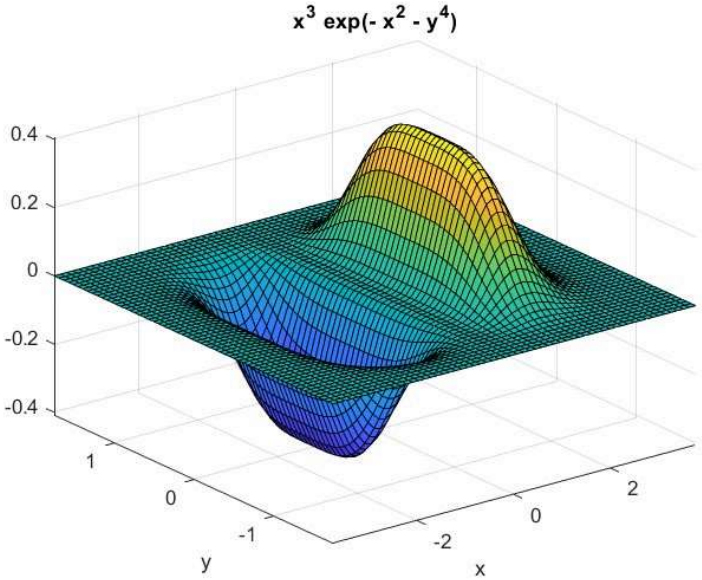
2η Εργαστηριακή Άσκηση:

Ελαχιστοποίηση κυρτής συνάρτησης πολλών μεταβλητών με τη χρήση παραγώγων

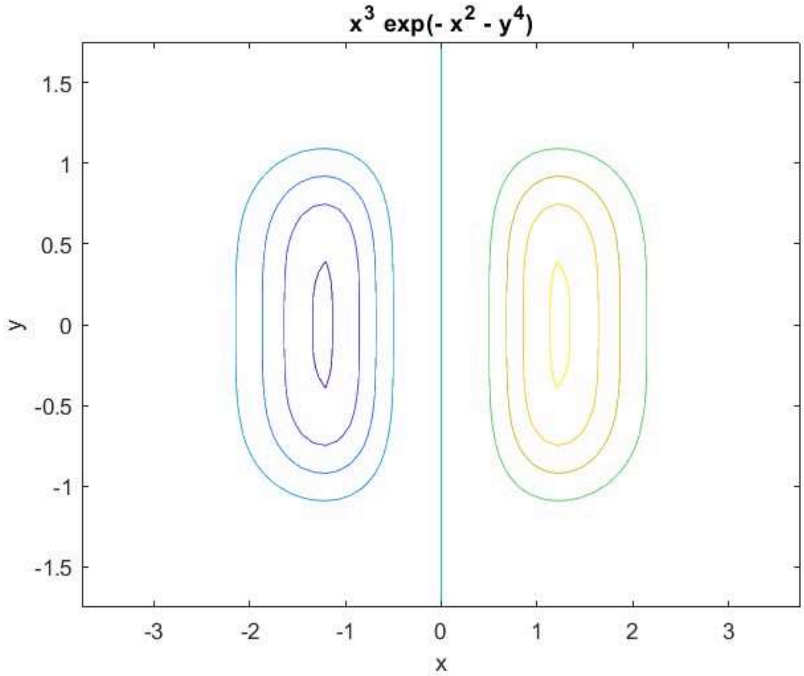
Σκοπός μας ήταν να ελαχιστοποιήσουμε την συνάρτηση

$$f(x,\, y) \;=\; x^3 e^{-x^2 - y^4}$$



Από το γράφημα μας βλέπουμε πως η συνάρτηση είναι κοίλη σε ένα σημείο και κυρτή σε ένα άλλο, καθώς και ότι έχει πολλά τοπικά μέγιστα και ελάχιστα. Αυτό μας προμηνύει ήδη ότι υπάρχει κίνδυνος εγκλωβισμού των αλγορίθμων σε κάποιο τοπικό ελάχιστο που δεν είναι το επιθυμητό.

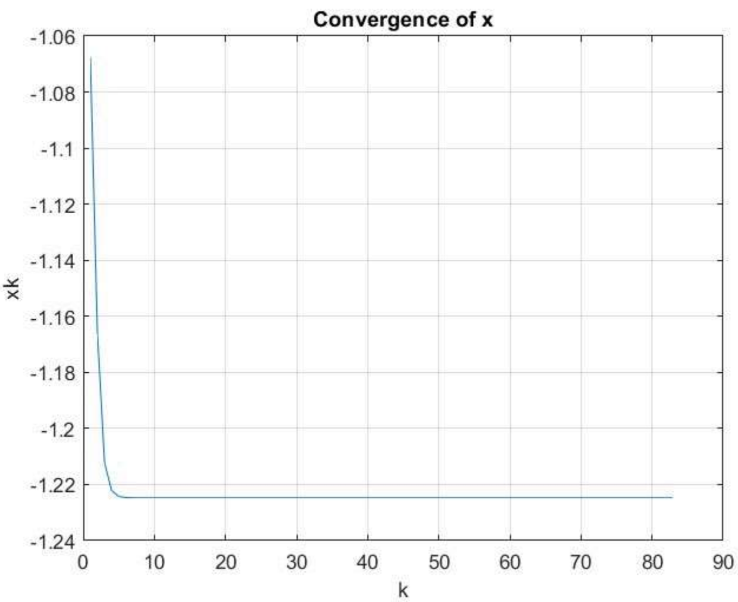
Δίπλα βλέπουμε και τις ισοϋψείς καμπύλες ώστε να έχουμε καλύτερη ιδέα του γύρω από ποιές τιμές θα βρούμε τελικά το τοπικό ελάχιστο.

