Τεχνικές Βελτιστοποίησης

Αναφορά για την 2η εργαστηριακή άσκηση

Θεόδωρος Λιούπης
ΑΕΜ 9733

Χειμερινό εξάμηνο 2022-2023

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΑΠΘ

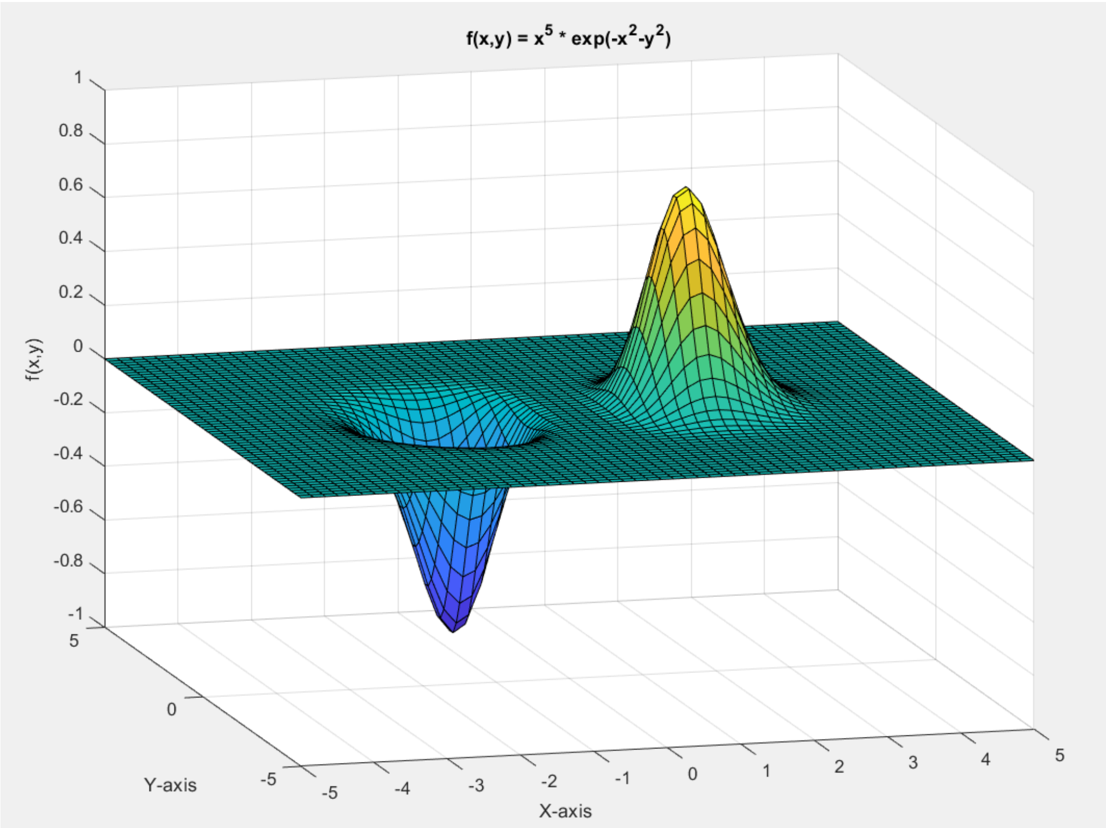
Ζητούμενο της 2ης εργαστηριακής άσκησης ήταν να ελαχιστοποιήσουμε μία δοσμένη συνάρτηση πολλών μεταβλητών με μεθόδους που χρησιμοποιούν παραγώγους και στηρίζονται στην επαναληπτική κάθοδο. Πιο συγκεκριμένα υλοποιήθηκαν οι αλγόριθμοι:

- Μέθοδος Μέγιστης Καθόδου (Steepest Descent)
- Μέθοδος Newton
- Μέθοδος Levenberg-Marquardt

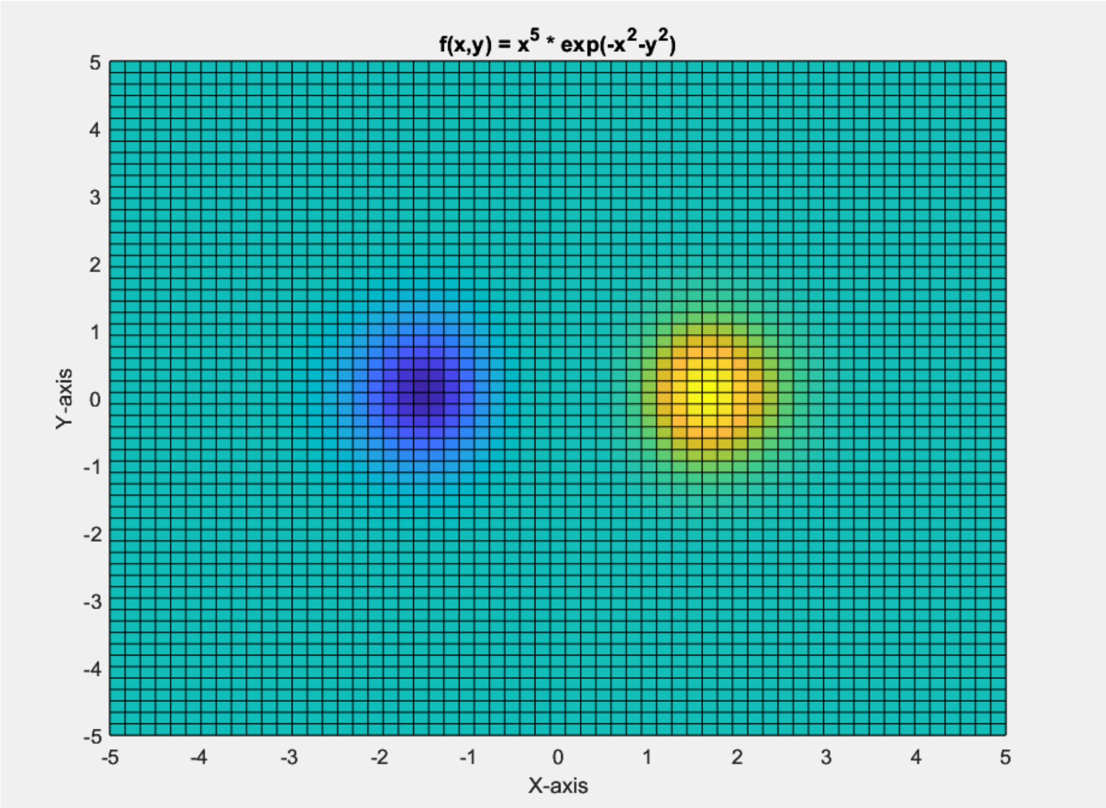
Η συνάρτηση που μας ζητήθηκε να ελαχιστοποιήσουμε ήταν η εξής:

$$f(x,y) = x^5 e^{-x^2-y^2}.$$

Στο πρώτο task της εργασίας έγινε plot η συνάρτηση ώστε να μπορέσουμε να την δούμε και να την καταλάβουμε καλύτερα.



