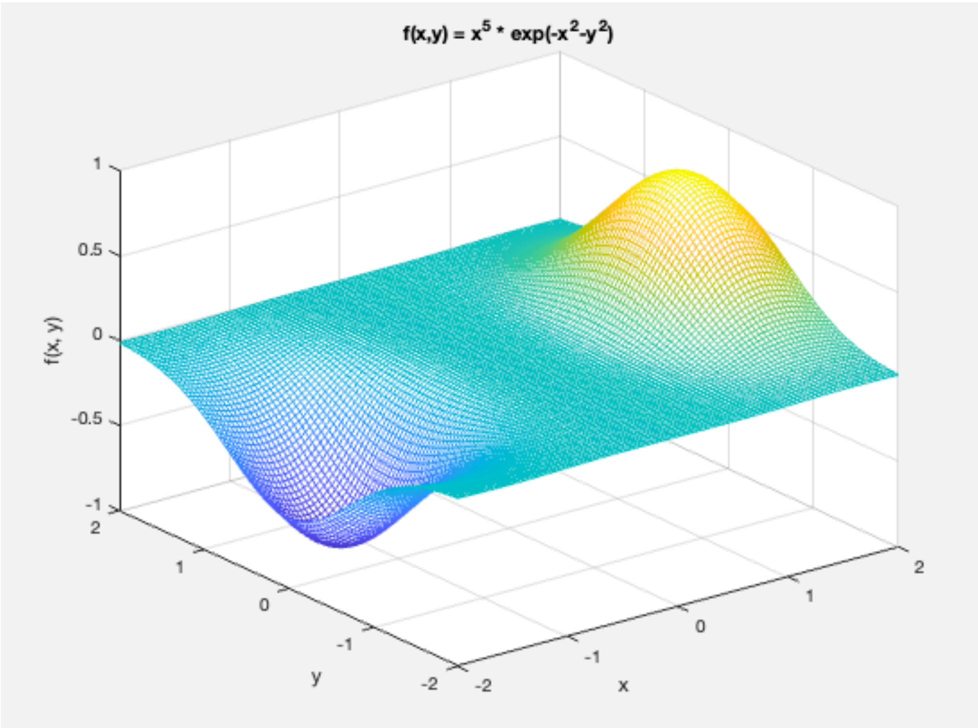
Τεχνικές Βελτιστοποίησης - 1η Εργαστηριακή Άσκηση

ΠΑΠΑΔΟΠΟΥΛΟΣ ΠΑΝΑΓΙΩΤΗΣ - 10697

Θέμα 1:

Η συνάρτηση που θα μελετηθεί είναι η $f(x, y) = x^5 e^{-x^2-y^2}$. Η γραφική της παράσταση φαίνεται στο παρακάτω σχήμα.



Όπως φαίνεται στο σχήμα, το ελάχιστο εμφανίζεται στο σημείο $(-1.581, 0)$. Επίσης, παρατηρείται ότι γύρω από το σημείο $(0,0)$ η συνάρτηση είναι εντελώς επίπεδη με μηδενική παράγωγο.

Θέμα 2:

- Εξήγηση της μεθόδου

Η μέθοδος της μέγιστης καθόδου αποτελεί μια αριθμητική διαδικασία βελτιστοποίησης, η οποία εφαρμόζεται ευρέως για τον εντοπισμό των ελαχίστων μιας συνάρτησης. Η κεντρική ιδέα έγκειται στην επιλογή ενός αρχικού σημείου x_0 στον χώρο των παραμέτρων και στη διαδοχική μετακίνηση προς την κατεύθυνση όπου η τιμή της συνάρτησης μειώνεται ταχύτερα, δηλαδή στην αντίθετη κατεύθυνση από την κλίση της.

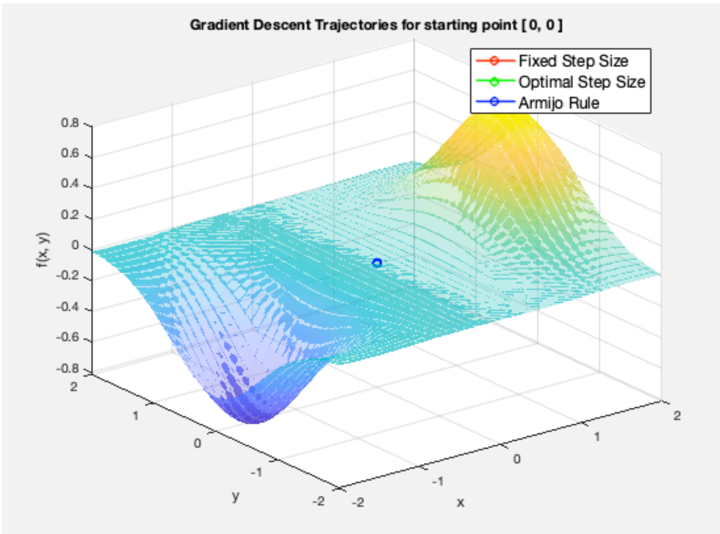
- Αλγόριθμος

Σε κάθε επαναληπτικό βήμα, η μέθοδος υπολογίζει την κλίση της συνάρτησης στο τρέχον σημείο, εκφραζόμενη ως το διάνυσμα $\nabla f(x_k)$, το οποίο υποδεικνύει την κατεύθυνση της μέγιστης αύξησης της συνάρτησης. Η ενημέρωση των παραμέτρων ακολουθεί τη σχέση $x_{k+1} = x_k - \gamma_k \nabla f(x_k)$, όπου k είναι το τρέχον επαναληπτικό βήμα και γ το βήμα. Το βήμα διαδραματίζει κρίσιμο ρόλο, καθώς εάν η τιμή του είναι υπερβολικά μεγάλη, η μέθοδος μπορεί να υπερβεί το ελάχιστο σημείο, ενώ μια υπερβολικά μικρή τιμή μπορεί να καταστήσει τη διαδικασία σύγκλισης ιδιαίτερα αργή. Η διαδικασία συνεχίζεται έως ότου επιτευχθεί ένα προκαθορισμένο κριτήριο τερματισμού, όπως η επίτευξη επαρκώς μικρής μεταβολής στην παράγωγο ή των αριθμών επαναλήψεων.

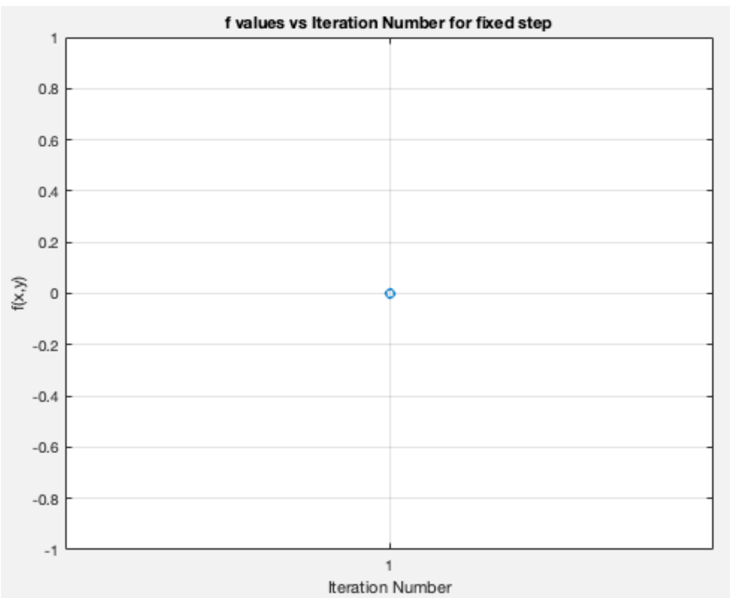
- Μετρήσεις

Τα παρακάτω αποτελέσματα προκύπτουν από υπολογισμούς μέσω του λογισμικού Matlab. Χρησιμοποιήθηκαν $\gamma = 0.5$, $\gamma_{armijo,0} = 1$, $\varepsilon = 10^{-4}$, $\alpha = 0.2$, $\beta = 0.01$ με 100 μέγιστες επαναλήψεις.