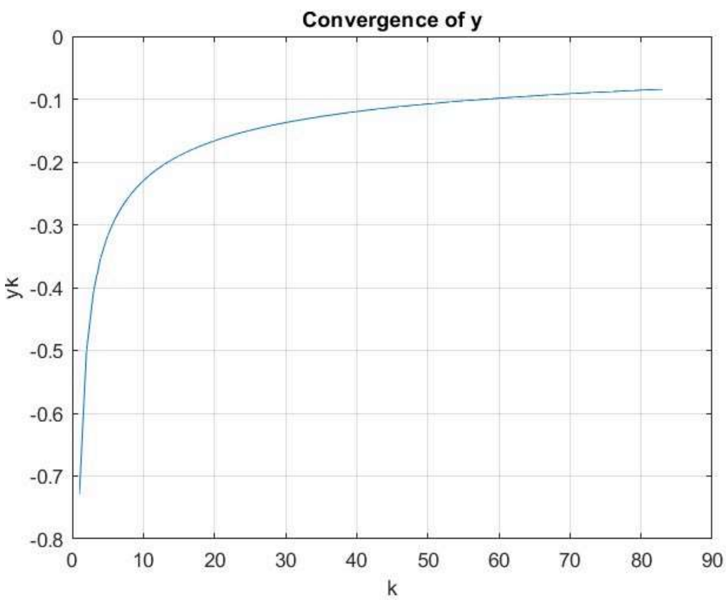


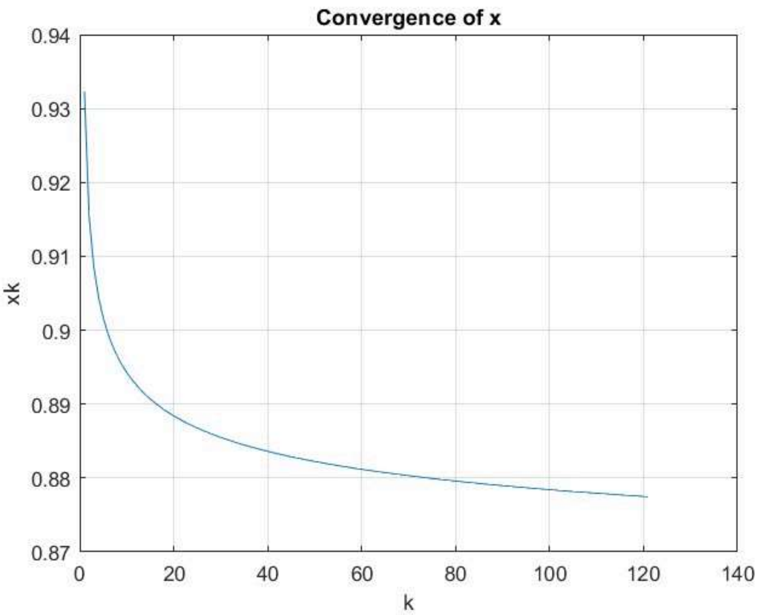
Από θεωρητική προσέγγιση, βρίσκουμε ότι τα κρίσιμα σημεία της f είναι τα $[\sqrt{(3/2)},\, 0]$ $[-\, \sqrt{(3/2)},\, 0]$ $[0,\, y \in R]$. Από αυτά με αντικατάσταση στον Εσσιανό πίνακα, αλλά και παρατηρώντας το γράφημα βλέπουμε πως το σημείο $[-\, \sqrt{(3/2)},\, 0]$ ή αλλιώς $[-\, 1.2247,\, 0]$ είναι το μικρότερο τοπικό ελάχιστο με τιμή -0.409916 και άρα αυτό μας ενδιαφέρει να βρίσκουμε μέσω των αλγορίθμων.

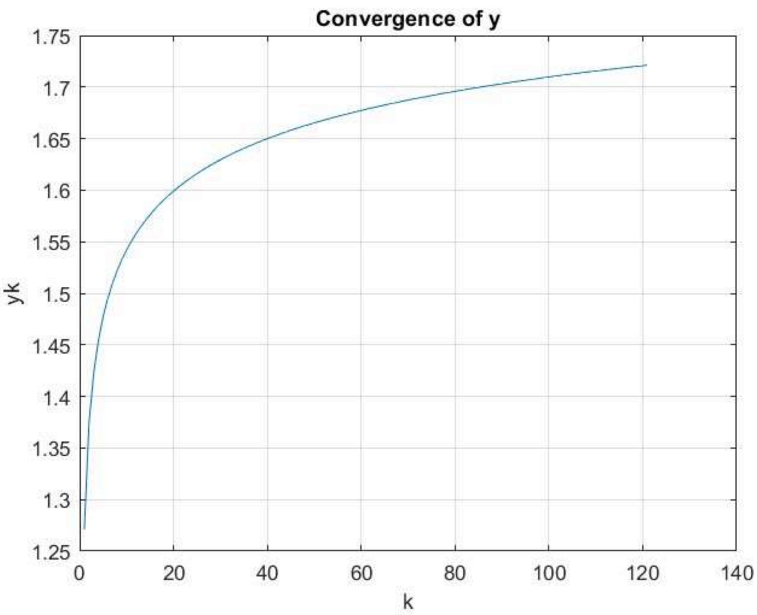
1. Μέθοδος της Μέγιστης Καθόδου

Η μέθοδος αυτή, σε αντίθεση με τις επόμενες που θα δούμε, χρησιμοποιεί μόνο το gradient της συνάρτησης για να βρει το μονοπάτι που θα ελαχιστοποιήσει την f. Θα δούμε πως αλλάζει το ελάχιστο στο οποίο σταθεροποιείται ο αλγόριθμος με βάση το σημείο εκκίνησης για $[0,\ 0]\quad[1,\ 1]\quad[-1,\ -1]$ με βήμα 0.5.