Για να αποκτήσουμε μια ακόμα καλύτερη εικόνα για το που βρίσκεται το ελάχιστο της συνάρτησης, περιστρέφουμε το γράφημα ώστε να το δούμε από «πάνω» στις δύο διαστάσεις.



Διαισθητικά μπορούμε να πούμε ότι η συνάρτηση ελαχιστοποιείται κοντά στο σημείο (x , y) = (-1.5 , 0).

Στην συνέχεια, στα tasks 2, 3 και 4 γίνεται η υλοποίηση των αλγορίθμων που αναφέραμε παραπάνω. Σε κάθε μέθοδο, μελετάμε επιπλέον τις εξής περιπτώσεις:

- Βήμα γ_k σταθερό (επιλέχθηκε ίσο με 0.1)
- Βήμα γ_k τέτοιο ώστε να ελαχιστοποιεί την $f(x_k + \gamma_k * d_k)$
- Βήμα γ_k βάσει του κανόνα Armijo

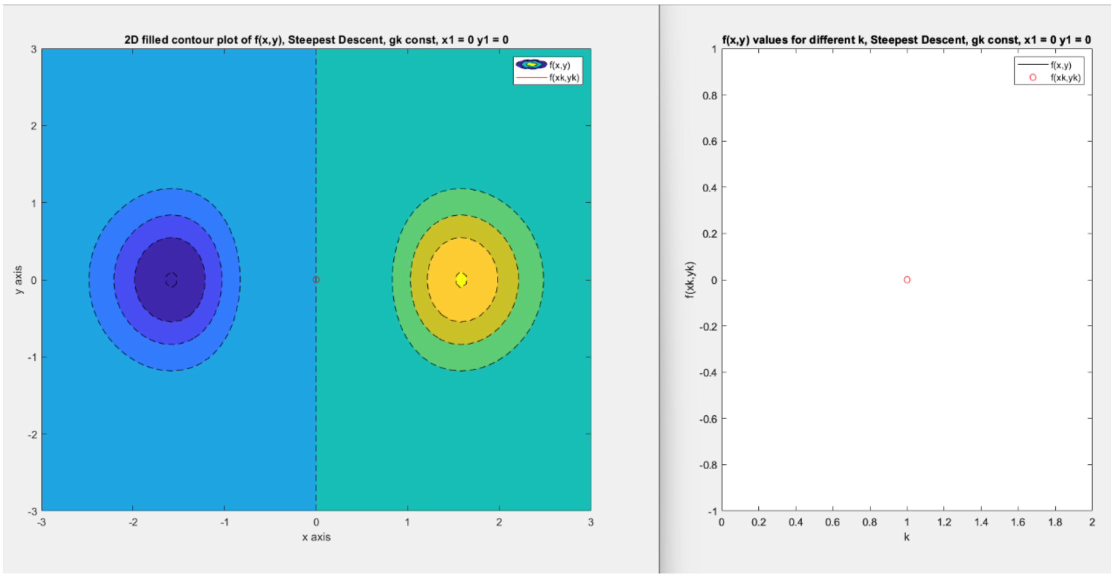
Τέλος, για κάθε μέθοδο και για κάθε περίπτωση συγκρίνουμε τα αποτελέσματα για διαφορετικά αρχικά σημεία έναρξης (x1,y1):

- (0,0)
- (-1,1)
- (1,-1)

Σε όλες τις μεθόδους επιλέχθηκε σταθερά τερματισμού $\varepsilon = 0.001$.

Η περίπτωση του σημείο έναρξης (0,0)

Όταν επιλέγουμε να ξεκινήσουμε από το σήμειο (0,0), κανένας αλγόριθμος δεν καταφέρνει να τρέξει καθώς «παγιδεύονται» εκεί. Αυτό συμβαίνει διότι, στο συγκεκριμένο σημείο, ισχύει $\nabla f\big((0,0)\big) = 0 < \varepsilon$ και έτσι τερματίζουν κατευθείαν. Τα αποτελέσματα που παίρνουμε είναι τα εξής:

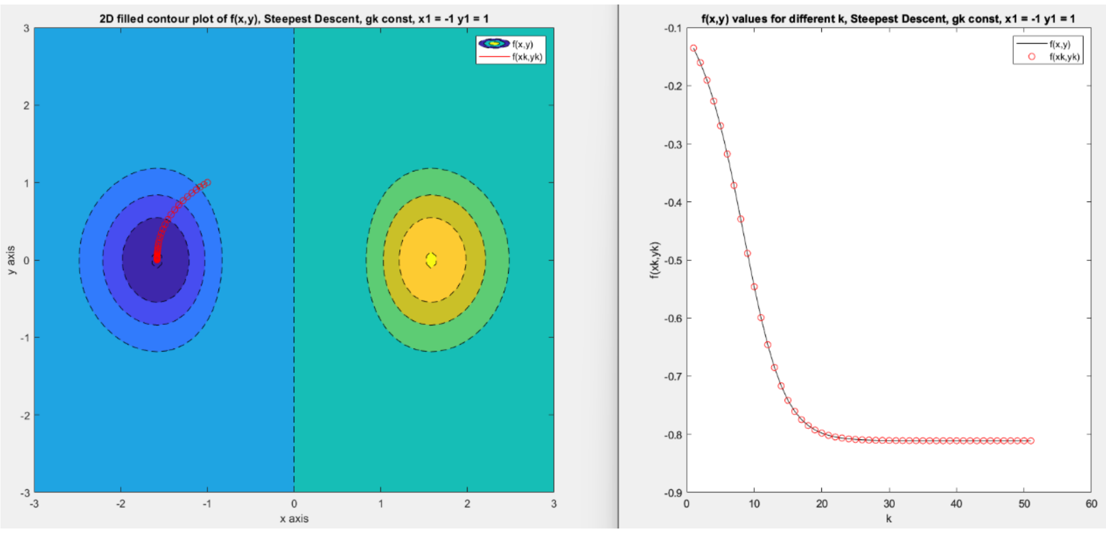


Το παραπάνω παράδειγμα είναι από την μέθοδο της μέγιστης καθόδου αλλά παίρνουμε ακριβώς το ίδιο και από τις δύο άλλες.

