζει να σημειωθεί ότι εμφανίζει ελάχιστο στο $\left(-\sqrt{\dfrac{5}{2}},\ 0\right) \approx (\,1.5811,\ 0)$, και ότι ο χώρος ανάμεσα από τα δύο «βουναλάκια» είναι σχεδόν επίπεδος που σημαίνει ότι θα μπορούσε να οδηγήσει σε προβλήματα κατά την αναζήτηση του ελαχίστου.

Ερώτημα 2ο: Μέθοδος Μέγιστης Καθόδου

Σε αυτήν την μέθοδο υπολογίζουμε το επόμενο σημείο ακολουθώντας κατεύθυνση αντίθετη από αυτή του gradient της συνάρτησης μας. Το κύριο πλεονέκτημα της είναι ότι δεν χρειάζεται να υπολογίσουμε τον εσσιανό πίνακα, αλλά μόνο το gradient της αντικειμενικής συνάρτησης. Άρα απαιτούνται λιγότεροι υπολογιστικοί πόροι, αλλά μάλλον και περισσότερες επαναλήψεις.

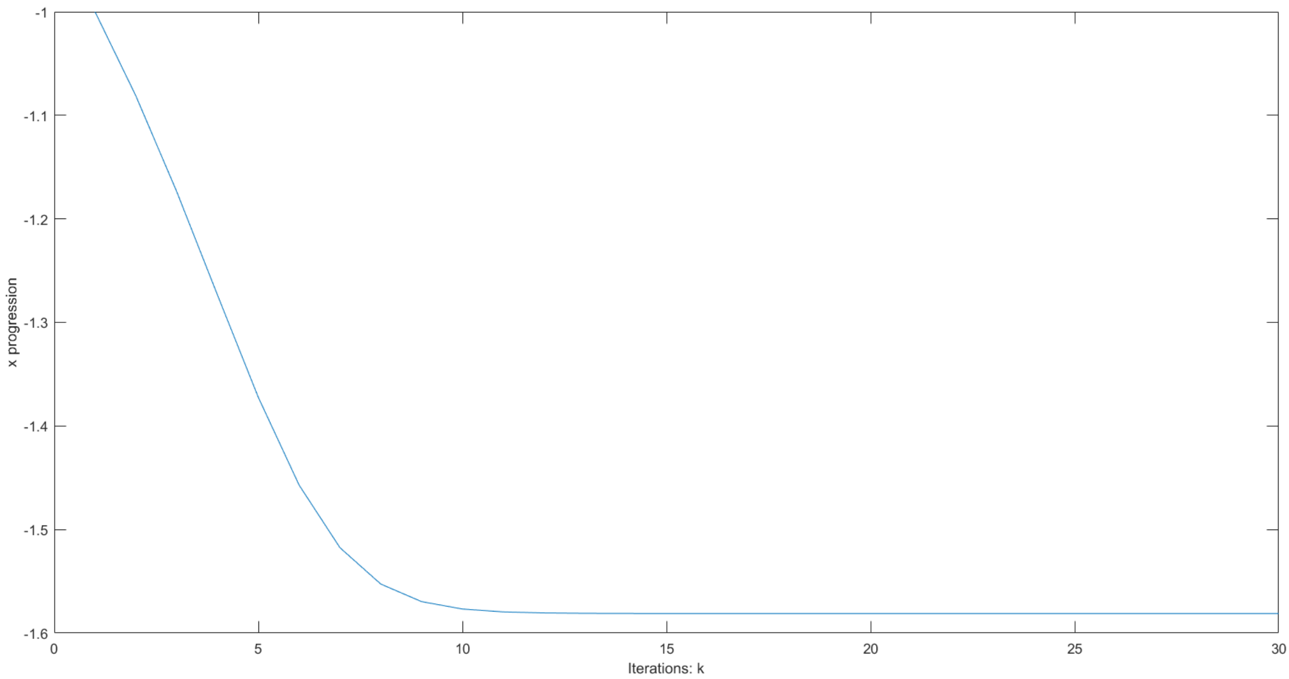
α) Μέθοδος μέγιστης καθόδου με σταθερό γ

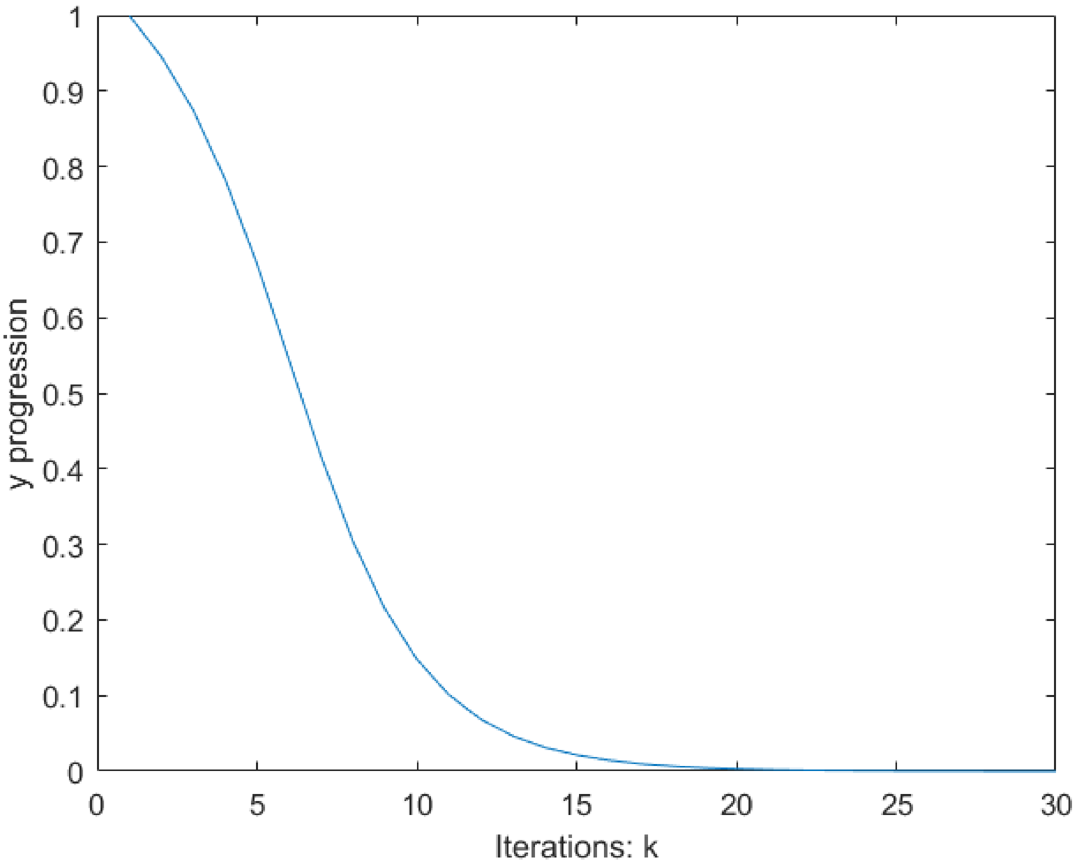
Αρχικά θα κρατάμε το γάμμα σταθερό σε όλες τις επαναλήψεις:

I. Με αρχικό σημείο το (0,0):

Προσπαθώντας να υπολογίσουμε με τον αλγόριθμο για (0,0), παρατηρούμε ότι το gradient σε εκείνο το σημείο είναι 0, οπότε το βήμα μας πολλαπλασιάζεται με 0 και συνεπώς δεν μετακινούμαστε καθόλου. Άρα το αποτέλεσμα είναι (0,0) με τιμή 0.