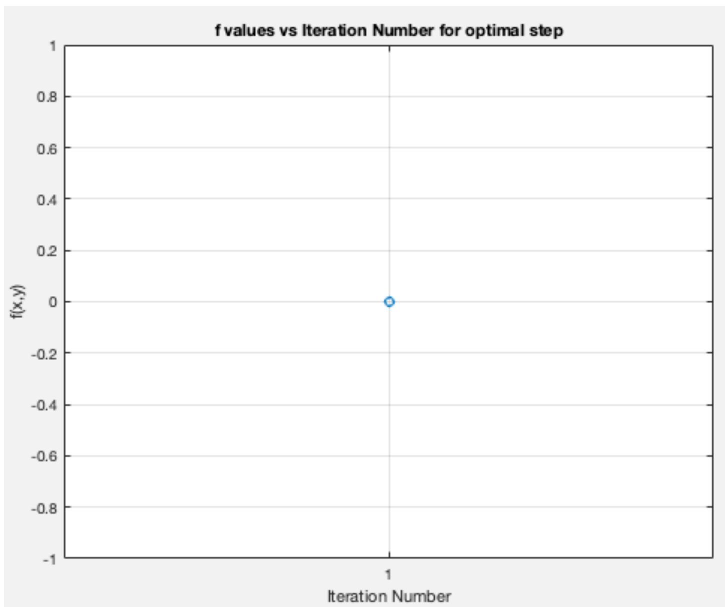
Για το σημείο $(0,0)$:



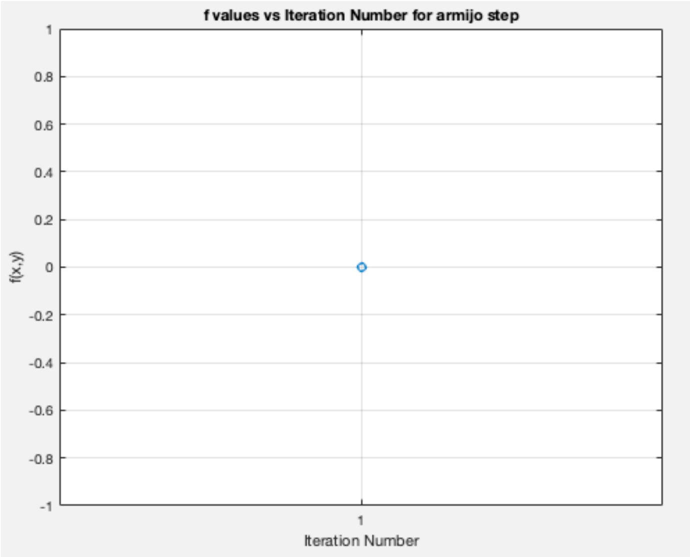
Γραφική Παράσταση της σύγκλισης της μεθόδου
με αρχικό σημείο το $(0,0)$



Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για σταθερό βήμα



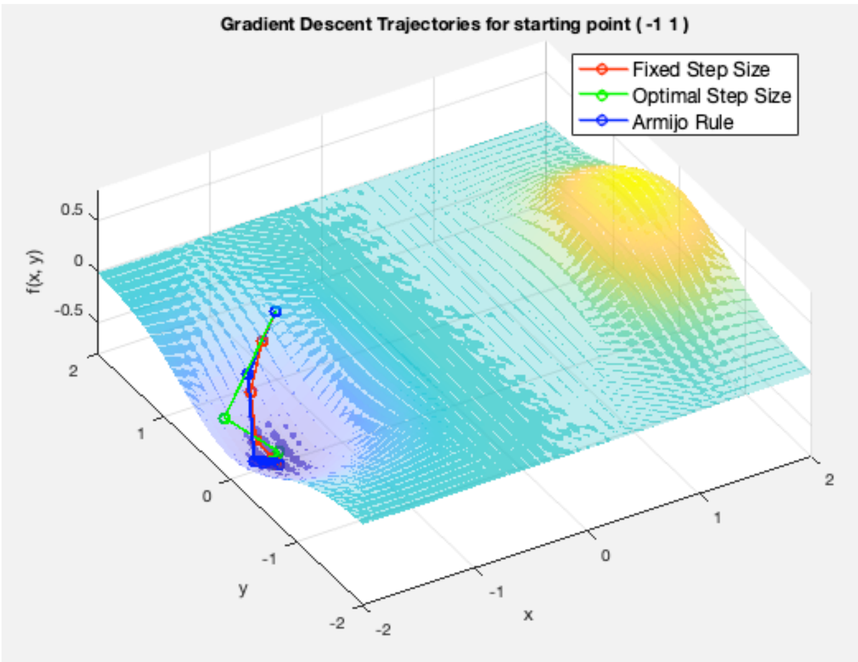
Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βέλτιστο βήμα



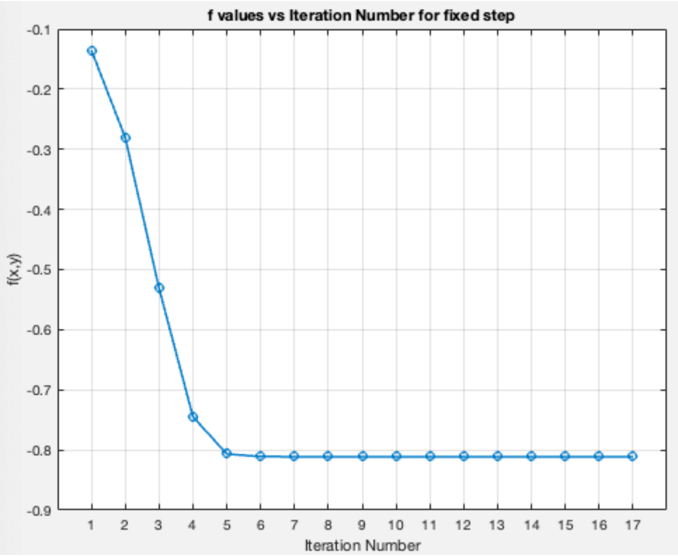
Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βήμα με κανόνα Armijo

Επειδή το αρχικό σημείο βρίσκεται σε επίπεδο σημείο, η κλίση της συνάρτησης είναι μηδενική, οπότε ο αλγόριθμος τερματίζει με την πρώτη επανάληψη, καθώς πληρείται η προϋπόθεση $|\nabla f(x_k)| < \epsilon$.

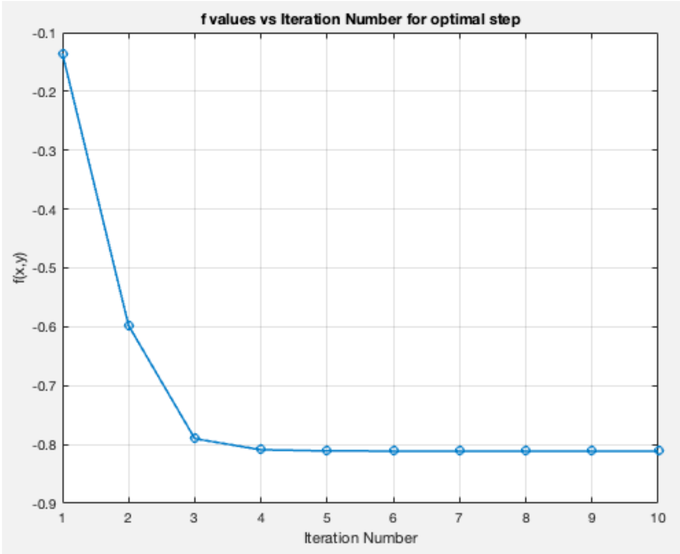
Για το σημείο $(-1,1)$:



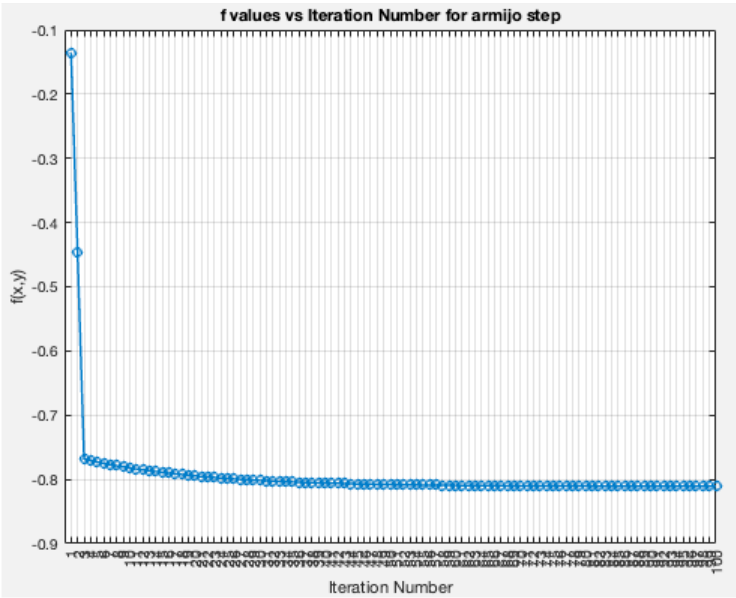
Γραφική Παράσταση της σύγκλισης της μεθόδου με αρχικό σημείο το $(-1,1)$



Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για σταθερό βήμα



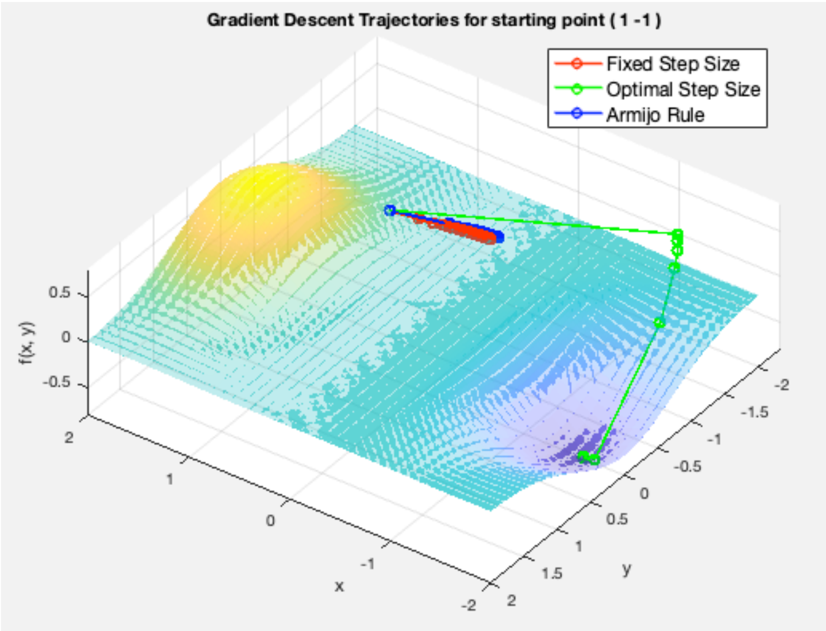
Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βέλτιστο βήμα



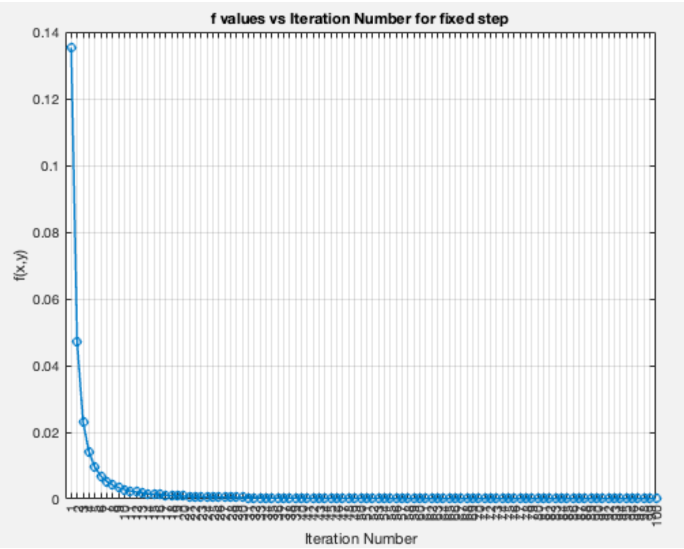
Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βήμα με κανόνα Armijo

Παρατηρείται ότι η επιλογή του βήματος με την ελαχιστοποίηση της $f(x_k + \gamma_k d_k)$ είναι σχεδόν δύο φορές πιο γρήγορη από την επιλογή σταθερού βήματος. Ωστόσο, και στις δυο περιπτώσεις ο αλγόριθμος συγκλίνει και τερματίζει λόγω του ανεκτού εύρους και όχι της υπέρβασης των επαναλήψεων. Από την άλλη πλευρά, η επιλογή βήματος με τον κανόνα Armijo αν και συγκλίνει στο σωστό σημείο, δεν φτάνει το επιθυμητό όριο ε που έχει τεθεί.

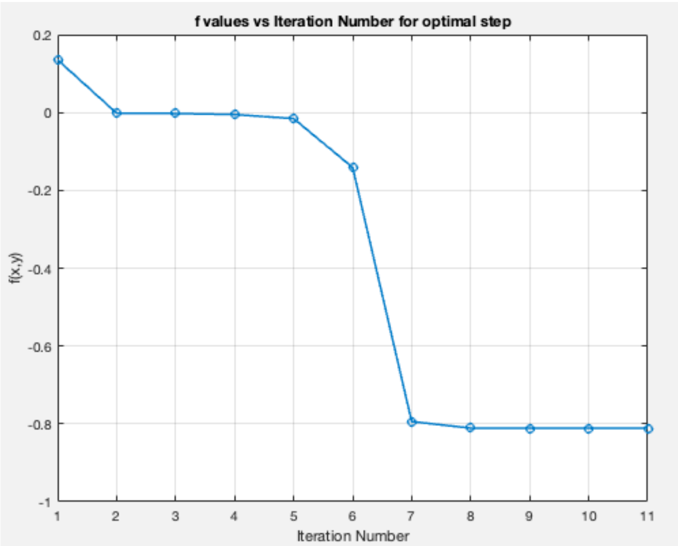
Για το σημείο $(1, -1)$:



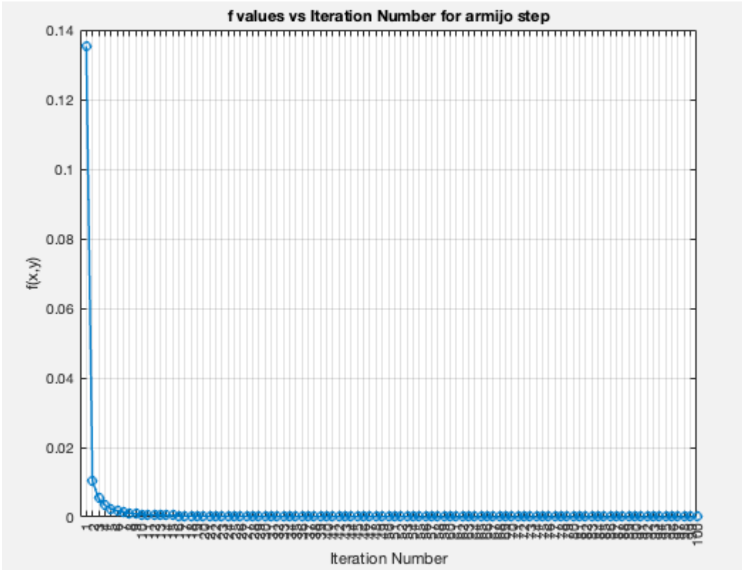
Γραφική Παράσταση της σύγκλισης της μεθόδου με αρχικό σημείο το $(1, -1)$



Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για σταθερό βήμα

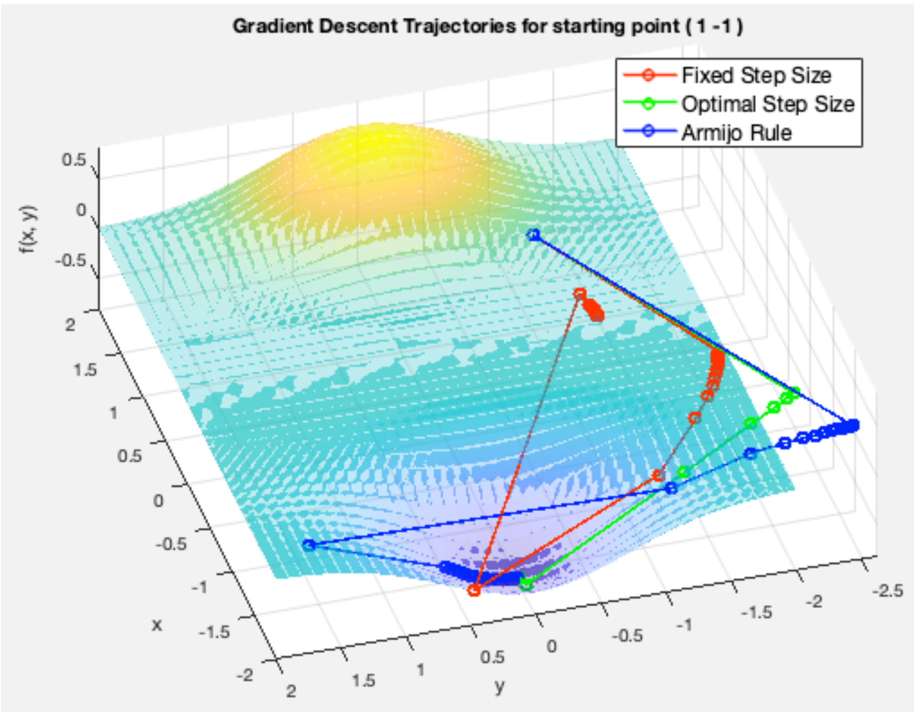


Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βέλτιστο βήμα

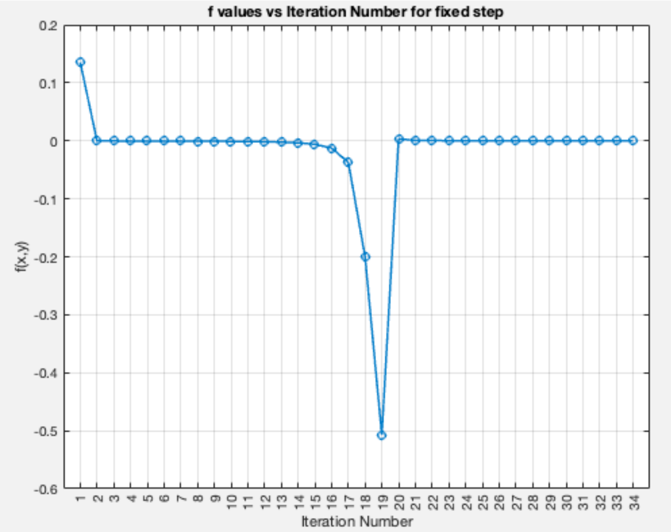


Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βήμα με κανόνα Armijo

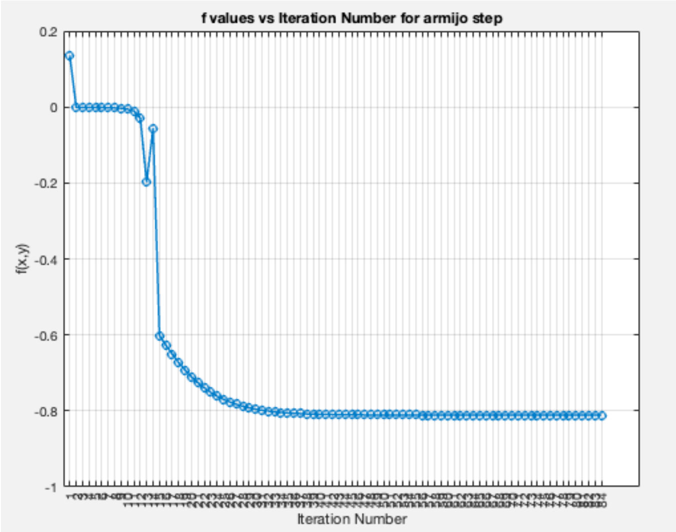
Είναι προφανές ότι για σταθερό βήμα και βήμα βάσει του κανόνα Armijo οι μέθοδοι δυσκολεύονται όταν συναντούν την επίπεδη περιοχή της συνάρτησης. Αντιθέτως, το βήμα με ελαχιστοποίηση της $f(x_k + \gamma_k d_k)$ συγκλίνει χωρίς κανένα πρόβλημα. Γίνονται δοκιμές και πράγματι για αρχικό βήμα $\gamma = 6$ η μέθοδος του κανόνα Armijo συγκλίνει μετά από 84 επαναλήψεις. Αντιθέτως, αυξάνοντας το σταθερό βήμα σε $\gamma = 3.5$, ο αλγόριθμος ξεφεύγει από την επίπεδη περιοχή και πλησιάζει το σημείο, επειδή όμως δεν μπορεί να ρυθμίσει το βήμα επανέρχεται σε επίπεδη περιοχή και τερματίζει.



Γραφική Παράσταση της σύγκλισης της μεθόδου με αρχικό σημείο το $(1, -1)$ με τροποποιημένα βήματα



Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για νέο σταθερό βήμα $\gamma = 3.5$



Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βήμα με κανόνα Armijo με αρχικό $\gamma = 6$

Καταληκτικά, οι αλγόριθμοι αυτοί εξαρτώνται σε μεγάλο βαθμό από το αρχικό σημείο. Αν το αρχικό σημείο βρίσκεται σε επίπεδη περιοχή, κανένας αλγόριθμος δεν μπορεί να ξεφύγει, καθώς δεν έχει υπόδειξη προς τα που να κινηθεί λόγω της μηδενικής παραγώγου. Αν το αρχικό σημείο έχει αρνητική κλίση και είναι κοντά στο ελάχιστο, τότε κανένας αλγόριθμος δεν δυσκολεύεται και βρίσκουν το σημείο με σχετική ευκολία, έχοντας εμφανώς πιο αποδοτικό αυτόν που ελαχιστοποιεί την $f(x_k + \gamma_k d_k)$ για να βρει το βήμα, ακολουθούμενος από αυτόν με σταθερό βήμα και τέλος αυτός με το βήμα του κανόνα Armijo. Στην περίπτωση που το αρχικό σημείο δεν είναι κοντά στο ελάχιστο αλλά έχει μη μηδενική παράγωγο, ο πρώτος αλγόριθμος το εντοπίζει με ευκολία. Αυτός που χρησιμοποιεί τον κανόνα Armijo χρειάζεται τις κατάλληλες παραμέτρους για να βρει τη σωστή μείωση ή αύξηση του βήματος για να ‘ξεφύγει’ από κάποια τυχόν επίπεδη περιοχή και να βρει αρνητική κλίση. Ο αλγόριθμος με σταθερό βήμα χρειάζεται ένα αρκετά μεγάλο βήμα για να αποφύγει τυχόν επίπεδες περιοχές, δεν έχει όμως τρόπο να μικρύνει το βήμα του καθώς πλησιάζει στο αναζητούμενο σημείο και απομακρύνεται από αυτό. Τελικά, η μέθοδος αυτή δεν καταλήγει πάντα στο σωστό αποτέλεσμα. Το αποτέλεσμα εξαρτάται πλήρως από το αρχικό σημείο, την ίδια τη συνάρτηση και τον τρόπο που επιλέγεται το βήμα γ_k.

Θέμα 3:

- Εξήγηση της μεθόδου

Η μέθοδος Newton στηρίζεται στη χρήση του gradient και του εσσιανού πίνακα. Η διαδικασία ξεκινά από ένα αρχικό σημείο στο χώρο των παραμέτρων και βασίζεται σε διαδοχικές βελτιώσεις του σημείου αυτού, με στόχο να οδηγηθεί στο τοπικό ελάχιστο της συνάρτησης. Σε κάθε βήμα, υπολογίζεται το gradient της συνάρτησης, το οποίο περιγράφει την κατεύθυνση της μεγαλύτερης αύξησής της, και ο εσσιανός πίνακας, ο οποίος προσφέρει πληροφορίες για την καμπυλότητα της συνάρτησης στο σημείο αυτό. Ο υπολογισμός της νέας θέσης γίνεται μέσω της επίλυσης ενός γραμμικού συστήματος που περιλαμβάνει τον εσσιανό πίνακα και το gradient, ώστε να καθοριστεί η βέλτιστη κατεύθυνση καθόδου. Απαιτείται προσοχή καθώς η μέθοδος απαιτεί ο εσσιανός πίνακας να είναι θετικά ορισμένος.