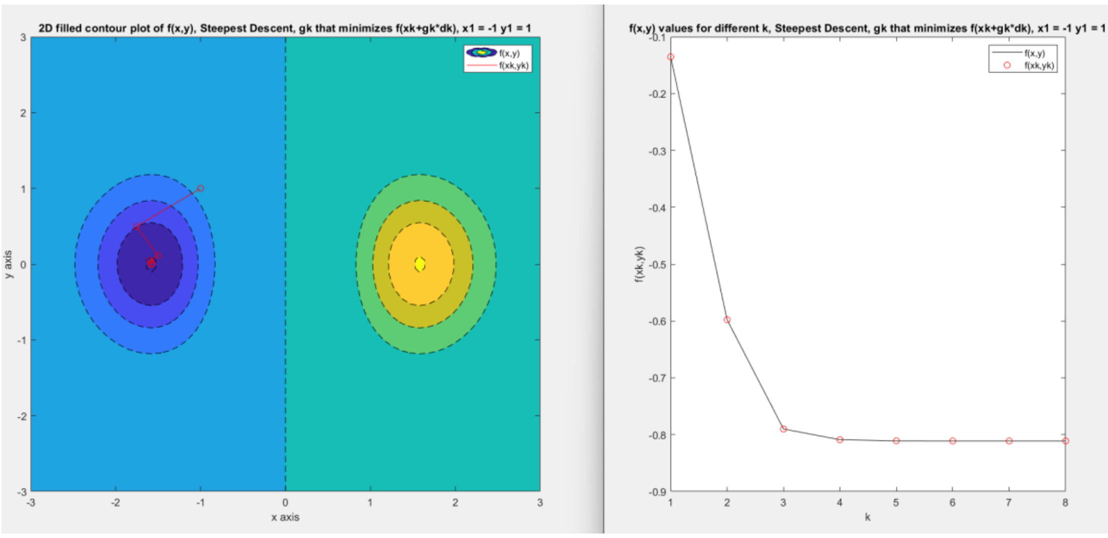
Μέθοδος Μέγιστης Καθόδου (Steepest Descent)

Αποτελέσματα για το σημείο έναρξης (-1,1):