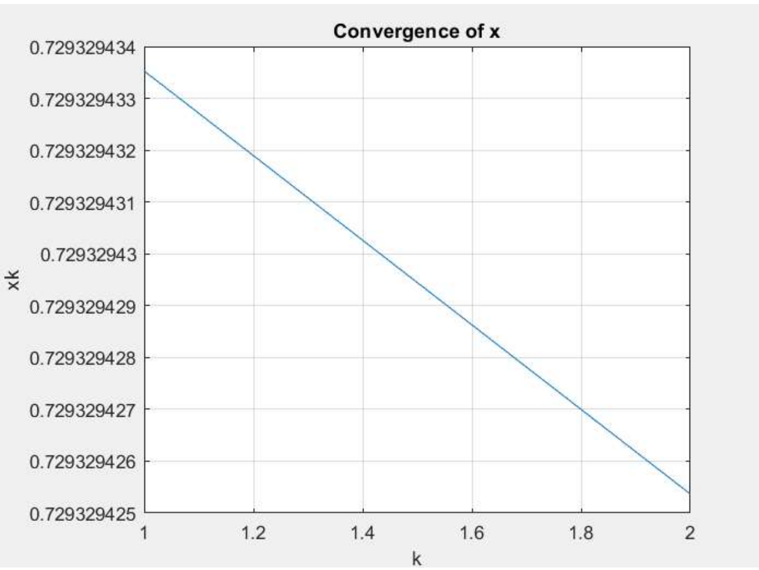


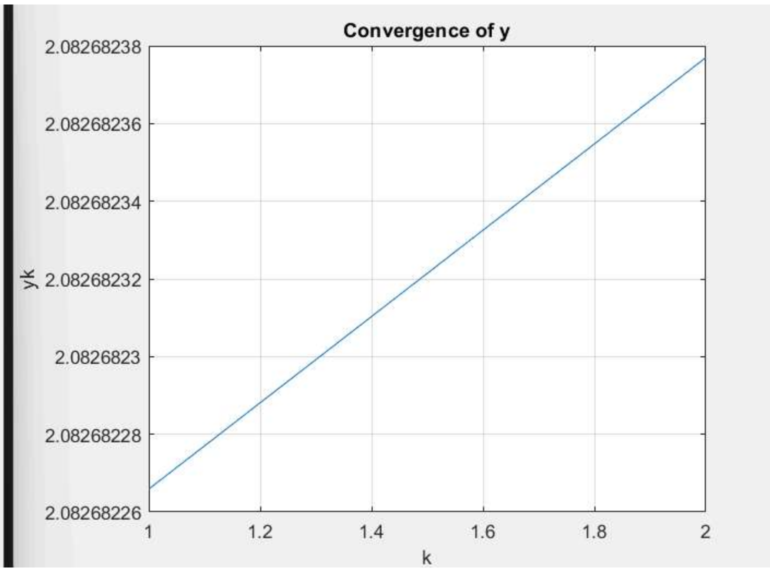


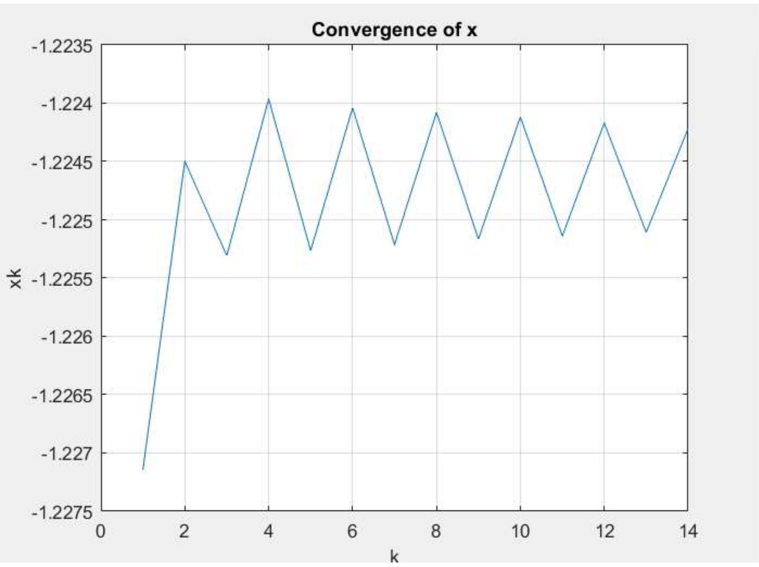


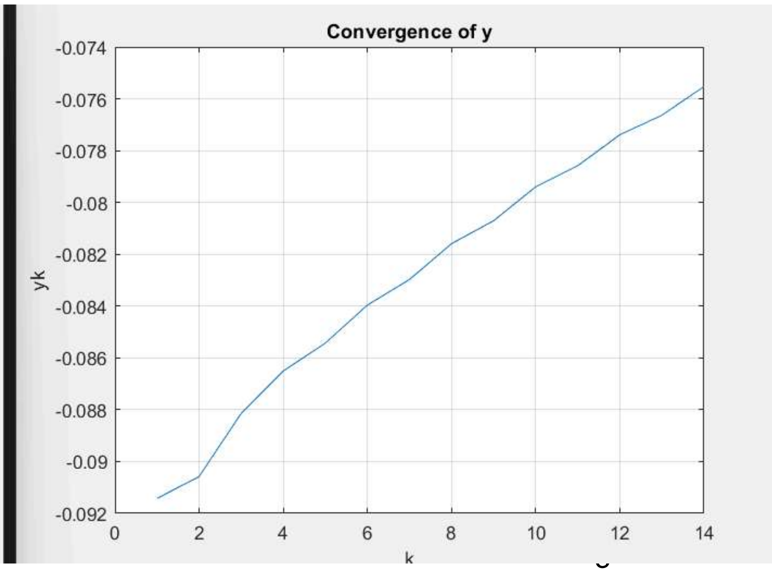
Παραπάνω βλέπουμε τα αποτελέσματα για [-1 ,-1] το οποίο έδωσε ελάχιστο -0.4099 και για [1, 1] το οποίο έδωσε ελάχιστο 4.8197e-05 (μια τιμή πολύ κοντά στο 0). Το [0, 0] δεν συμπεριλήφθηκε καθώς ο αλγόριθμος δεν έφυγε ποτέ από αυτό το σημείο δίνοντας τελικό ελάχιστο 0. Αν κοιτάξουμε ξανά στο αρχικό γράφημα της συνάρτησης, παρατηρούμε πως για αυτές τις δύο τιμές που είχαμε ως αποτέλεσμα το 0 ο αλγόριθμος μπαίνει σε ένα τοπικό ελάχιστο (το 0) από το οποίο μετά δε βγαίνει, για αυτό και τα διαφορετικά αποτελέσματα απο το αρχικό σημείο [-1, -1] το οποίο μάλιστα ήταν και πολύ καλή επιλογή αφού έδωσε το αποτέλεσμα που θέλαμε με μεγάλη ακρίβεια.