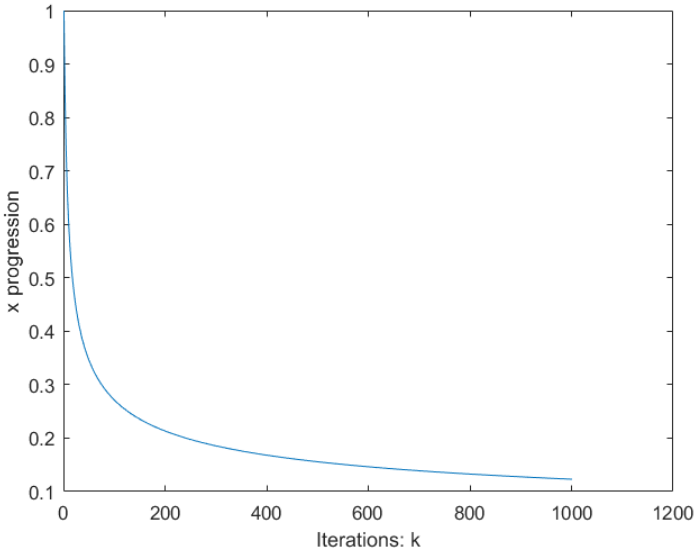
II. Με αρχικό σημείο το (-1,1):

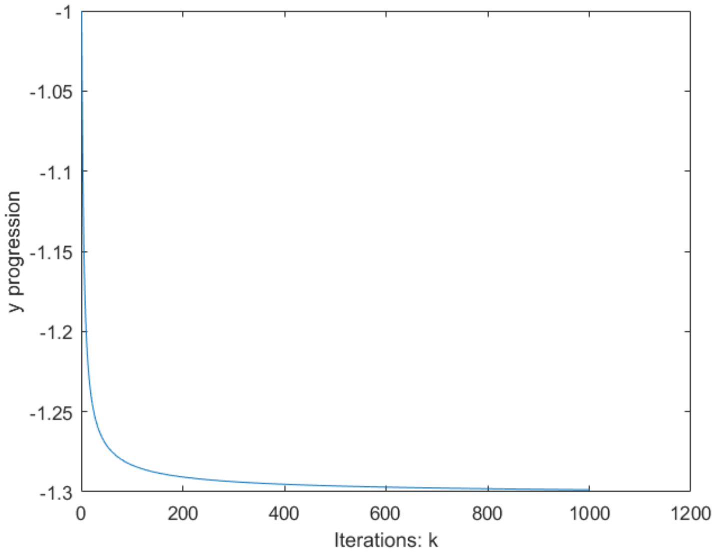




Στο παραπάνω διάγραμμα φαίνεται πως εξελίσσεται η τιμή των x,y συναρτήσει του αριθμού των επαναλήψεων k. Όπως φαίνεται φτάνουμε στις σωστές τιμές (-1.5811,0) με την τιμή της αντικειμενικής να είναι -0.8112

III. Με αρχικό σημείο το (-1,1):





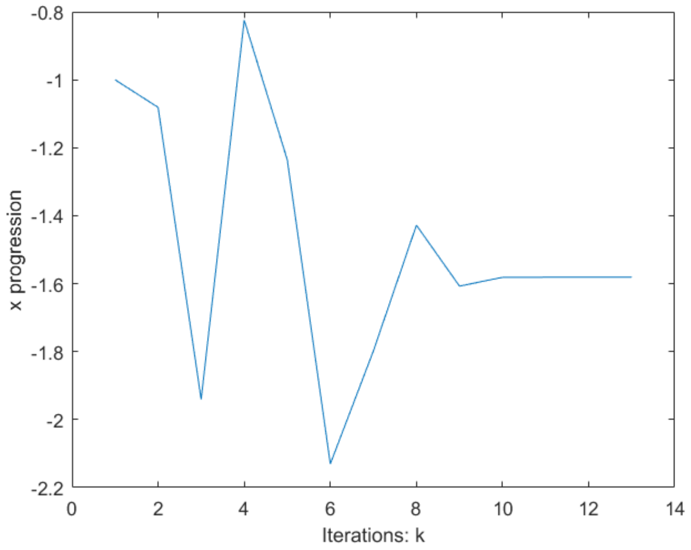
Όπως φαίνεται στα παραπάνω διαγράμματα οι τιμές που παίρνουμε για x,y είναι (0.1, -1.3), αυτή δεν είναι η ελάχιστη τιμή, αλλά ο λόγος που καταλήγουμε εκεί είναι καθώς το σημείο αυτό ανήκει στην πεδιάδα ανάμεσα από το μέγιστο και το ελάχιστο, σε αυτό το σημείο το gradient είναι πολύ κοντά στο 0 οπότε η μετατόπιση είναι πάρα πολύ μικρή, άρα δεν μπορούμε να ξεφύγουμε από αυτήν την πεδιάδα. Αυτό φαίνεται και από το γεγονός ότι φτάνουμε στις 1000 επαναλήψεις στις οποίο ο αλγόριθμος τερματίζει για να μην τρέχει ατέρμονα.