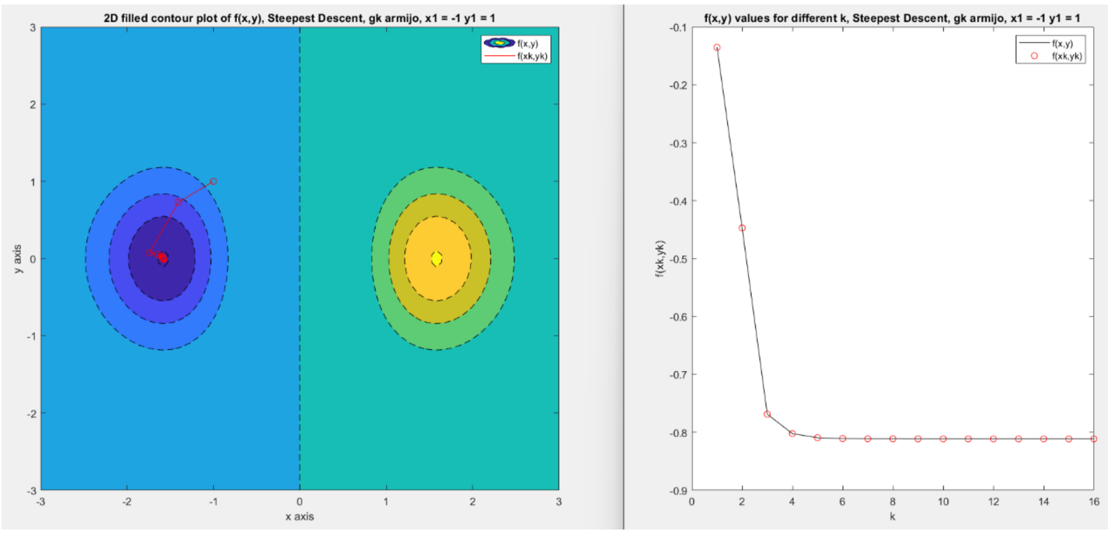
1) Βήμα γ_k σταθερό 0.1



2) Βήμα γ_k τέτοιο ώστε να ελαχιστοποιεί την $f(x_k + \gamma_k * d_k)$

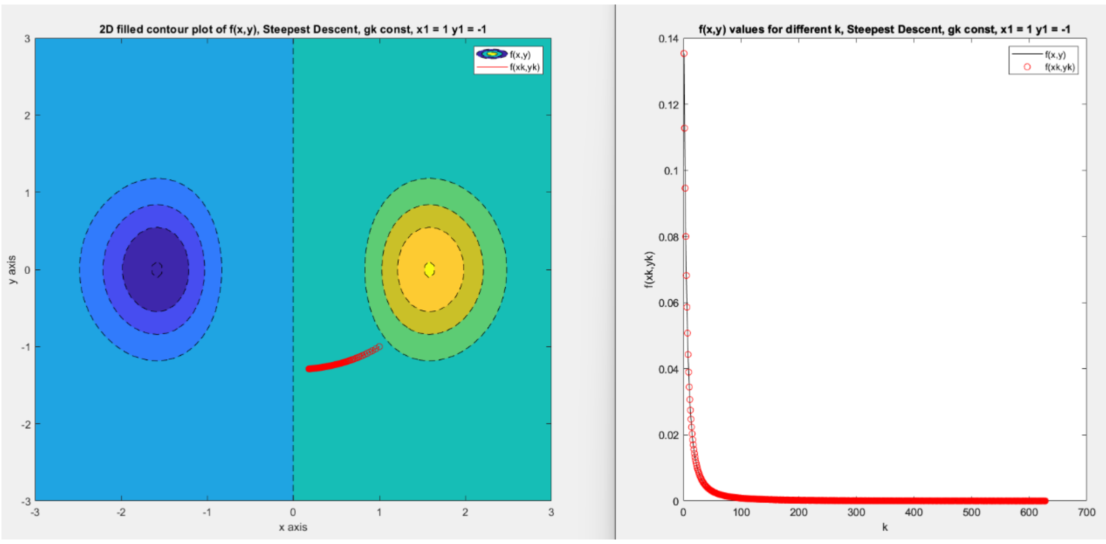


3) Βήμα γ_k βάσει του κανόνα Armijo

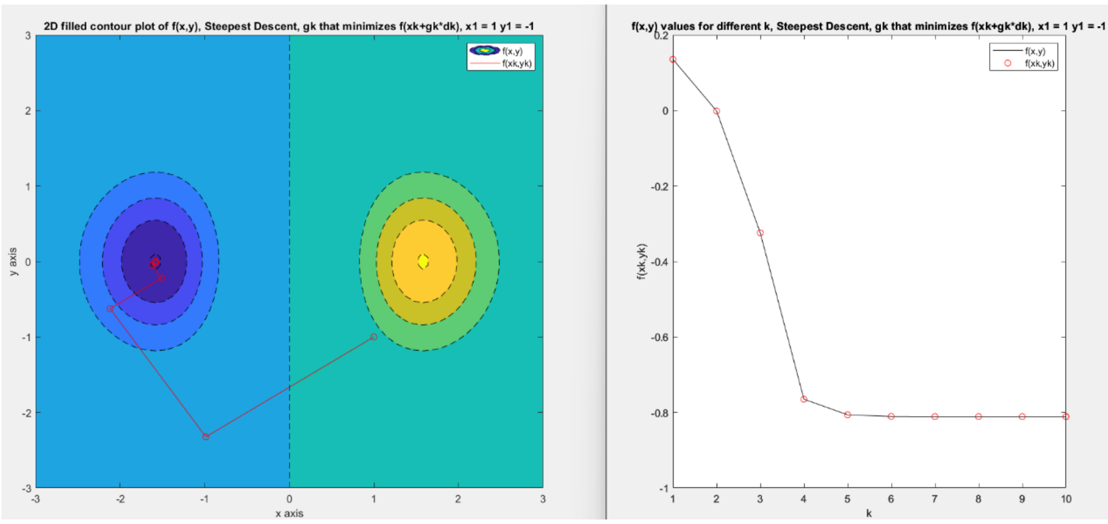


Αποτελέσματα για το σημείο έναρξης (1,-1):

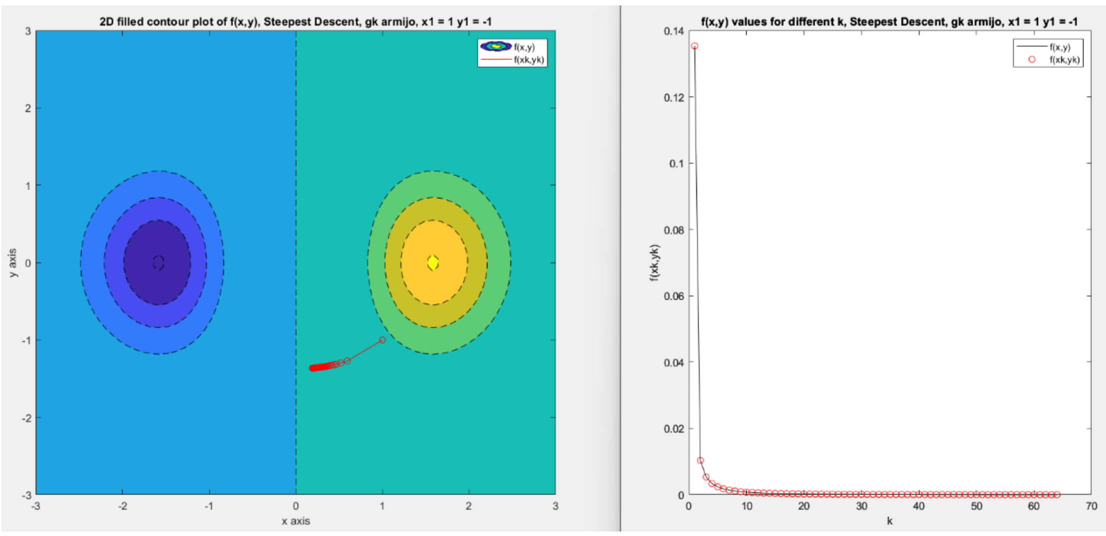
1) Βήμα γ_k σταθερό 0.1



2) Βήμα γ_k τέτοιο ώστε να ελαχιστοποιεί την $f(x_k + \gamma_k * d_k)$



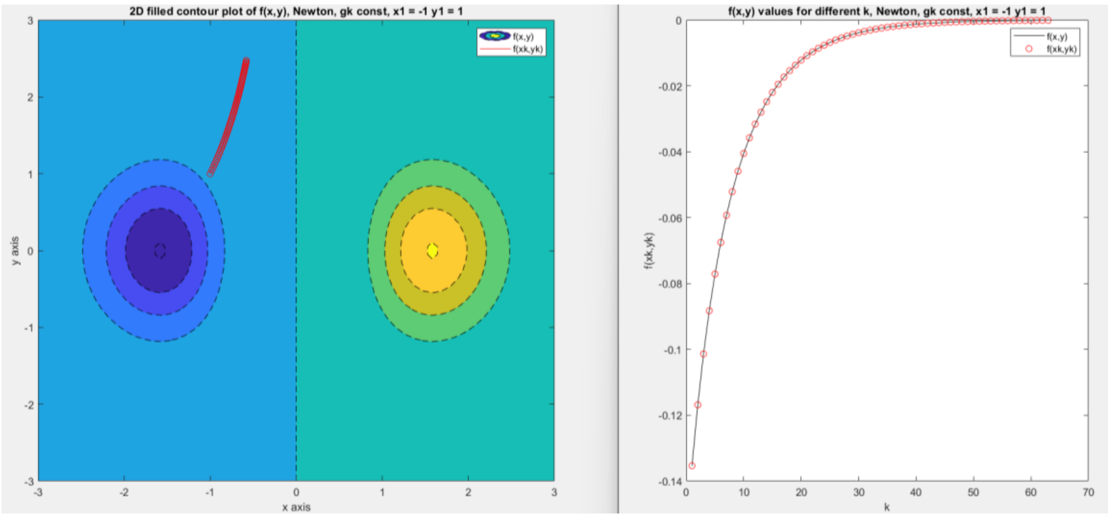
3) Βήμα γ_k βάσει του κανόνα Armijo



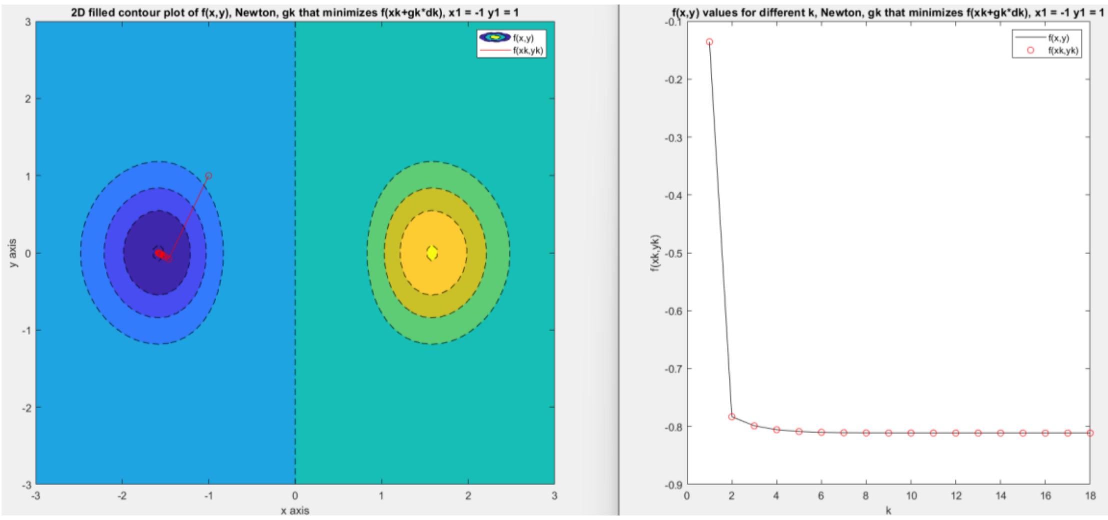
Μέθοδος Newton

Αποτελέσματα για το σημείο έναρξης (-1,1):