Αν για step αντί για κάποιον τυχαίο, αυθαίρετο αριθμό θεωρήσουμε ένα γ τέτοιο ώστε να ελαχιστοποιεί την f(χκ + γκdκ) παίρνουμε τα εξής αποτελέσματα:





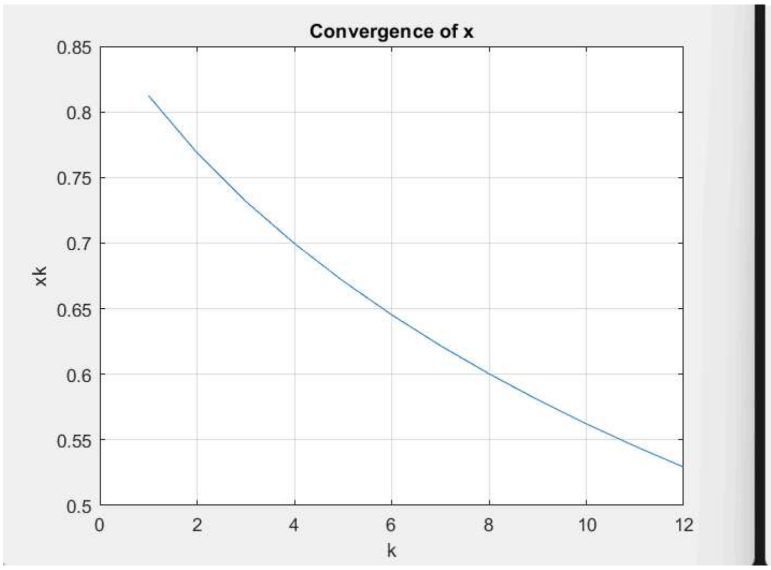


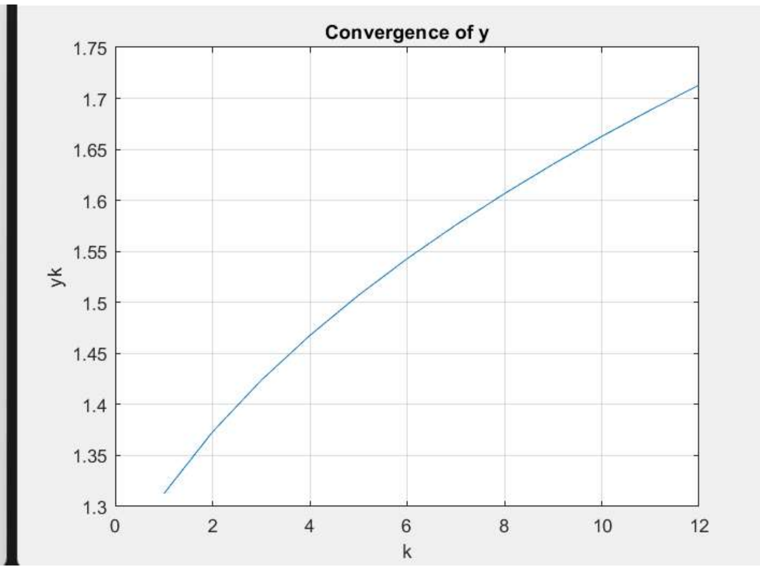


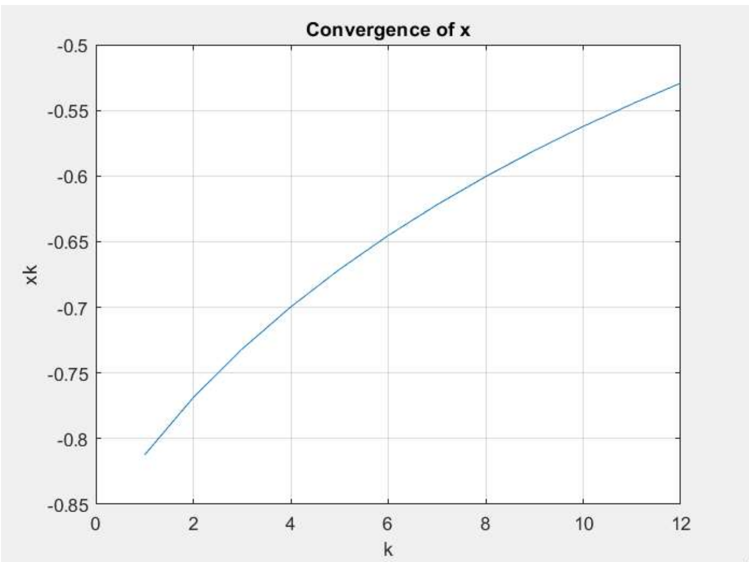
Βλέπουμε πως αν και το τελικό αποτέλεσμα δεν αλλάζει σημαντικά (και τα τρεία σημείο δίνουν το ίδιο τελικό ελάχιστο με πριν) αλλάζει σημαντικά ο τρόπος με τον οποίο συγκλίνουν στο σημείο αυτό. Το πιο προφανές είναι ότι η σύγκλιση γίνεται σε πολύ λιγότερα βήματα. Για το σημείο [1, 1] ειδικά βλέπουμε να έχουν αλλάξει τα τελικά x, y χωρίς όμως να αλλάξει αυτό τη σύγκλιση στο 0.