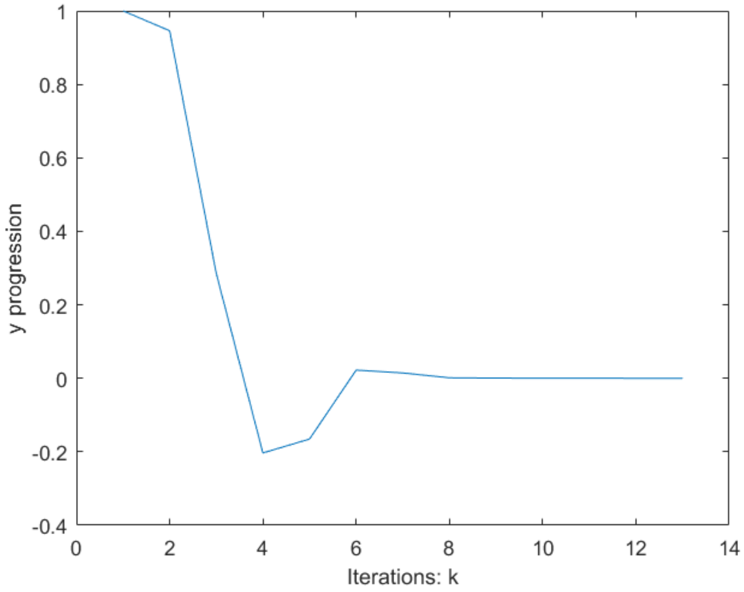
β) Μέθοδος Μέγιστης καθόδου με επιλογή του γ που ελαχιστοποιεί την $f(x_\kappa+\gamma_\kappa\delta_\kappa)$

I. Με αρχικό σημείο το (0,0):

Προσπαθώντας να υπολογίσουμε με τον αλγόριθμο για (0,0), παρατηρούμε ότι το gradient σε εκείνο το σημείο είναι 0, οπότε το βήμα μας πολλαπλασιάζεται με 0 και συνεπώς δεν μετακινούμαστε καθόλου. Άρα το αποτέλεσμα είναι (0,0) με τιμή 0.

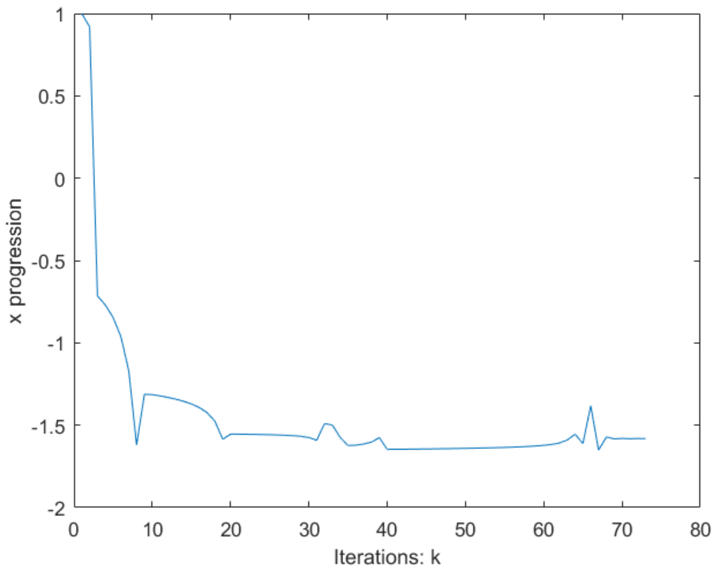
II. Με αρχικό σημείο το (-1,1):

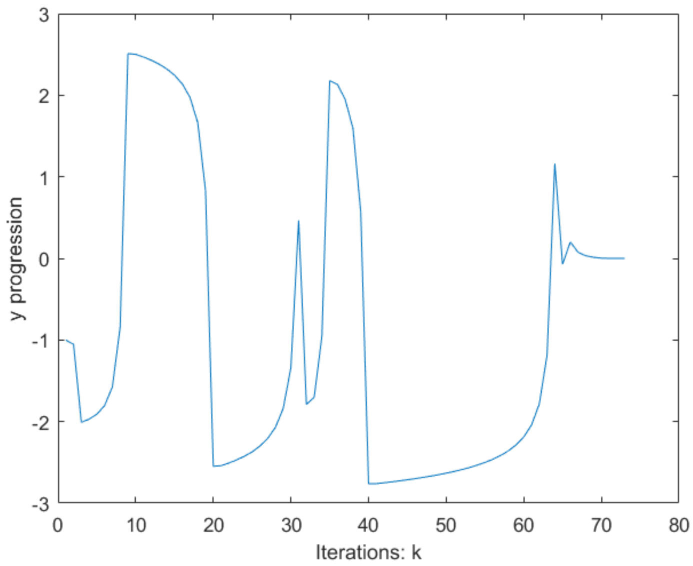




Όπως και πριν βγάζουμε σωστό αποτέλεσμα (-1.5811, 0) με την τιμή της αντικειμενικής να είναι -0.8112, αλλά σε σχέση με πριν έχουμε φτάσει σε λιγότερα βήματα με λιγότερο όμως ομαλό τρόπο.

III. Με αρχικό σημείο (1,-1)





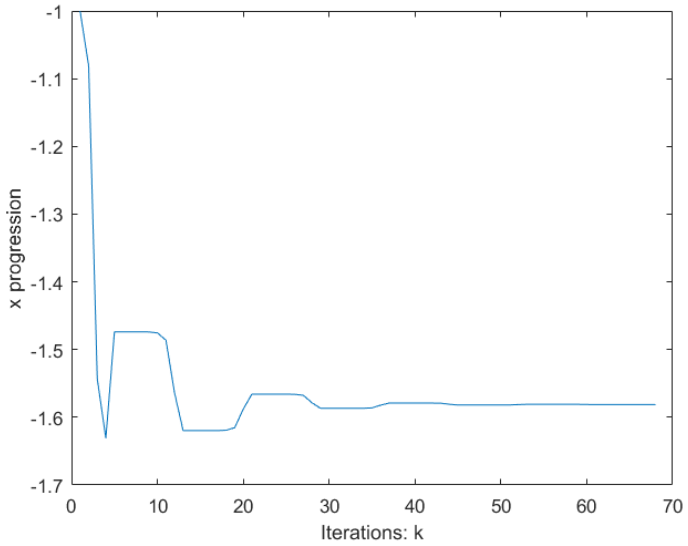
Όπως φαίνεται από τα διαγράμματα με αυτήν την μέθοδο επιλογής γ, καταφέρνουμε να ξεφύγουμε από την ενδιάμεση πεδιάδα και να φτάσουμε στο ολικό μέγιστο αφού το γάμμα μπορούσε να γίνει αρκετά μεγαλύτερο απ' ότι πριν. Βέβαια για να το πετύχουμε αυτό έπρεπε να χρησιμοποιήσουμε και μια εσωτερική διαδικασία εύρεσης ελαχίστου για μονοδιάστατη συνάρτηση η οποία είναι κάπως χρονοβόρα.