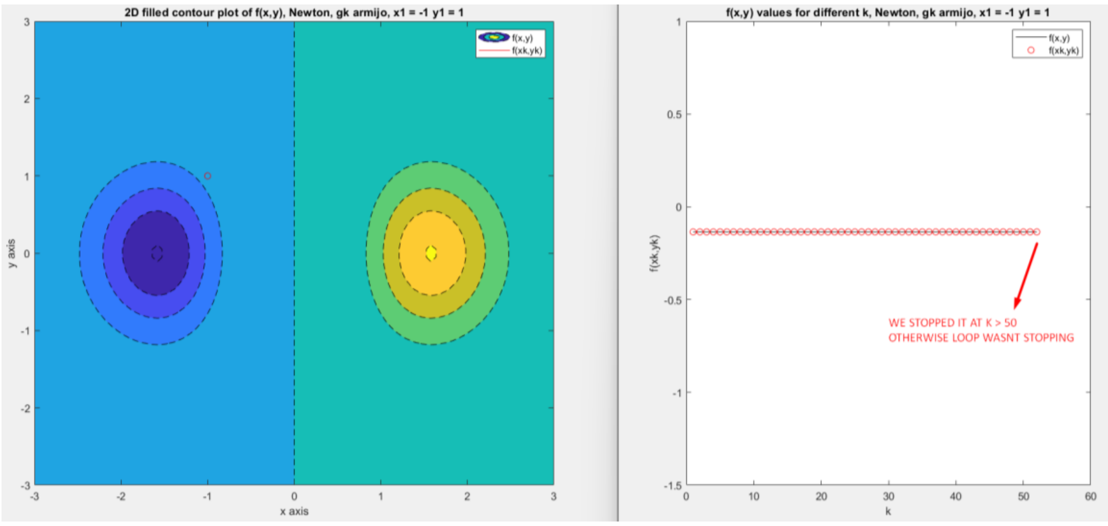
1) Βήμα γ_k σταθερό 0.1



2) Βήμα γ_k τέτοιο ώστε να ελαχιστοποιεί την $f(x_k + \gamma_k * d_k)$

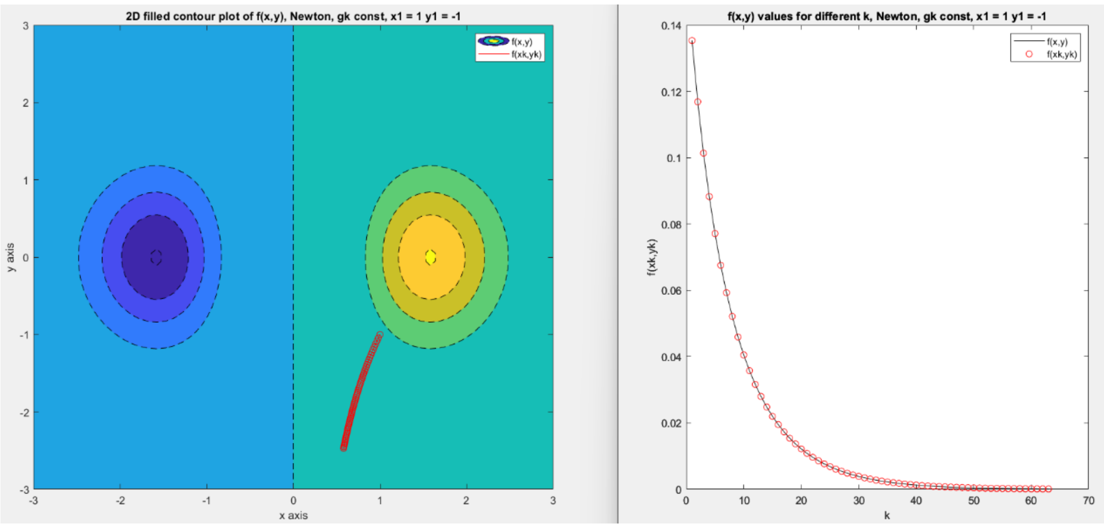


3) Βήμα γk βάσει του κανόνα Armijo

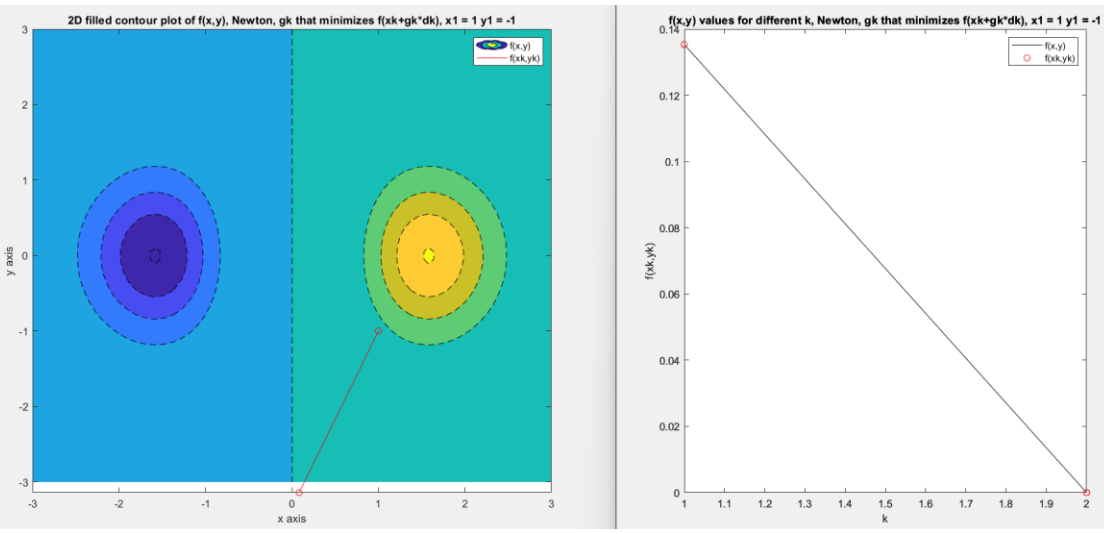


Αποτελέσματα για το σημείο έναρξης (1,-1):