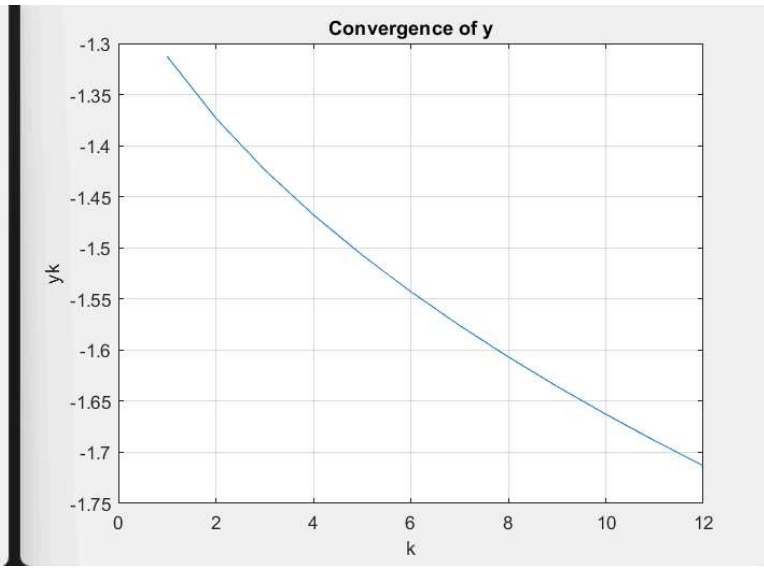
2. Μέθοδος του Newton

Αυτή η μέθοδος χρησιμοποιεί πέρα απο το gradient και τον Εσσιανό πίνακα και συγκεκριμένα τον αντίστροφο του πολλαπλασιασμένο με το gradient σε κάθε σημείο ώστε να βρει την επόμενη βέλτιστη κατεύθυνση. Με τα προηγούμενα σημεία παίρνουμε τα ελάχιστα για [-1, -1] -2.0348e-05, για [1, 1] 2.0348e-05 και για [0, 0] πάλι 0.