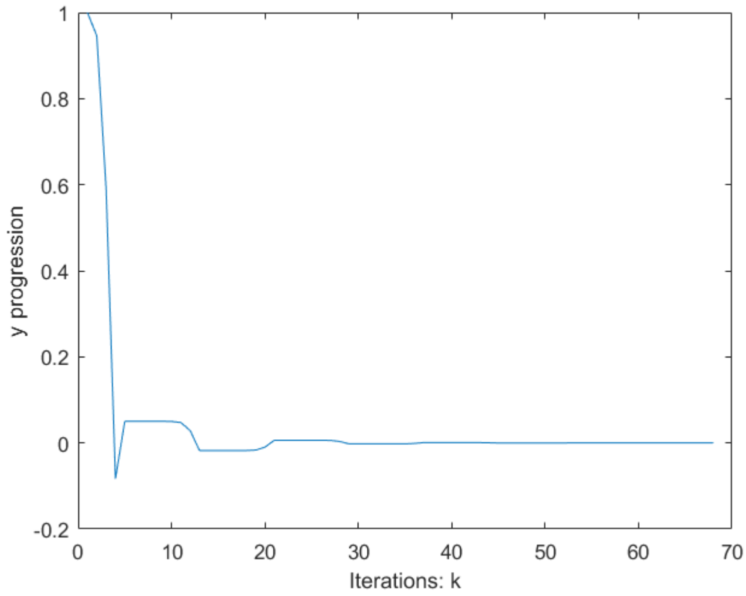
γ) Μέθοδος Μέγιστης καθόδου με επιλογή του γ από τον κανόνα Armijo

I. Με αρχικό σημείο (0,0):

Προσπαθώντας να υπολογίσουμε με τον αλγόριθμο για (0,0), παρατηρούμε ότι το gradient σε εκείνο το σημείο είναι 0, οπότε το βήμα μας πολλαπλασιάζεται με 0 και συνεπώς δεν μετακινούμαστε καθόλου. Άρα το αποτέλεσμα είναι (0,0) με τιμή 0.

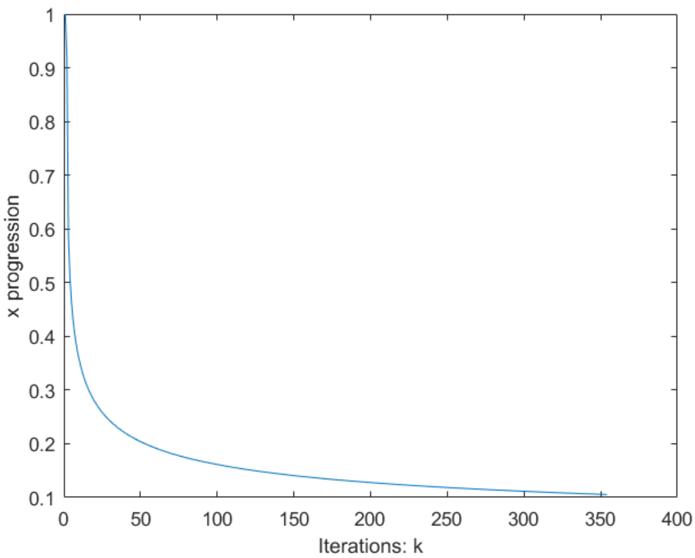
II. Με αρχικό σημείο (-1,1):

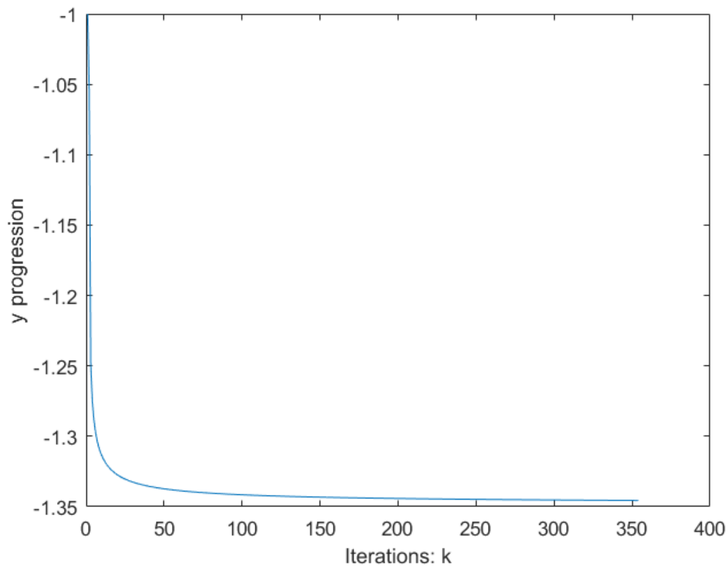




Όπως και πριν βγάζουμε σωστό αποτέλεσμα (-1.5811, 0) με την τιμή της αντικειμενικής να είναι -0.8112. Φαίνεται πώς ξεκινάμε με μεγάλο βήμα και όσο περισσότερες επαναλήψεις κάνουμε τόσο περισσότερο μειώνεται. Αυτό είναι και το πλεονέκτημα του κανόνα Armijo ότι στην αρχή μπορούμε να κάνουμε μεγάλα βήματα για να ξεφύγουμε από πιθανά τοπικά ελάχιστα και όταν φτάσουμε στο ολικό να το προσεγγίσουμε με μεγαλύτερη ακρίβεια.