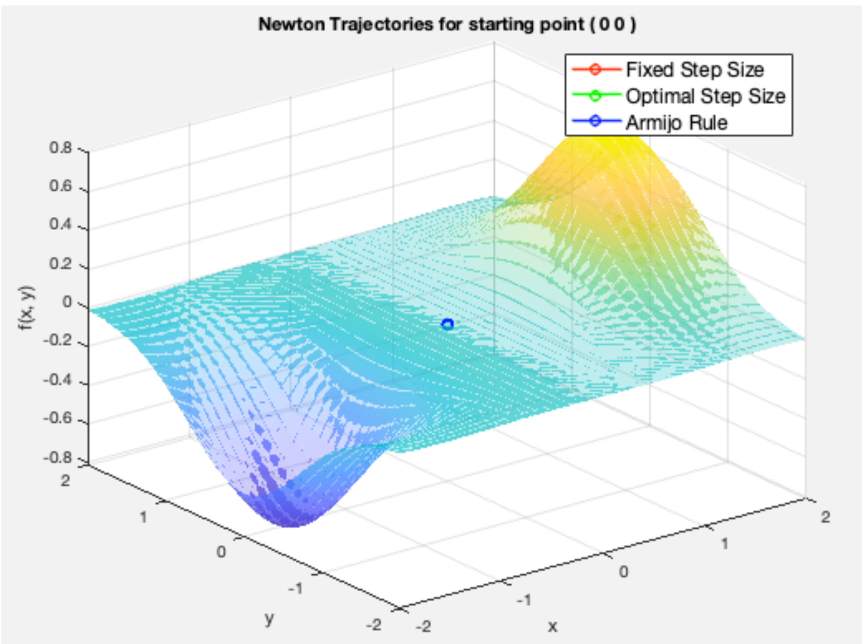
- Αλγόριθμος

Ο αλγόριθμος της ξεκινά από ένα αρχικό σημείο και χρησιμοποιεί επαναληπτική διαδικασία για να φτάσει σε ένα τοπικό ελάχιστο. Στο πρώτο βήμα, υπολογίζεται το gradient της συνάρτησης στο σημείο εκκίνησης, το οποίο δίνει την κατεύθυνση της μεγαλύτερης αύξησης της συνάρτησης, και ο εσσιανός πίνακας, ο οποίος περιγράφει την καμπυλότητα της συνάρτησης και προσδιορίζει πώς μεταβάλλεται το gradient στον χώρο. Η βέλτιστη διόρθωση, δηλαδή η κατεύθυνση του επόμενου βήματος, υπολογίζεται με την επίλυση του γραμμικού συστήματος $d_k = -\,[\nabla^2 f(x_k)]^{-1}\nabla f(x_k)$. Το νέο σημείο υπολογίζεται προσθέτοντας στη θέση του τρέχοντος σημείου την κατεύθυνση καθόδου πολλαπλασιασμένη με το βήμα, το οποίο μπορεί να είναι είτε σταθερό είτε δυναμικά υπολογιζόμενο. Ο αλγόριθμος συνεχίζει να επαναλαμβάνει αυτή τη διαδικασία για κάθε σημείο, υπολογίζοντας εκ νέου το gradient, τον εσσιανό πίνακα, και την κατεύθυνση καθόδου, μέχρι να πληρούνται τα κριτήρια τερματισμού. Τα κριτήρια τερματισμού περιλαμβάνουν είτε το gradient να έχει προσεγγίσει το μηδέν, δηλαδή το σημείο να είναι αρκετά κοντά στο τοπικό ελάχιστο, είτε ο εσσιανός πίνακας να μην είναι θετικά ορισμένος είτε να έχει υπερβεί το προκαθορισμένο πλήθος επαναλήψεων.

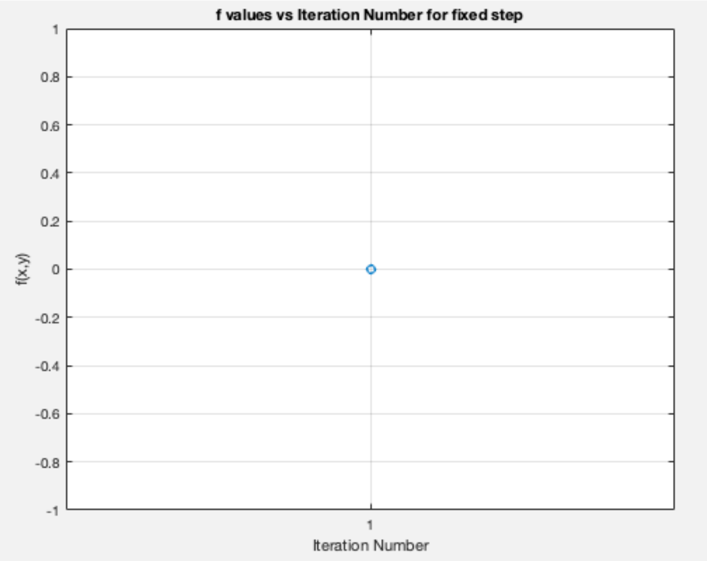
- Μετρήσεις

Τα παρακάτω αποτελέσματα προκύπτουν από υπολογισμούς μέσω του λογισμικού Matlab. Χρησιμοποιήθηκαν $\gamma_{fixed} = 0.5$, $\gamma_{armijo,0} = 1$, $\varepsilon = 10^{-4}$, $\alpha = 0.2$, $\beta = 0.01$ με 100 μέγιστες επαναλήψεις.

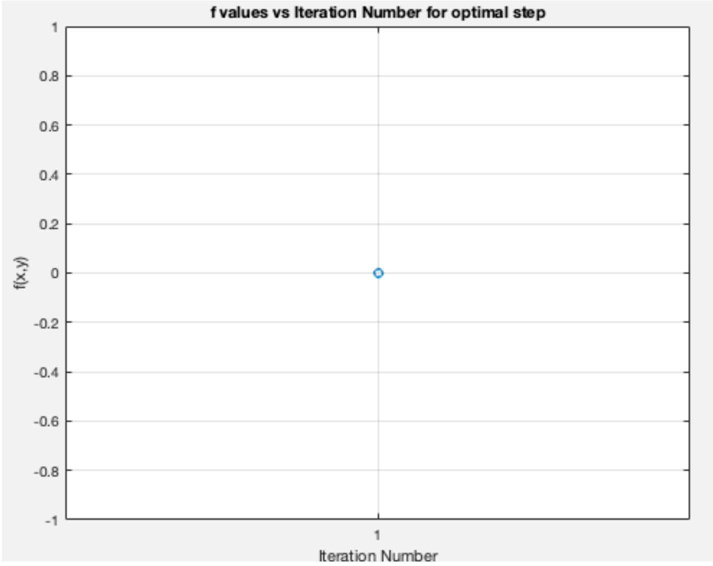
Για το σημείο $(0{,}0)$:



Γραφική Παράσταση της σύγκλισης της μεθόδου με αρχικό σημείο το $(0{,}0)$



Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για σταθερό βήμα

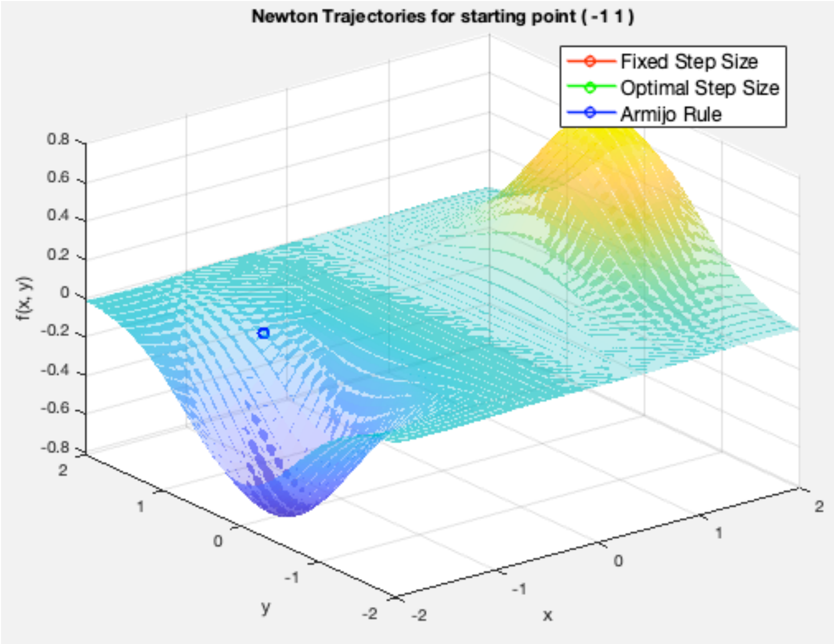


Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βέλτιστο βήμα

Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βήμα με κανόνα Armijo

Ομοίως με την προηγούμενη μέθοδο, επειδή το αρχικό σημείο βρίσκεται σε επίπεδο σημείο, πληρείται από την πρώτη κιόλας επανάληψη η προϋπόθεση $|\nabla f(x_k)| < \epsilon$ επειδή η κλίση της συνάρτησης είναι μηδενική, οπότε ο αλγόριθμος τερματίζει κατευθείαν..