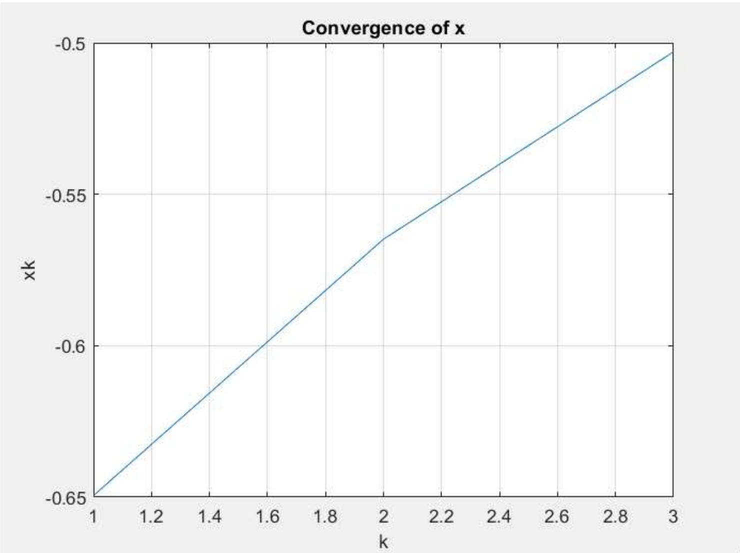


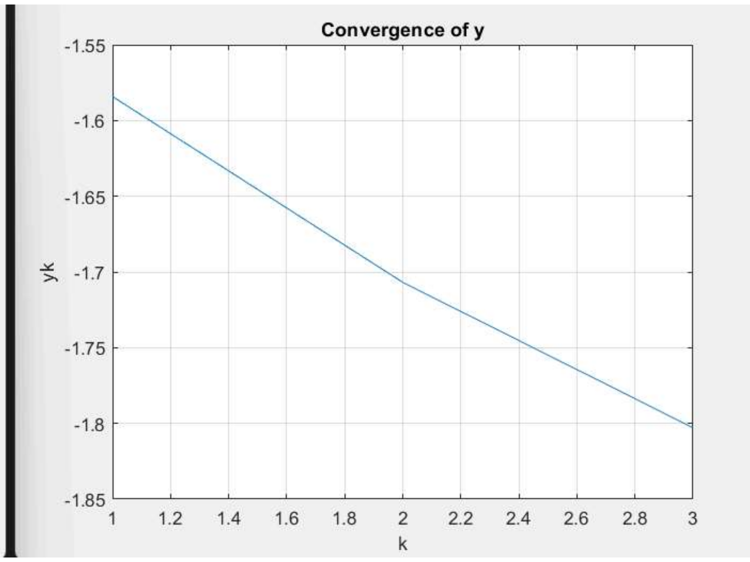


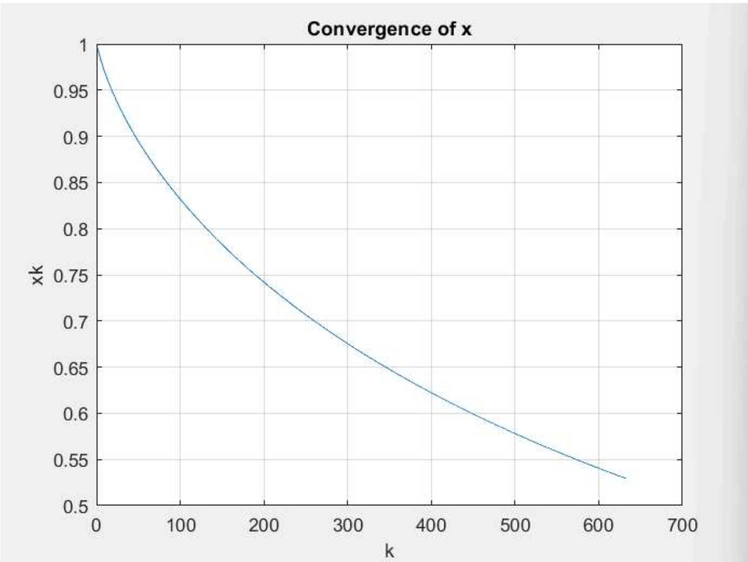


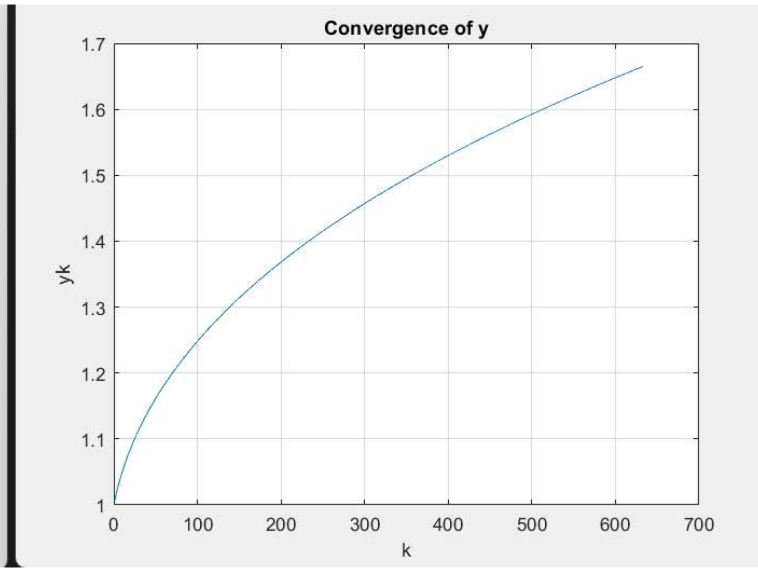
Περιέργως, παρότι θα περιμέναμε να δούμε αποτελέσματα παρόμοια με τα προηγούμενα (αν όχι ίδια) παρατηρούμε πως με τη μέθοδο Newton η συνάρτηση για όλα τα δοσμένα αρχικά σημεία συγκλίνει στο τοπικό ελάχιστο 0 και έπειτα δεν μπορεί να βγει εκτός αυτού. Αυτό μας λέει πως ίσως η μέθοδος αυτή δεν είναι η ιδανική για τη συγκεκριμένη συνάρτηση παρά την γρήγορη σύγκλιση της, όπως φαίνεται από τον μειωμένο αριθμό επαναλήψεων.

Με step που δίνεται από την ελαχιστοποίση του f(χκ + γκdκ):







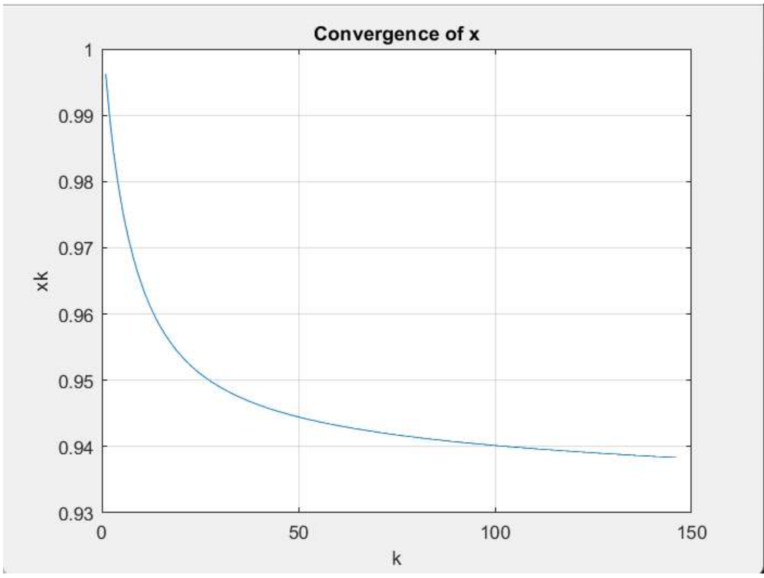


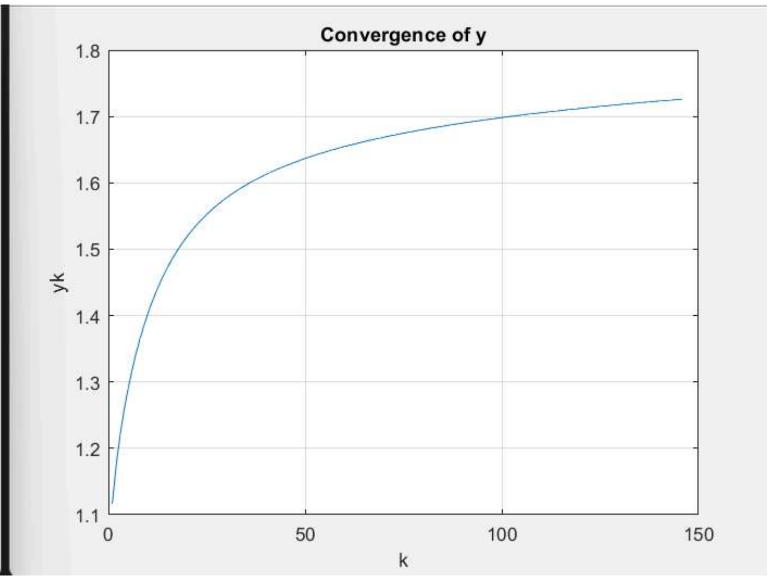
Βλέπουμε πως για το αρχικό σημείο [-1, -1] η σύγκλιση έγινε απίστευτα γρήγορα ενώ για το [1, 1] οι απαραίτητες επαναλήψεις ανέβηκαν εκθετικά. Αυτό το παράδοξο πιθανότατα να οφείλεται στο πόσο πιο κοντά είναι το