III. Με αρχικό σημείο (1, -1):



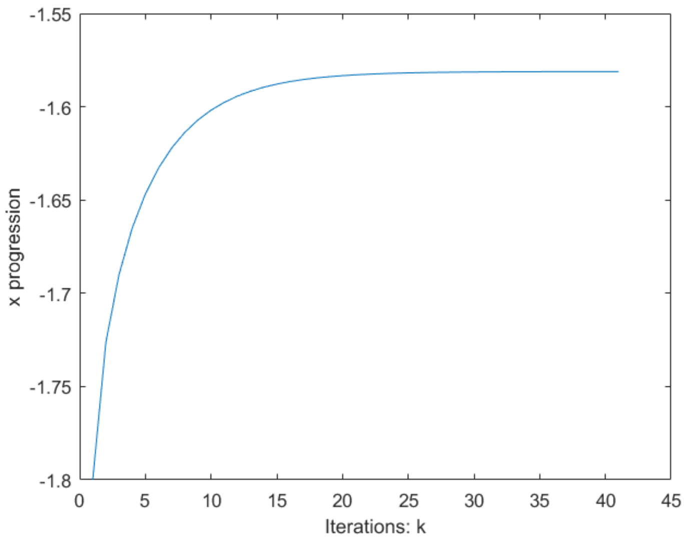


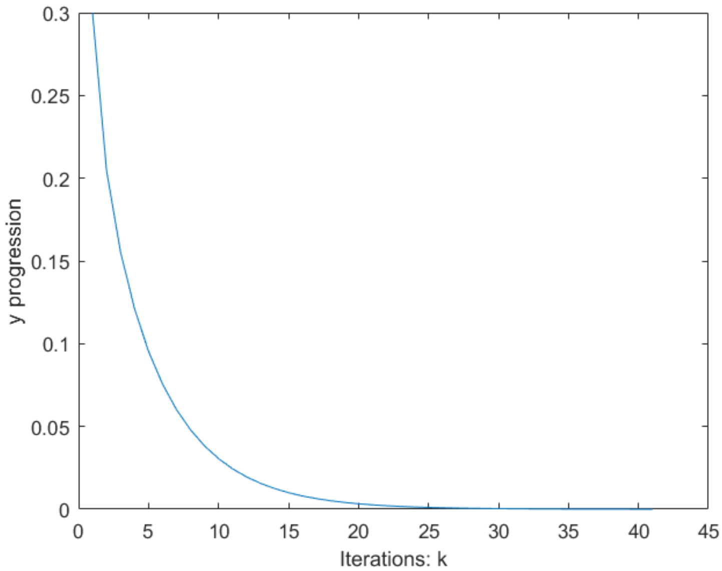
Όπως φαίνεται στα παραπάνω διαγράμματα οι τιμές που παίρνουμε για x,y είναι (0.1, -1.3), αυτή δεν είναι η ελάχιστη τιμή, αλλά ο λόγος που καταλήγουμε εκεί είναι καθώς το σημείο αυτό ανήκει στην πεδιάδα ανάμεσα από το μέγιστο και το ελάχιστο, σε αυτό το σημείο το gradient είναι πολύ κοντά στο 0 οπότε η μετατόπιση είναι πάρα πολύ μικρή, άρα δεν μπορούμε να ξεφύγουμε από αυτήν την πεδιάδα.

Ερώτημα 3ο: Μέθοδος Newton

Στην μέθοδο Newton υπολογίζουμε την κατεύθυνση του επόμενου βήματος με βάση τον εσσιανό πίνακα της αντικειμενικής συνάρτησης. Αυτό έχει ορισμένα πλεονεκτήματα και ορισμένα μειονεκτήματα. Από την μία γνωρίζοντας τον εσσιανό μπορείς να δεις ένα βήμα μπροστά και να ξέρεις με μεγαλύτερη σιγουριά προς τα πού να κινηθείς αλλά και να φτάσεις εκεί πιο γρήγορα. Από την άλλη όχι μόνο είναι υπολογιστικά πιο χρονοβόρο να υπολογίσεις τον εσσιανό πίνακα, αλλά ταυτόχρονα πρέπει να είναι και θετικά ορισμένος για να μπορείς να το χρησιμοποιήσεις για ώστε να λειτουργεί σωστά ο αλγόριθμος.

Και για τα 3 αρχικά σημεία ο εσσιανός πίνακας είναι μη θετικά ορισμένος, άρα το αποτέλεσμα που θα πάρουμε και για τις 9 περιπτώσεις είναι το αρχικό σημείο. Αν παίρναμε ένα σημείο που να έχει θετικά ορισμένο εσσιανό, πχ (-1.8, 0.3) θα βγάζαμε σωστό αποτέλεσμα:





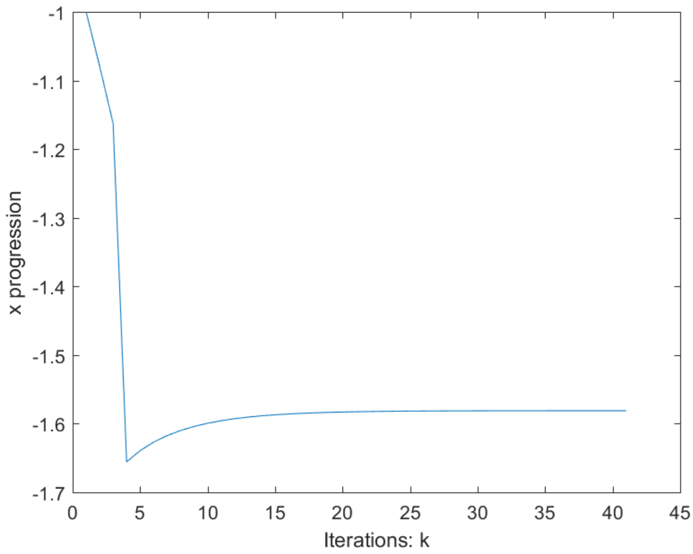
Ερώτημα 4ο: Μέθοδος Levenberg-Marquardt

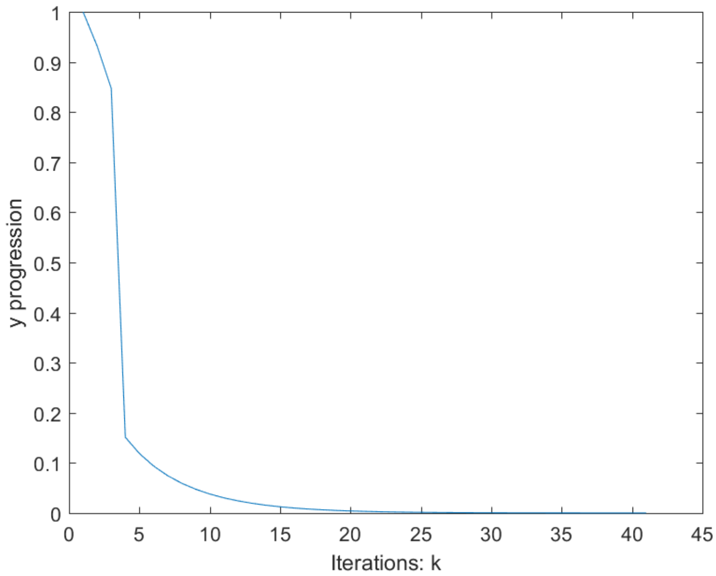
Η μέθοδος Levenberg-Marquardt λειτουργεί σαν την μέθοδο Newton όταν ο εσσιανός είναι θετικά ορισμένος, ενώ όταν δεν είναι προσθέτουμε σε αυτόν τον μικρότερο ακέραιο πολλαπλάσιο του μοναδιαίου πίνακα μέχρι να γίνει θετικά ορισμένος και χρησιμοποιώντας αυτόν για τους υπολογισμούς μας. Με αυτόν τον τρόπο αξιοποιούμε τα πλεονεκτήματα της μεθόδου newton αλλά χωρίς να χρειάζεται να έχουμε θετικά ορισμένο εσσιανό. Όσο λιγότερες φορές προσθέτουμε τον μοναδιαίο πίνακα, τόσο πιο κοντά είναι αυτή η μέθοδος στην μέθοδο newton ενώ όσο πιο πολλές φορές προσθέτουμε τον μοναδιαίο, τόσο περισσότερο αυτή η μέθοδος τείνει να ταυτιστεί με την μέθοδο της μέγιστης καθόδου.