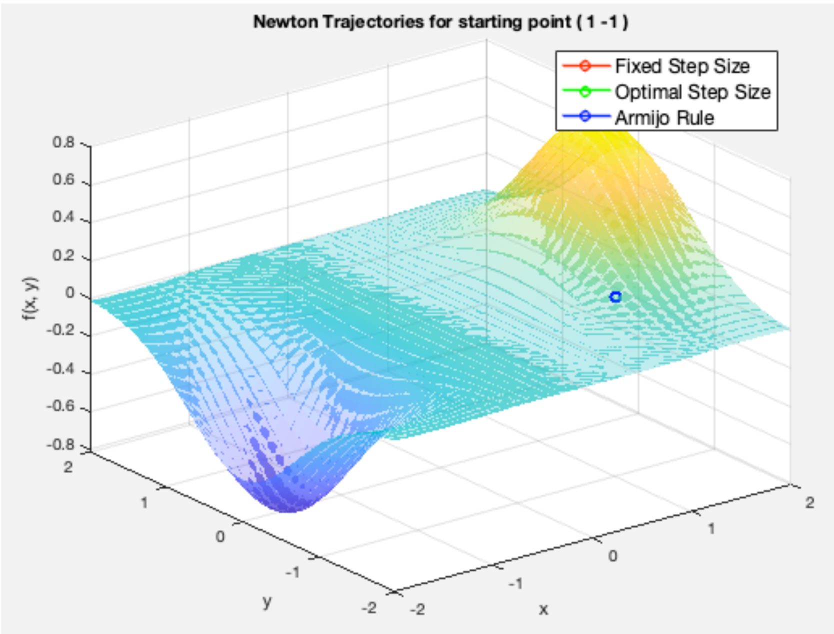
Για το σημείο $(-1,1)$:



Γραφική παράσταση της σύγκλισης της μεθόδου με αρχικό σημείο το $(-1,1)$

Από την πρώτη κιόλας επανάληψη ο εσσιανός πίνακας δεν είναι θετικά ορισμένος. Επομένως η μέθοδος δεν μπορεί να εφαρμοστεί σε αυτή την περίπτωση.

Για το σημείο $(1, -1)$:



Γραφική παράσταση της σύγκλισης της μεθόδου με αρχικό σημείο το $(1, -1)$

Ομοίως με το προηγούμενο αρχικό σημείο, ο εσσιανός πίνακας δεν είναι θετικά ορισμένος και η μέθοδος δεν μπορεί να εφαρμοστεί καθώς δεν πληρούνται οι προϋποθέσεις που αναφέρθηκαν αρχικά.

Τελικά, η μέθοδος Newton δεν οδηγεί πάντα σε σωστό αποτέλεσμα. Η μέθοδος έχει περισσότερες προϋποθέσεις που πρέπει να πληρούνται για να λειτουργήσει. Πράγματι, εξαρτάται από το αρχικό σημείο, καθώς για το αρχικό σημείο $(0,0)$ ο αλγόριθμος σύγκλινε, με τον ίδιο όμως τρόπο όπως η προηγούμενη μέθοδος λόγω της επίπεδης επιφάνειας γύρω του σημείου. Ακόμη, η ίδια η συνάρτηση επηρρεάζει τη λειτουργία του αλγορίθμου, καθώς ο εσσιανός πίνακας πρέπει σε κάθε εξεταζόμενο σημείο να είναι θετικά ορισμένος και να μην υπάρχει επίπεδη περιοχή γύρω από το αρχικό σημείο. Η επιλογή του βήματος δεν μπόρεσε να εξεταστεί.

Θέμα 4:

- Εξήγηση της μεθόδου

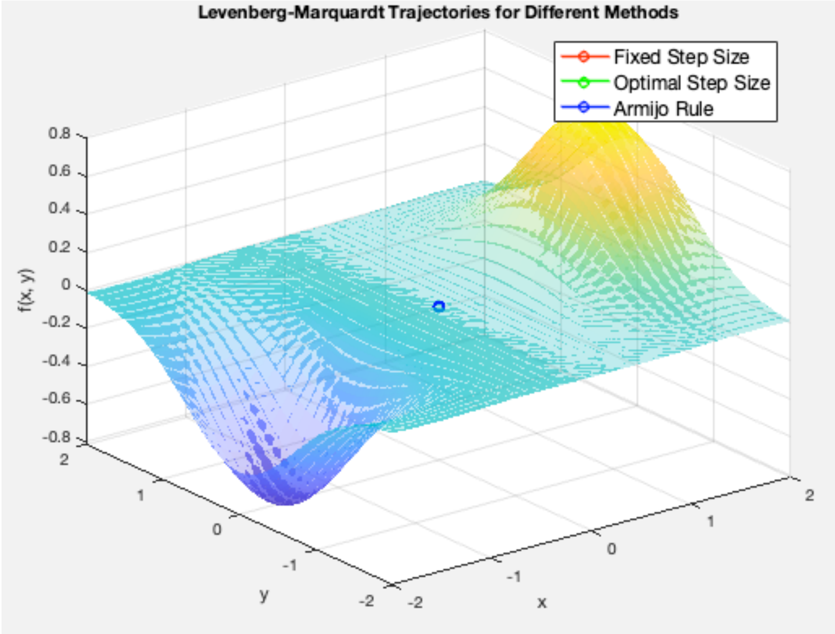
Η μέθοδος Levenberg-Marquardt είναι ένας υβριδικός αλγόριθμος βελτιστοποίησης που συνδυάζει στοιχεία της μεθόδου Newton και της μέγιστης καθόδου για την ελαχιστοποίηση μη γραμμικών συναρτήσεων. Στόχος της είναι η σταθερότερη και αποδοτικότερη σύγκλιση σε σχέση με την απλή μέθοδο Newton. Η μέθοδος ξεκινά από ένα αρχικό σημείο και επαναληπτικά βελτιώνει τη λύση, χρησιμοποιώντας έναν τροποποιημένο υπολογισμό του εσσιανού πίνακα. Στη βασική της ιδέα, ενσωματώνει έναν παράγοντα κανονικοποίησης μ_κ, που προσαρμόζει τη βηματική κατεύθυνση ανάμεσα στη μέθοδο Νεύτωνα και της μέγιστης καθόδου. Ο αλγόριθμος περιλαμβάνει την επίλυση του συστήματος $(J^T J + \mu I)\Delta x = -J^T r$, όπου J είναι ο πίνακας Jacobian των παραγώγων της συνάρτησης, r είναι το διάνυσμα των υπολοίπων (δηλαδή της διαφοράς μεταξύ των δεδομένων και της εκτιμώμενης τιμής), και μ είναι ένας θετικός παράγοντας που ελέγχει την επιρροή της κανονικοποίησης. Η διαδικασία λειτουργεί ως εξής: όταν μ είναι μεγάλος, το $J^T J + \mu I$ κυριαρχείται από τον όρο μI, κάτι που καθιστά τον αλγόριθμο παρόμοιο με την μέθοδο μέγιστης καθόδου, διασφαλίζοντας σταθερότητα όταν το σημείο εκκίνησης απέχει πολύ από τη λύση. Όταν μ είναι μικρός, το $J^T J$ κυριαρχεί, επιτρέποντας στον αλγόριθμο να συμπεριφέρεται όπως η μέθοδος του Newton, που έχει ταχύτερη σύγκλιση κοντά στο βέλτιστο σημείο.

- Αλγόριθμος

Ο αλγόριθμος της μεθόδου Levenberg-Marquardt ξεκινά με την επιλογή ενός αρχικού σημείου x_0 και ενός αρχικού παράγοντα τακτικοποίησης $\mu > 0$. Η διαδικασία είναι επαναληπτική και σε κάθε βήμα ο υπολογισμός βασίζεται στην επίλυση του προσαρμοσμένου γραμμικού συστήματος $(J^T J + \mu I)\Delta x = -J^T r$ που αναφέρθηκε παραπάνω. Το νέο σημείο υπολογίζεται ως $x_{k+1} = x_k + \Delta x$. Για τον υπολογισμό του μ, αποδεικνύεται ότι χρειάζεται να είναι μεγαλύτερος της απόλυτης τιμής της μεγαλύτερης ιδιοτιμης του $\nabla^2 f(x)$, μ_k. Επομένως ισχύει $\mu = \mu_k + \lambda$, όπου λ σταθερά.