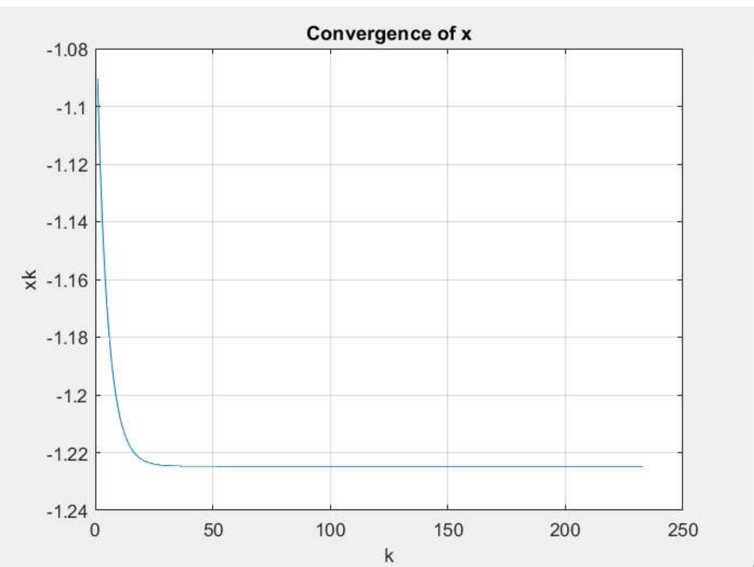
[-1, -1] στη θέση του τελικού ελαχίστου ή ίσως σε κακή υλοποίηση της υπορουτίνας που βρίσκει το κατάλληλο γ.

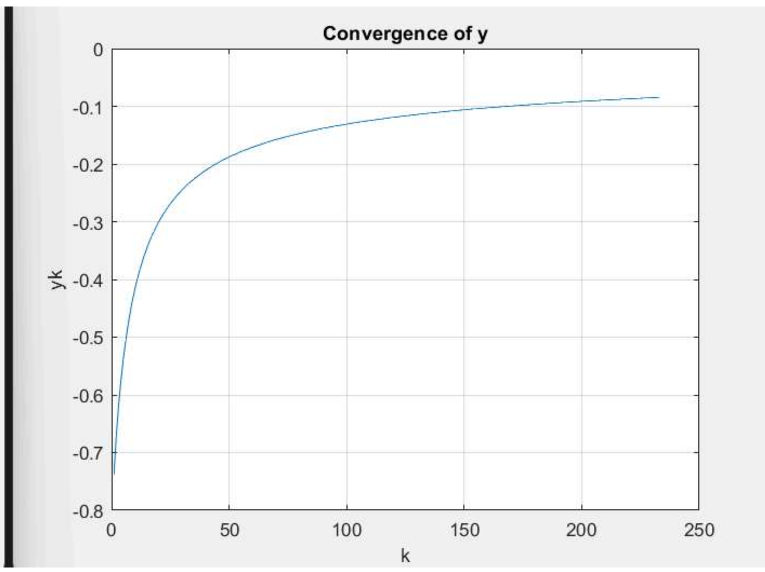
3. Μέθοδος Levenberg-Marquardt

Η τελευταία μέθοδος πάει να γεφυρώσει το κένο ανάμεσα στη μέθοδο της μέγιστης καθόδου και στη μέθοδο Newton. Οι νέες τιμές που παίρνουν τα x, y κάθε φορά εξαρτώνται από έναν παράγοντα μ ο οποίος και καθορίζει αν η μέθοδος θα δρα περισσότερο σαν την μια ή την αλλή προαναφερθείσες μεθόδους. Τα αποτελέσματα που παίρνουμε είναι: 4.7746e-05 από το [1, 1], -0.4099 από το [-1, -1] και 0 από το [0, 0]