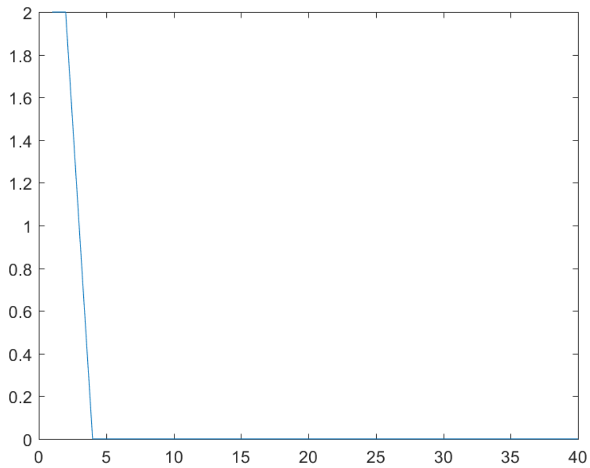
α) Μέθοδος Levenberg-Marquardt με σταθερό γ

I. Με αρχικό σημείο (0, 0):
   Για αρχικό σημείο το (0,0) το gradient του f γίνεται μηδέν και συνεπώς η λύση της εξίσωσης για την εύρεση του d_k θα βγάλει και αυτή 0, οπότε εγκλωβιζόμαστε στο (0,0) με αυτήν την μέθοδο.

II. Με αρχικό σημείο (-1, 1):

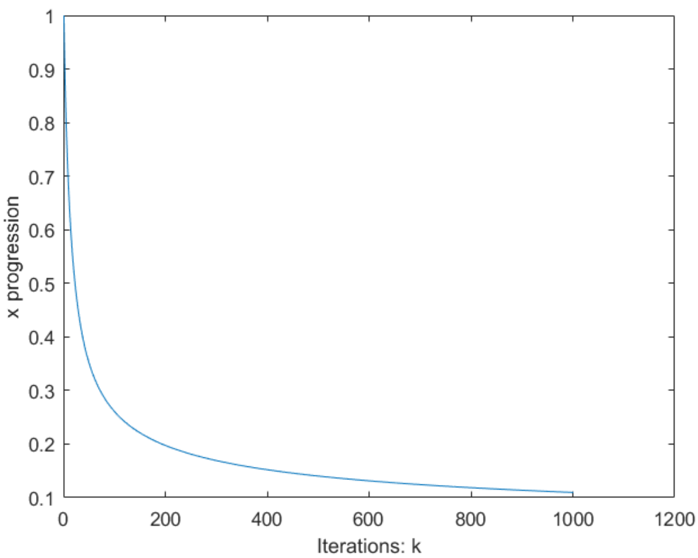


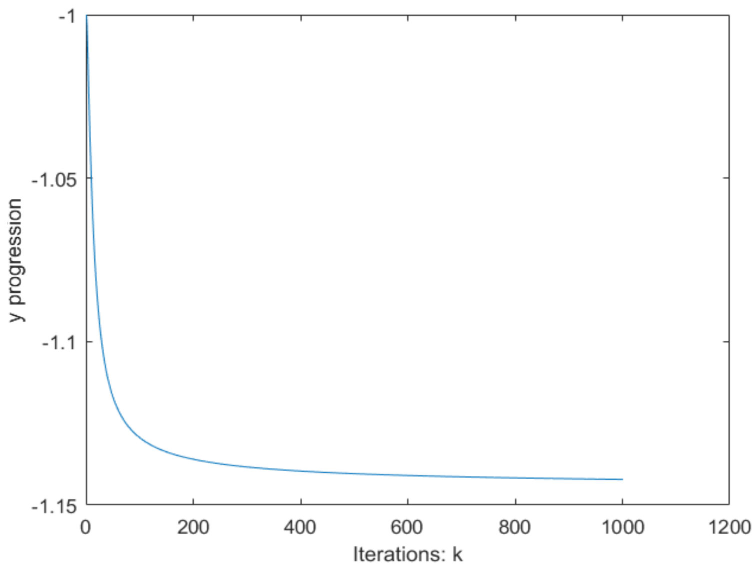




Όπως βλέπουμε και στα διαγράμματα ο αλγόριθμος σωστά καταλήγει στο (-1.5811, 0) με τιμή -0.8112 μετά από αρκετές επαναλήψεις. Στο τρίτο διάγραμμα βλέπουμε την τιμή του mk συναρτήσει των επαναλήψεων. Από αυτό γίνεται εμφανές ότι μόνο στις πρώτες επαναλήψεις ήταν απαραίτητο να προστεθεί να μία ή δυο φορές ο μοναδιαίος πίνακας για να ‘ξεκολλήσει’ στην αρχή ο αλγόριθμος και στην συνέχεια ό εσσιανός ήταν θετικός.