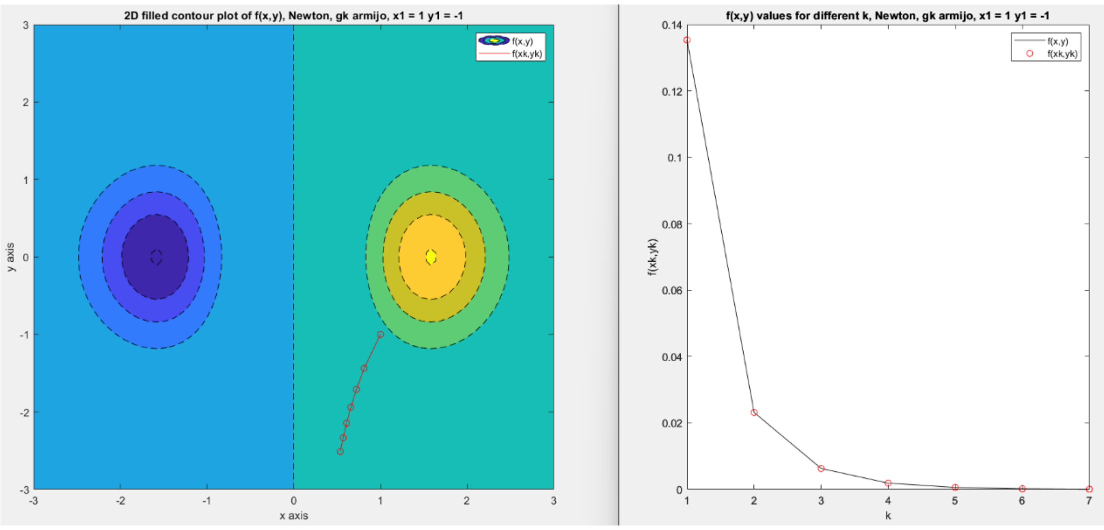
1) Βήμα γ_k σταθερό 0.1



2) Βήμα γ_k τέτοιο ώστε να ελαχιστοποιεί την $f(x_k+ \gamma_k*d_k)$



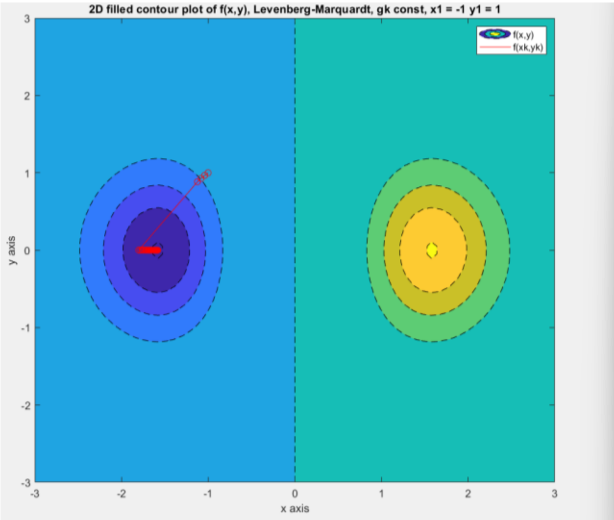
3) Βήμα γ_k βάσει του κανόνα Armijo

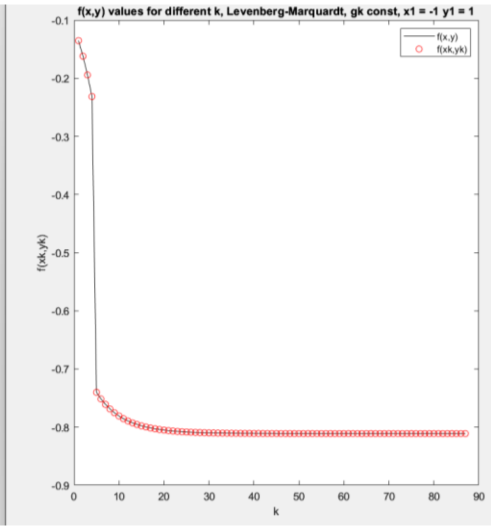


Μέθοδος Levenberg-Marquardt

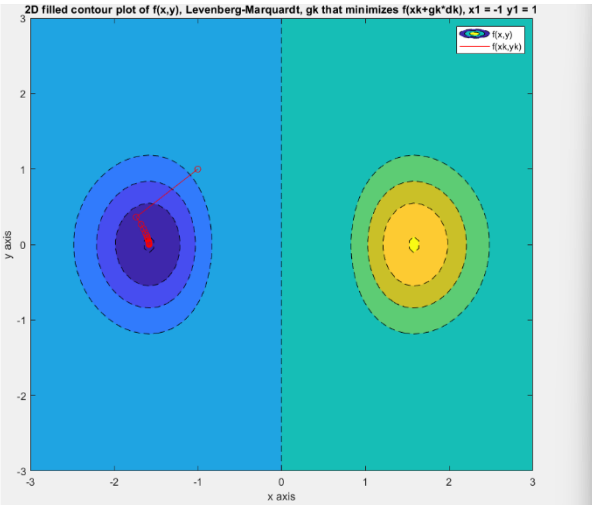
Αποτελέσματα για το σημείο έναρξης (-1,1):