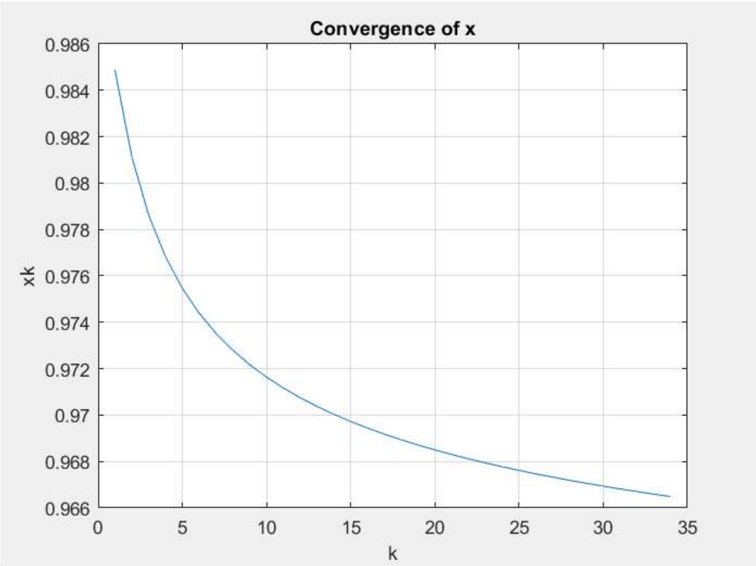


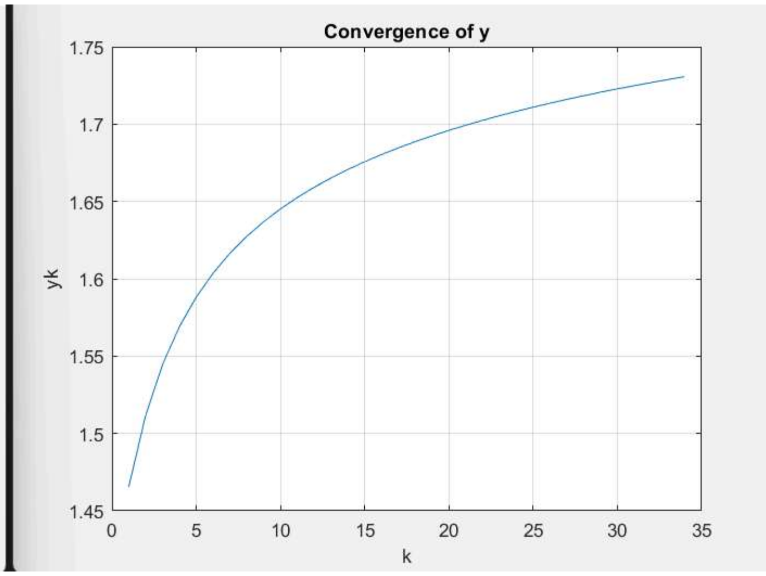


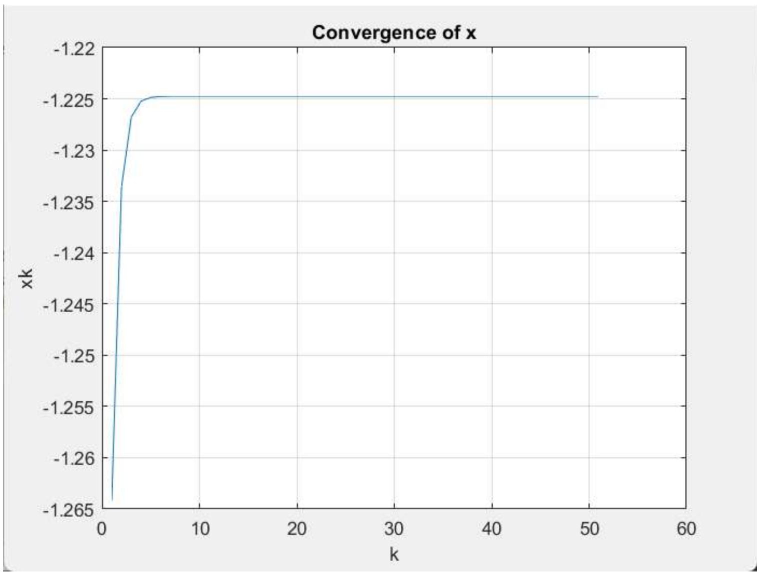


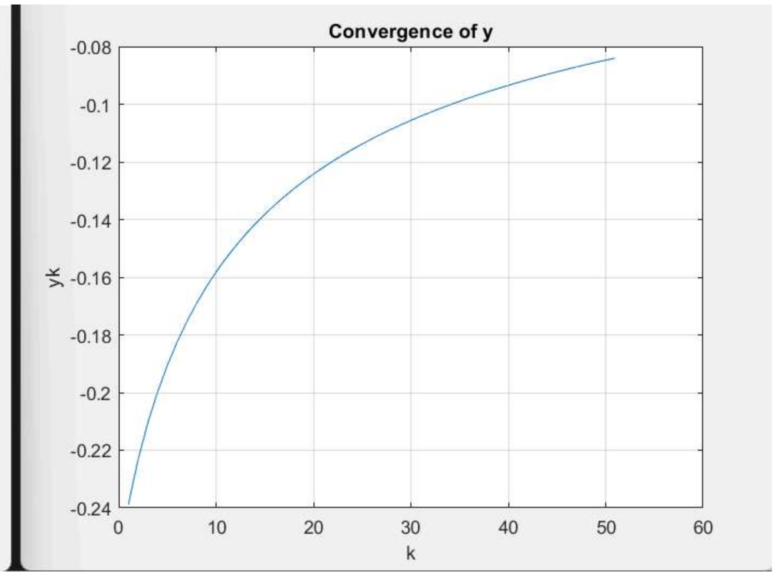
Αυτή η μέθοδος μας δίνει επίσης μια πολύ καλή ακρίβεια, ωστόσο με κόστος σε χρόνο, αφού παίρνει αρκετές παραπάνω επαναλήψεις.

Με step που δίνεται από την ελαχιστοποίση του f(χκ + γκdκ):









Βλέπουμε σημαντική μείωση του αριθμού επαναλήψεων ενώ τα αποτελέσματα δεν άλλαξαν καθόλου.

Συμπερασματικά μπορούμε να πούμε πως ο αλγόριθμος της μέγιστης καθόδου ξεχωρίζει για την απλότητα και ευκολία, ο αλγόριθμος Newton για την ταχύτητα σύγκλισης και ο αλγόριθμος Levenberg-Marquardt για την ακρίβεια του. Φυσικά, η αποδοτικότητα των μεθόδων εξαρτάται πολύ από την ίδια τη συνάρτηση που έχουν ως είσοδο, καθώς και από το βήμα και το αρχικό σημείο που εμείς διαλέγουμε. Είδαμε συγκεκριμένα πως για βήμα που βγαίνει από μια διαδικασία έναντι τυχαιότητας έχουμε γρηγορότερη σύγκλιση και επίσης είδαμε ότι το σημείο [0, 0] λόγω της θέσης του πάνω σε τοπικό ελάχιστο εγκλωβίστηκε σε αυτό.