- Μετρήσεις

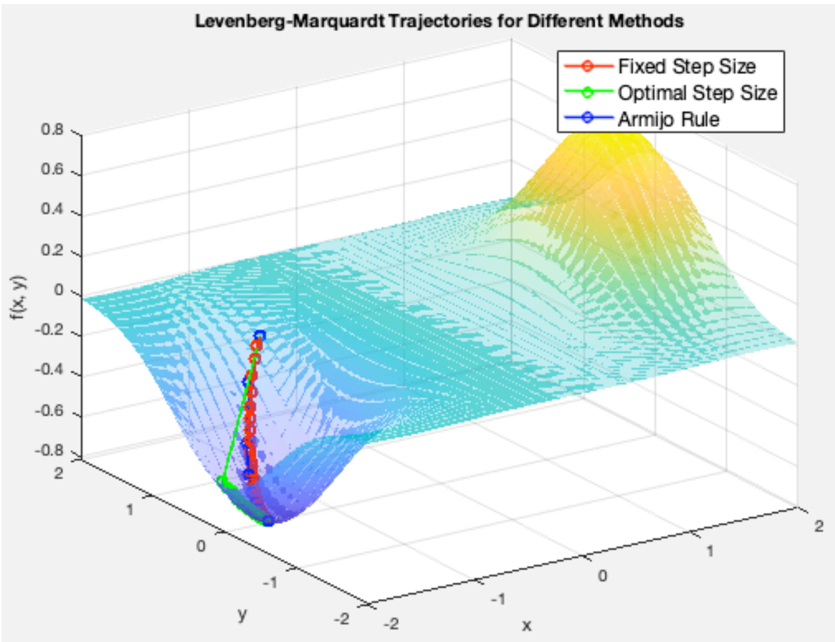
Για το σημείο $(0,0)$:



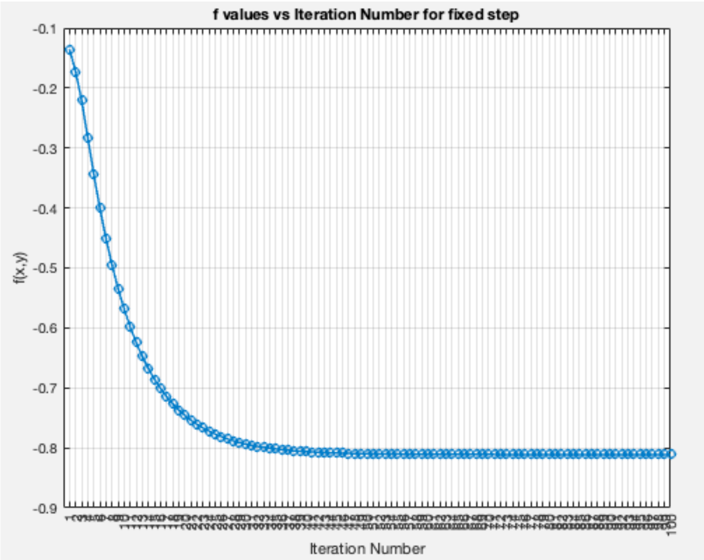
Γραφική Παράσταση της σύγκλισης της μεθόδου με
αρχικό σημείο το $(0,0)$

Είναι προφανές ότι η μέθοδος δεν συγκλίνει για κανένα υπολογισμό του βήματος όταν το αρχικό σημείο είναι το $(0,0)$ διότι η παραγωγός είναι μηδενική στη γύρω περιοχή.

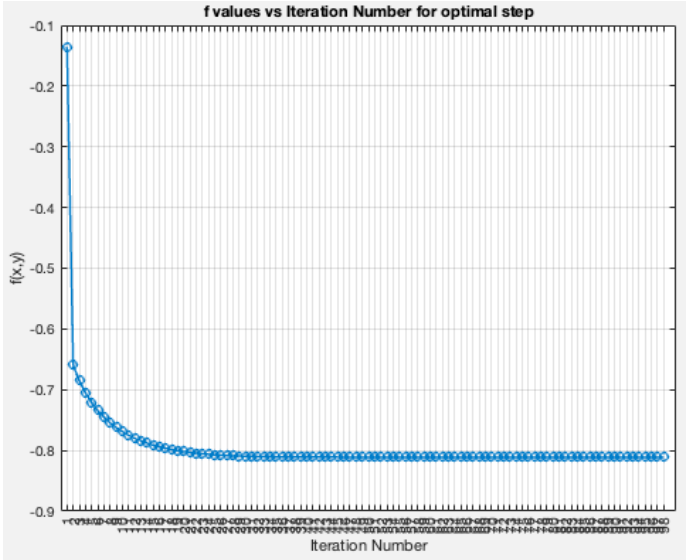
Για το σημείο $(-1,1)$:



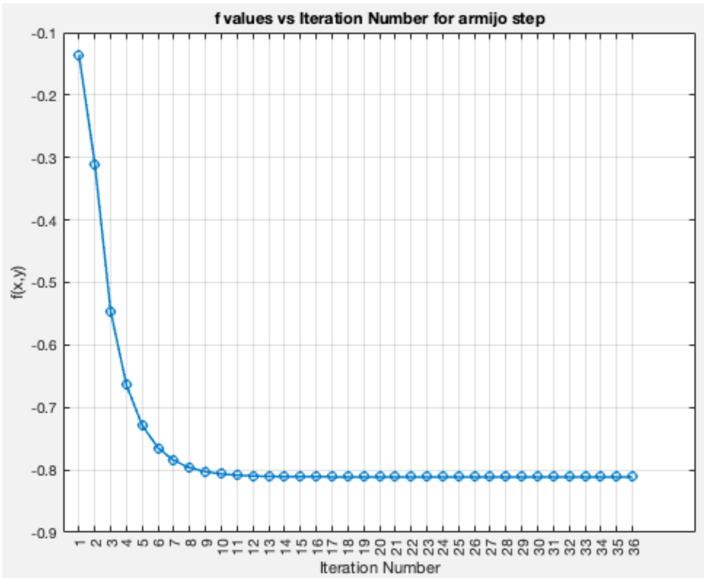
Γραφική Παράσταση της σύγκλισης της μεθόδου με
αρχικό σημείο το $(-1,1)$



Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για σταθερό βήμα

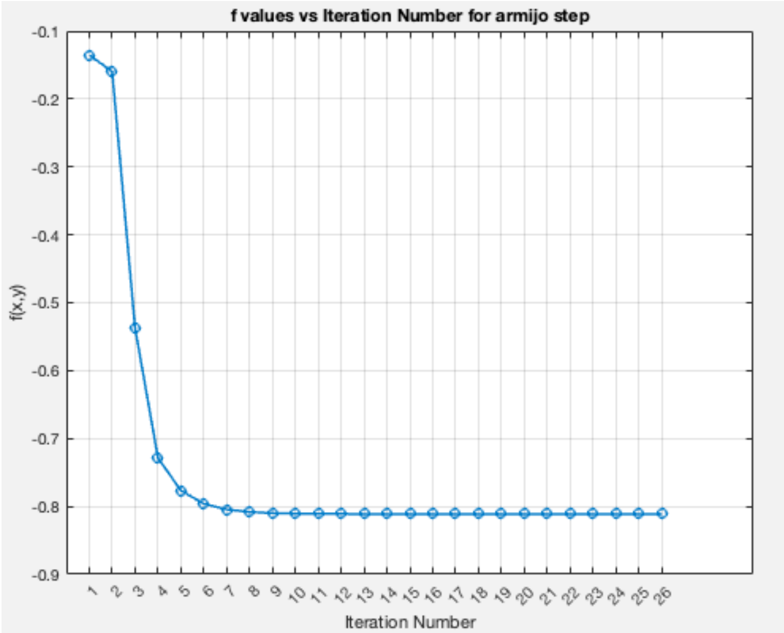


Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βέλτιστο βήμα



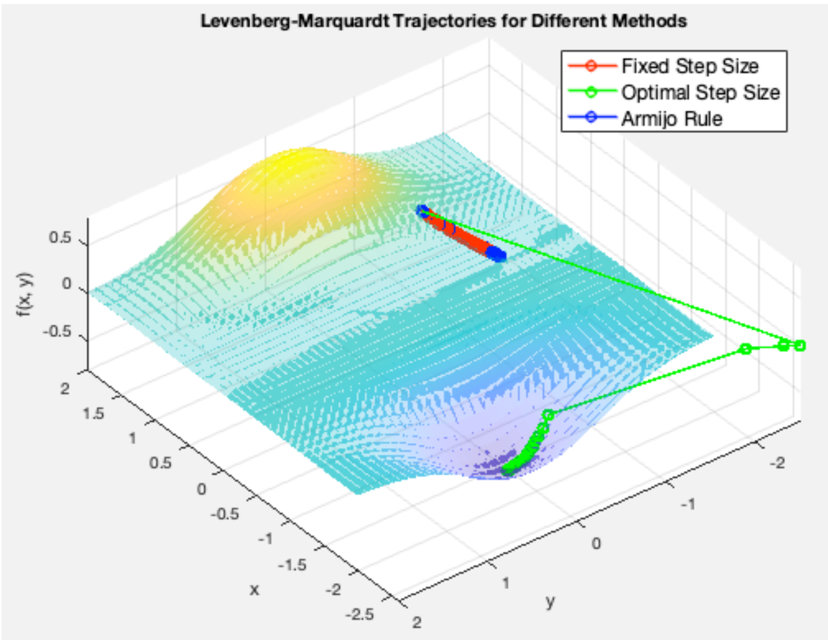
Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βήμα με κανόνα Armijo