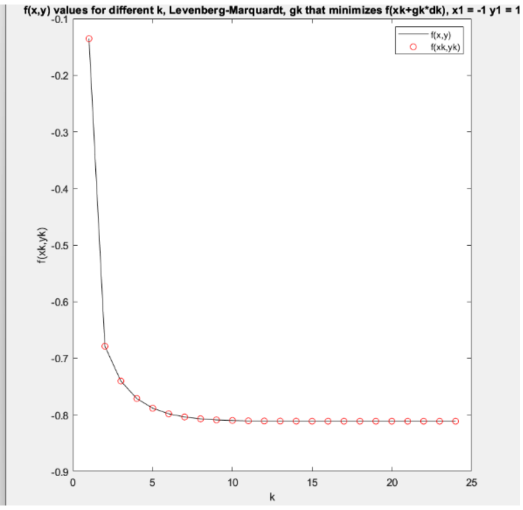
1) Βήμα γ_k σταθερό 0.1



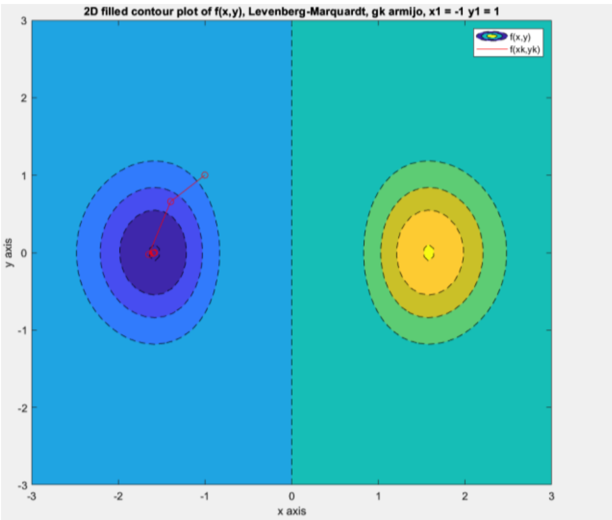


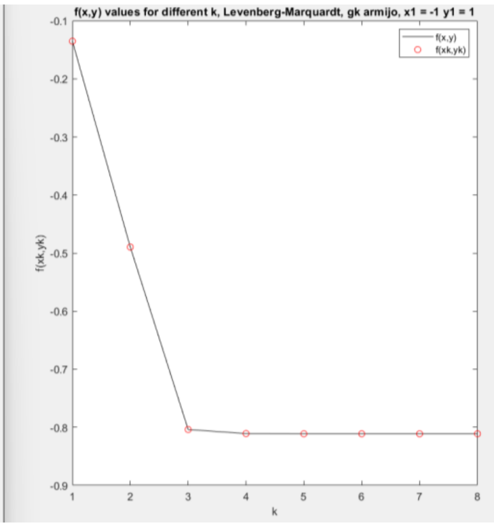
2) Βήμα γ_k τέτοιο ώστε να ελαχιστοποιεί την $f(x_k + \gamma_k * d_k)$





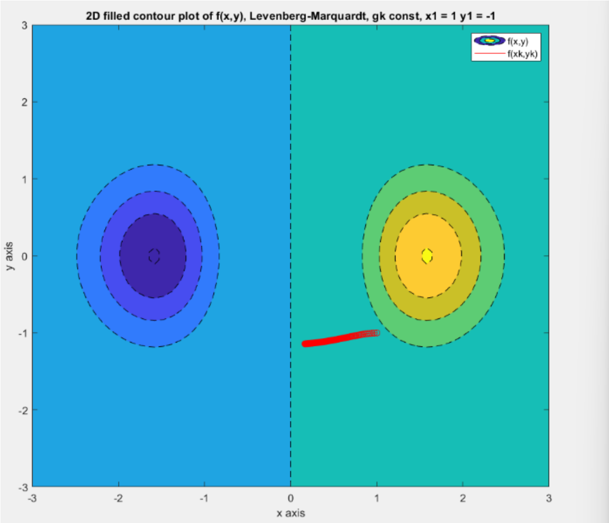
3) Βήμα γk βάσει του κανόνα Armijo

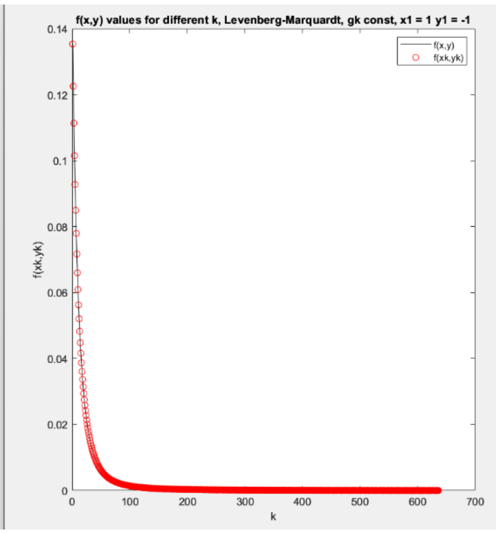




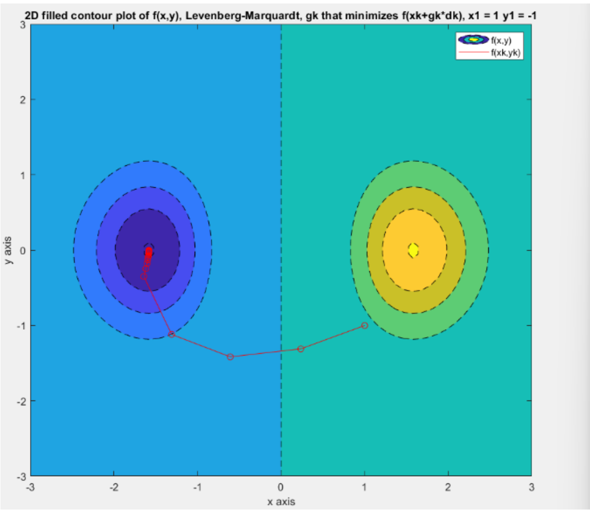
Αποτελέσματα για το σημείο έναρξης (1,-1):