III. Με αρχικό σημείο (1, -1):



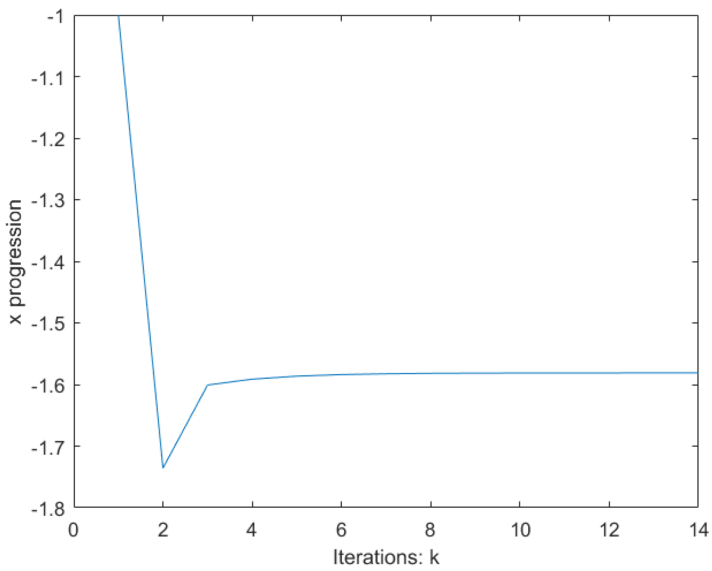


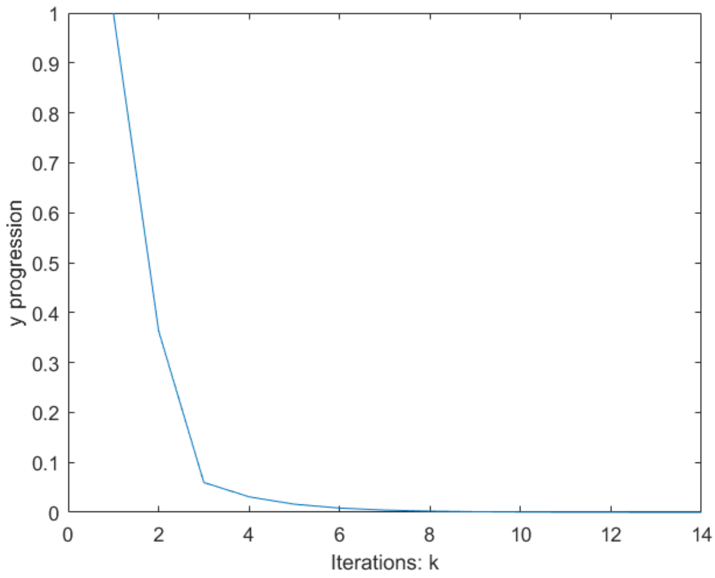
Η λύση για ακόμη μια φορά καταλήγει σε σημείο με τιμή 0 και μηδενική κλίση από το οποίο δεν μπορεί να ξεφύγει.

β) Μέθοδος Levenberg-Marquardt με γ που ελαχιστοποιεί την $f(x_κ+γ_κδ_κ)$

I. Με αρχικό σημείο (0, 0):
Για αρχικό σημείο το (0,0) το gradient του f γίνεται μηδέν και συνεπώς η λύση της εξίσωσης για την εύρεση του d_k θα βγάλει και αυτή 0, οπότε εγκλωβιζόμαστε στο (0,0) με αυτήν την μέθοδο.

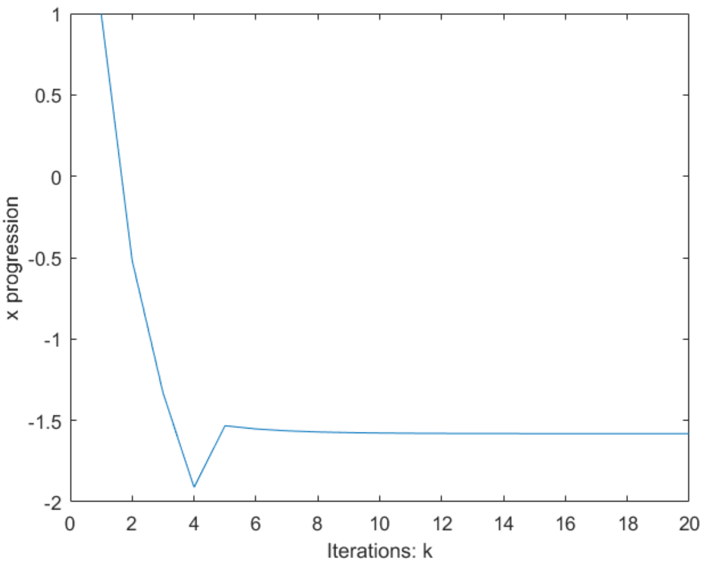
II. Με αρχικό σημείο (-1, 1):

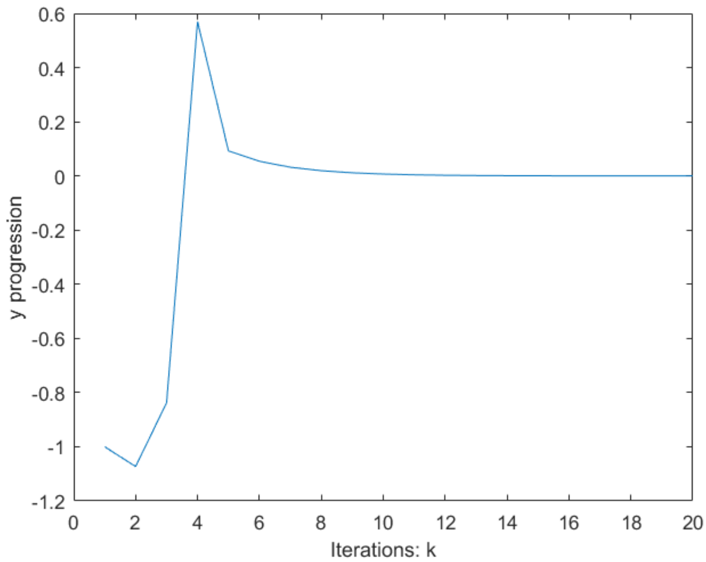




Καταλήγουμε σε σωστό αποτέλεσμα σε πολύ λιγότερες επαναλήψεις σε σχέση με το σταθερό γ.

III. Με αρχικό σημείο (1, -1):



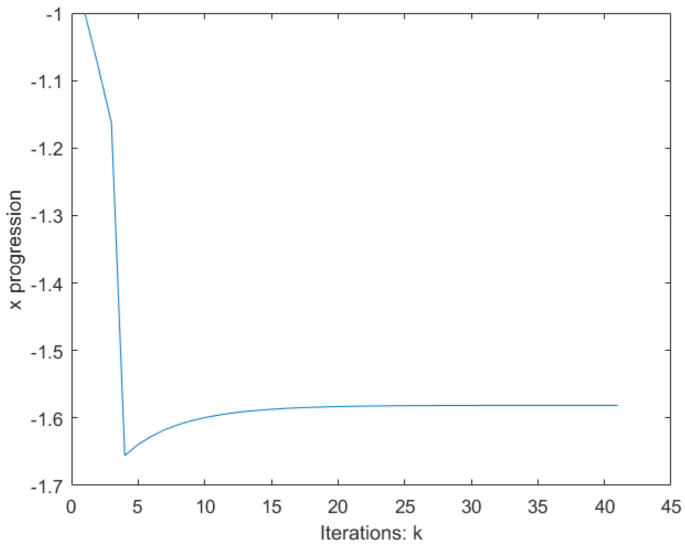


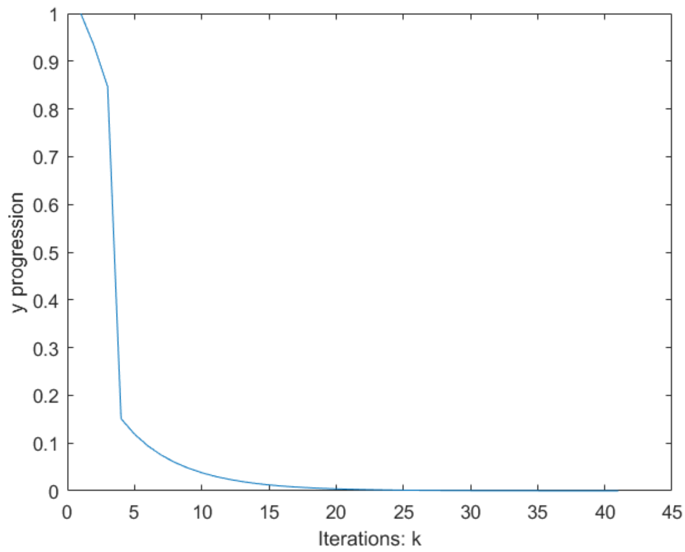
Με αυτήν την επιλογή γάμμα καταφέρνουμε να φτάσουμε στο σωστό ολικό ελάχιστο ανεξάρτητα από την επιλογή αρχικού σημείου, αρκεί να μην έχει μηδενική κλίση, αφού το γάμμα μπορεί να επιλεγεί αρκετά μεγάλο και να ξεφύγει από πιθανά τοπικά ελάχιστα.