Οι παραπάνω μετρήσεις έγιναν με $\gamma_{fixed} = 0.5$, $\gamma_{armijo,0} = 1$, $\varepsilon = 10^{-4}$, $\alpha = 0.2$, $\beta = 0.01$ και 100 μέγιστες επαναλήψεις. Το μ υπολογίστηκε ως $\mu_k + 1.7$. Για σταθερά μικρότερη του 1.7 ο αλγόριθμος με το βήμα υπολογισμένο από την ελαχιστοποίηση της $f(x_k + \gamma_k d_k)$ δεν σύγκλινε. Προτιμάται ο καθορισμός του βήματος με τον κανόνα Armijo, καθώς σύγκλινε στο σημείο μετά από 36 επαναλήψεις, ενώ οι άλλοι δυο χρειάστηκαν τις ολόκληρες 100. Για τιμές $\mu \to \mu_k$ ο καθορισμός του βήματος με τον κανόνα Armijo γίνεται ακόμη πιο αποτελεσματικός (ανεξαρτήτως της σταθεράς $\gamma_{armijo,0}$), ενώ οι άλλοι δυο μέθοδοι δεν συγκλίνουν.

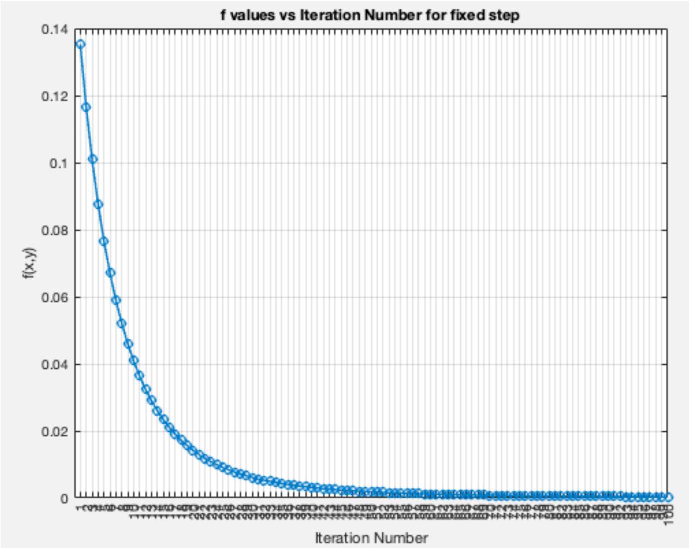


Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βήμα με κανόνα Armijo για $\mu \rightarrow \mu_k$

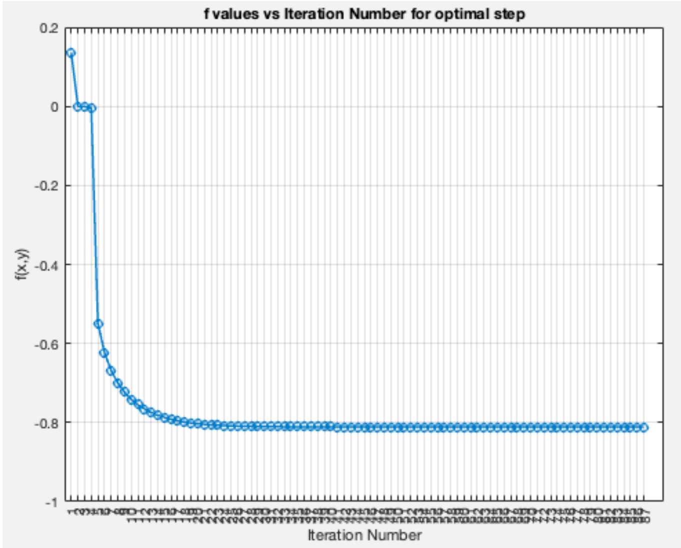
Για το σημείο $(1, -1)$:



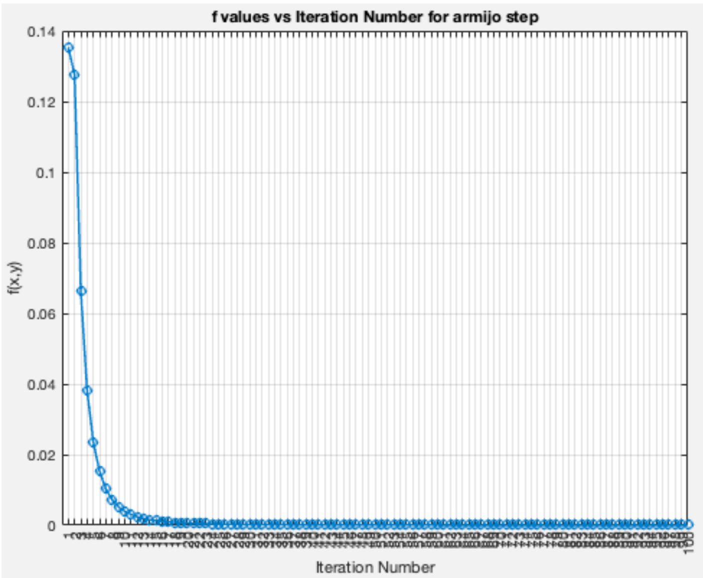
Γραφική Παράσταση της σύγκλισης της μεθόδου με αρχικό σημείο το $(1, -1)$



Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για σταθερό βήμα



Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βέλτιστο βήμα



Διάγραμμα τιμών της συνάρτησης f συναρτήσει του αριθμού επαναλήψεων για βήμα με κανόνα Armijo

Οι παραπάνω μετρήσεις έγιναν με $\gamma_{fixed} = 0.5$, $\gamma_{armijo,0} = 1$, $\varepsilon = 10^{-4}$, $\alpha = 0.2$, $\beta = 0.01$ και 100 μέγιστες επαναλήψεις. Το μ υπολογίστηκε ως $\mu_k + 1$. Παρατηρείται ότι μόνο η εύρεση του βήματος με την ελαχιστοποίηση της $f(x_k + \gamma_k d_k)$ συγκλίνει, ενώ οι άλλες δυο τεχνικές δεν μπορούν να ξεφύγουν από το επίπεδο σημείο. Για τιμές $\mu \to \mu_k$ και $\mu \to \infty$ καμία μέθοδος δεν συγκλίνει. Ο αλγόριθμος δεν συγκλίνει ποτέ όταν το βήμα είναι σταθερό ή υπολογίζεται με τον κανόνα Armijo.

Καταληκτικά, ο αλγόριθμος αυτός δεν συγκλίνει πάντοτε. Το αρχικό σημείο συμβάλει στο αποτέλεσμα, καθώς ξεκινώντας από μια επίπεδη περιοχή η παράγωγος είναι μηδενική και ο αλγόριθμος τερματίζει. Επίσης, αν το αρχικό σημείο είναι κοντά στο ελάχιστο, τότε συγκλίνουν και οι 3 μέθοδοι για κατάλληλη κανονικοποίηση του μη θετικού εσσιανού πίνακα. Από την άλλη, όταν το αρχικό σημείο είναι μακριά από το σημείο αναζήτησης, μόνο η μέθοδος ελαχιστοποίησης της

$f(x_k + \gamma_k d_k)$ συγκλίνει για κατάλληλο θετικό παράγοντα μ. Καταληκτικά, ο αλγόριθμος συγκλίνει μόνο όταν το αρχικό σημείο δεν είναι σε επίπεδη περιοχή και έχει γίνει σωστή κανονικοποίηση του εσσιανού πίνακα με την μέθοδο αυτή. Αν το αρχικό σημείο είναι κοντά στο ελάχιστο, τότε συγκλίνουν και οι άλλοι μέθοδοι.