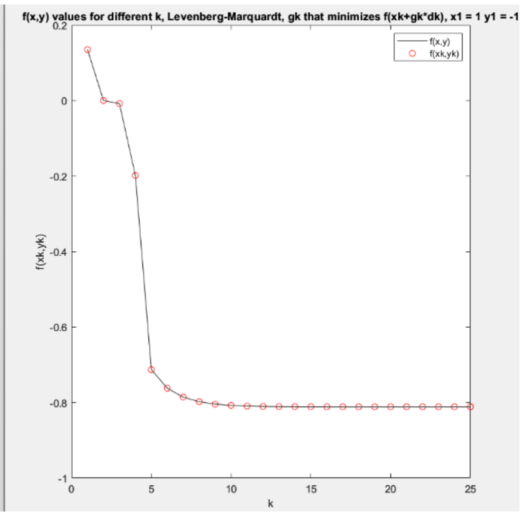
1) Βήμα γ_k σταθερό 0.1



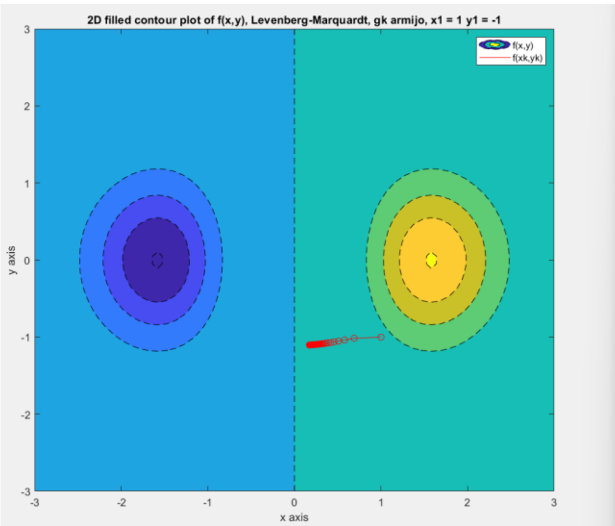


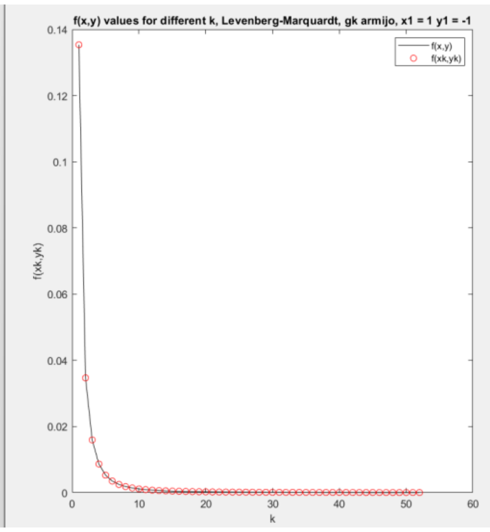
2) Βήμα γ_k τέτοιο ώστε να ελαχιστοποιεί την $f(x_k+ \gamma_k * d_k)$





3) Βήμα γ_k βάσει του κανόνα Armijo





Συμπεράσματα

Μέθοδος Μέγιστης Καθόδου (Steepest Descent)

- (0,0): Ο αλγόριθμος εγκλωβίζεται για τον λόγο που αναφέραμε στην αρχή $\nabla f\big((0,0)\big) = 0 < \varepsilon$.
- (-1,1): Ο αλγόριθμος συγκλίνει στο ολικό ελάχιστο της f(x,y) και στις τρεις περιπτώσεις επιλογής του βήματος γ_k.
- (1,-1):
    - γ_k σταθερό: Ο αλγόριθμος εγκλωβίζεται σε τοπικό ελάχιστο όπου f(x,y) = 0 και $\nabla f\big((x,y)\big) = 0 < \varepsilon$.
    - γ_k τέτοιο ώστε να ελαχιστοποιεί την $f(x_k + \gamma_k * d_k)$: Ο αλγόριθμος συγκλίνει στο ολικό ελάχιστο της f(x,y).
    - γ_k από Armijo: Ο αλγόριθμος εγκλωβίζεται σε τοπικό ελάχιστο όπου f(x,y) = 0 και $\nabla f\big((x,y)\big) = 0 < \varepsilon$.

Μέθοδος Newton

- (0,0): Ο αλγόριθμος εγκλωβίζεται για τον λόγο που αναφέραμε στην αρχή $\nabla f\big((0,0)\big) = 0 < \varepsilon$.
- (-1,1):
    - γ_k σταθερό: Ο αλγόριθμος εγκλωβίζεται σε τοπικό μέγιστο όπου f(x,y) = 0 και $\nabla f\big((x,y)\big) = 0 < \varepsilon$.
    - γ_k τέτοιο ώστε να ελαχιστοποιεί την $f(x_k + \gamma_k * d_k)$: Ο αλγόριθμος συγκλίνει στο ολικό ελάχιστο της f(x,y).
    - γ_k από Armijo: Ο αλγόριθμος δεν τρέχει καθόλου (μένει στο ίδιο σημείο) καθώς η ανισότητα του κανόνα Armijo δεν ικανοποιείται ποτέ.
- (1,-1): Ο αλγόριθμος εγκλωβίζεται σε τοπικό ελάχιστο όπου f(x,y) = 0 και $\nabla f\big((x,y)\big) = 0 < \varepsilon$ και στις τρεις περιπτώσεις επιλογής του βήματος γ_k.

Μέθοδος Levenberg-Marquardt

- (0,0): Ο αλγόριθμος εγκλωβίζεται για τον λόγο που αναφέραμε στην αρχή $\nabla f\big((0,0)\big) = 0 < \varepsilon$.
- (-1,1): Ο αλγόριθμος συγκλίνει στο ολικό ελάχιστο της f(x,y) και στις τρεις περιπτώσεις επιλογής του βήματος γ_k.
- (1,-1):
    - γ_k σταθερό: Ο αλγόριθμος εγκλωβίζεται σε τοπικό ελάχιστο όπου f(x,y) = 0 και $\nabla f\big((x,y)\big) = 0 < \varepsilon$.
    - γ_k τέτοιο ώστε να ελαχιστοποιεί την $f(x_k + \gamma_k * d_k)$: Ο αλγόριθμος συγκλίνει στο ολικό ελάχιστο της f(x,y).
    - γ_k από Armijo: Ο αλγόριθμος εγκλωβίζεται σε τοπικό ελάχιστο όπου f(x,y) = 0 και $\nabla f\big((x,y)\big) = 0 < \varepsilon$.

Σύγκριση των αλγορίθμων ως προς τις επαναλήψεις (k) που χρειάστηκαν για να τερματίσουν, άσχετα από την σύγκλιση τους ή όχι στο ολικό ελάχιστο:

Σημέιο (-1,1)	**Steepest Descent**	**Newton**	**Levenberg-Marquardt**
γ_k σταθερό	k = 51	k = 63	k = 88
γ_k για min f	k = 8	k = 18	k = 24
γ_k Armijo	k = 16	k = -	k = 8
Σημέιο (1,-1)			
γ_k σταθερό	k = ~620	k = 63	k = ~635
γ_k για min f	k = 10	k = 2	k = 25
γ_k Armijo	k = 64	k = 7	k = 52

Επιπλέον σχόλια

Ο αλγόριθμος Newton φαίνεται να υστερεί ως προς την σύγκλιση στο ολικό ελάχιστο της συνάρτησης αφού καταφέρνει να «φτάσει» μόνο στην περίπτωση της έναρξης από το σημείο (-1,1) και με βήμα γ_k τέτοιο ώστε να ελαχιστοποιεί την $f(x_k + \gamma_k * d_k)$.

Από τις τρεις περιπτώσεις επιλογής του βήματος γ_k η πιο αποτελεσματική στην δική μας περίπτωση, είναι εκείνη που βασίζεται στην ελαχιστοποίηση της f στο $x_k + \gamma_k * d_k$. Αυτό το συμπεραίνουμε διότι, με αυτήν την επιλογή, όλοι οι αλγόριθμοι καταφέρνουν να «φτάσουν» στο ολικό ελάχιστο της συνάρτησης και χρειάζονται λιγότερες επαναλήψεις k για να τερματίσουν.