γ) Μέθοδος Levenberg-Marquardt με γ που επιλέγεται με βάση του κανόνα Armijo

I. Με αρχικό σημείο (0, 0):
Για αρχικό σημείο το (0,0) το gradient του f γίνεται μηδέν και συνεπώς η λύση της εξίσωσης για την εύρεση του d_k θα βγάλει και αυτή 0, οπότε εγκλωβιζόμαστε στο (0,0) με αυτήν την μέθοδο.

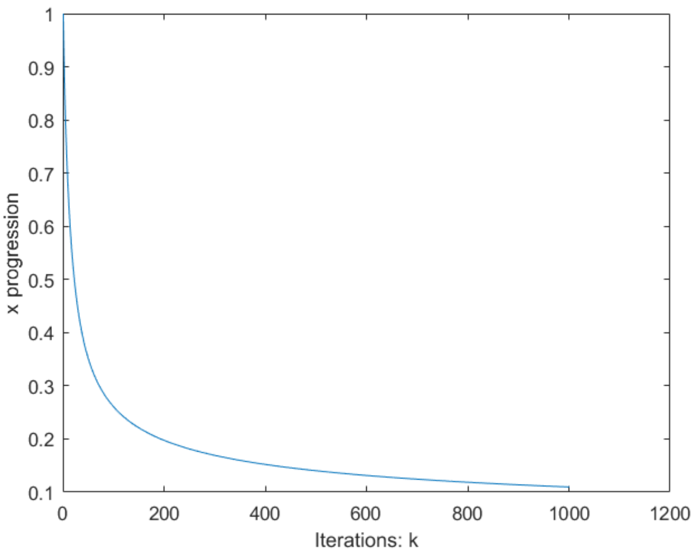
II. Με αρχικό σημείο (-1, 1):

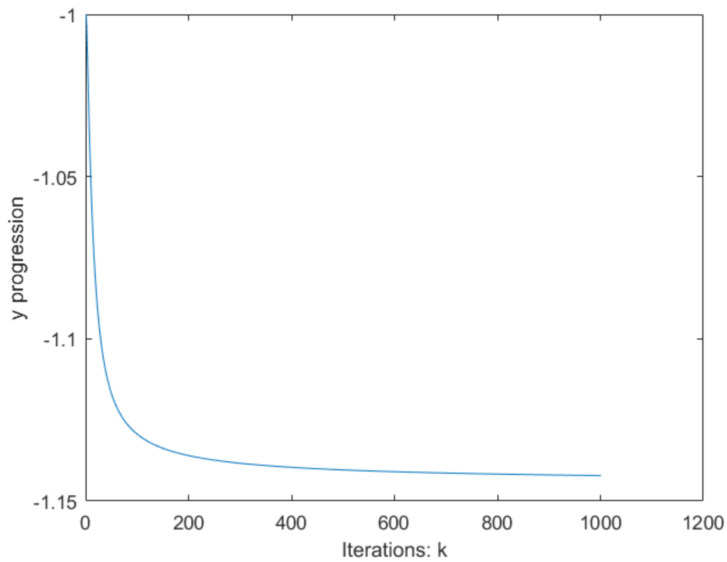




Καταλήγουμε σωστά στο ολικό ελάχιστο.

III. Με αρχικό σημείο (1, -1):





Συμπεράσματα

Η μέθοδος της μέγιστης καθόδου θα μπορούσε να παρομοιαστεί με άπληστο αλγόριθμο αφού επιλέγει το βήμα της με βάση μόνο την κλίση στο συγκεκριμένο σημείο, αλλά είναι και πιο εύκολος και γρήγορος ο υπολογισμός της. Αντίθετα η μέθοδος Newton κοιτάει ένα βήμα μπροστά αφού βασίζει την απόφαση της στην δεύτερη παράγωγο, αλλά είναι αρκετά πιο δύσκολο να εφαρμοστεί αφού ο εσσιανός της αντικειμενικής συνάρτησης πρέπει να είναι θετικά ορισμένος. Για αυτόν τον λόγο χρησιμοποιείται η μέθοδος Levenberg-Marquadt η οποία είναι η μέση λύση. Όταν ο εσσιανός είναι θετικό ορισμένος λειτουργεί σαν την μέθοδο Newton ενώ αν είναι αρνητικά ορισμένος προσθέτοντας πολλαπλάσια του μοναδιαίου τείνει να λειτουργεί σαν την μέθοδο μέγιστης καθόδου. Όσον αφορά την επιλογή του γ, η επιλογή του σαν σταθερό, αν και απλή, για την σωστή εφαρμογή της απαιτείται εμπειρία, πολύ μικρή τιμή θα οδηγήσει σε μεγάλη καθυστέρηση ή λάθος αποτέλεσμα ενώ πολύ μεγάλη σε λάθος αποτέλεσμα. Η επιλογή του γ που ελαχιστοποιεί την επόμενη τιμή που θα πάρει η αντικειμενική συνάρτηση μπορεί να είναι πολύ αποτελεσματική αφού το γάμμα μπορεί να πάρει αρκετά μεγάλες τιμές για να ξεφύγει από τοπικά ελάχιστα, αλλά μπορεί να χρειαστούν παραπάνω βήματα για τον υπολογισμό αυτό, ενώ είναι και απαραίτητη η χρήση μιας μεθόδου για την εύρεση της τιμής του γ που ελαχιστοποιεί την f (κάτι που απαιτεί επιπλέον χρόνο και υπολογιστικούς πόρους). Η μέθοδος Armijo είναι μια μέθοδος στην οποία το βήμα φθίνει συνεχώς με το σκεπτικό ότι στην αρχή μπορούμε να κάνουμε μεγάλα βήματα για να φτάσουμε γρήγορα κοντά στο ελάχιστο και μετά να το υπολογίσουμε με ακρίβεια.