

Τεχνικές Βελτιστοποίησης - Εργασία 2

Ρουσομάνης Γεώργιος (ΑΕΜ: 10703)

Νοέμβριος 2024

Εισαγωγή

Ο στόχος της παρούσας εργασίας είναι η ελαχιστοποίηση μιας συνάρτησης πολλών μεταβλητών χωρίς περιορισμούς με τη χρήση παραγώγων και η μελέτη της αποτελεσματικότητας διαφορετικών αλγορίθμων βελτιστοποίησης. Ειδικότερα, η αντικειμενική συνάρτηση που θα εξεταστεί είναι:

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

Οι μέθοδοι που θα χρησιμοποιηθούν για την αναζήτηση του ελαχίστου περιλαμβάνουν:

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

Η κατεύθυνση καθόδου d_k καθορίζεται διαφορετικά σε κάθε μέθοδο, ενώ το βήμα γ_k μπορεί να επιλεγεί είτε σταθερό, είτε με ελαχιστοποίηση της $f(x_k + \gamma_k d_k)$, είτε βάσει του κανόνα Armijo. Σε κάθε επανάληψη, το νέο σημείο υπολογίζεται ως $x_{k+1} = x_k + \gamma_k d_k$ ενώ οι αλγόριθμοι τερματίζουν όταν $|\nabla(x_k)| < \epsilon$ όπου $\epsilon > 0$.

Στόχος είναι να μελετηθούν και να συγκριθούν οι παραπάνω αλγόριθμοι, τόσο ως προς την αποδοτικότητά τους όσο και ως προς τον αριθμό επαναλήψεων που απαιτούνται για την σύγκλισή τους. Παράλληλα, θα εξεταστούν η επίδραση του σημείου εκκίνησης (x_0, y_0) καθώς και η επιλογή του βήματος γ_k στη σύγκλιση και την αποδοτικότητα του κάθε αλγορίθμου.

Τέλος, θα διερευνηθούν οι περιπτώσεις όπου οι μέθοδοι δεν καταλήγουν στο σωστό αποτέλεσμα, προσδιορίζοντας πιθανούς λόγους όπως εγκλωβισμό σε τοπικά άκρα.

Μέθοδος Μέγιστης Καθόδου

Σε αυτή τη μέθοδο, το διάνυσμα κατεύθυνσης d_k επιλέγεται ως $d_k = -\frac{\nabla f(x_k)}{|\nabla f(x_k)|}$. Η χρήση του μοναδιαίου διανύσματος για το d_k εξασφαλίζει ότι το διάστημα αναζήτησης κατά την ελαχιστοποίηση της $f(x_k + \gamma_k d_k)$ εκφράζεται σε φυσικές μονάδες, διατηρώντας έτσι τη συμβατότητα των υπολογισμών με το πρόβλημα βελτιστοποίησης. Ένα επιπλέον πλεονέκτημα της κανονικοποίησης του d_k είναι ότι, σε σημεία όπου η κλίση είναι πολύ μικρή, το διάστημα αναζήτησης και το βήμα γ_k παραμένουν μικρά, οδηγώντας σε μικρότερα βήματα και συνεπώς σε περισσότερες επαναλήψεις. Ενώ η μέθοδος μέγιστης καθόδου μας παρέχει ισχυρή εγγύηση για την σύγκλισή της, δεν είναι υπολογιστικά αποτελεσματική όπως θα διαπιστώσουμε στη συνέχεια.

Σταθερό βήμα

Στο Σχήμα 1 παρατηρούμε τα διαδοχικά σημεία που υπολογίζει ο αλγόριθμος σε κάθε επανάληψη για $\gamma_k = 0.001$. Είναι σημαντικό να τονίσουμε ότι στην περίπτωση του σταθερού βήματος το γ_k πρέπει να είναι αρκούντως μικρό διαφορετικά μπορεί να υπάρξει αδυναμία σύγκλισης ή ταλαντώσεις.

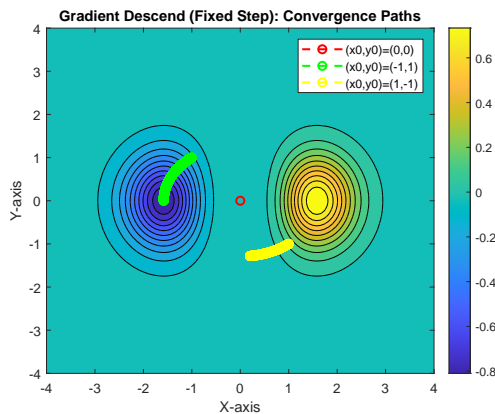
Προφανώς για το αρχικό σημείο $(0,0)$ ισχύει $\nabla f = 0$, επομένως ο αλγόριθμος τερματίζει αμέσως. Για το $(-1,1)$ έχουμε επιτυχή σύγκλιση καθώς βρισκόμαστε αρκούντως κοντά στο ολικό ελάχιστο. Ωστόσο για το $(1,-1)$ έχουμε αδυναμία σύγκλισης λόγω της επίπεδης επιφάνειας που εισέρχεται η μέθοδος όπου $\nabla f \approx 0$. Όπως θα δούμε στην συνέχεια το πρόβλημα αυτό αίρεται με την ελαχιστοποίηση της $f(x_k + \gamma_k d_k)$ ή με τον κανόνα Armijo.

Επίσης παρατηρούμε ότι η τροχιά που διαγράφουν τα διαδοχικά σημεία που υπολογίζει ο αλγόριθμος είναι κάθετη στις ισοδυναμικές επιφάνειες. Αυτό συμβαίνει επειδή το d_k έχει την κατεύθυνση της μέγιστης μείωσης της συνάρτησης, και σε κάθε σημείο πάνω στην ισοδυναμική επιφάνεια η συνάρτηση παραμένει σταθερή, άρα η παράγωγος της συνάρτησης σε κάθε κατεύθυνση πάνω στην επιφάνεια είναι μηδέν.

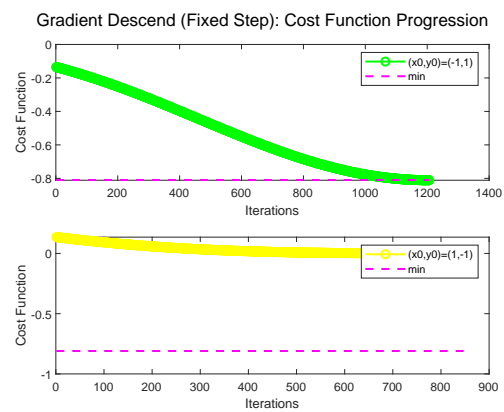
Στο Σχήμα 2 φαίνεται η σύγκλιση της μεθόδου σε κάθε επανάληψη για τα αρχικά σημεία $(-1,1)$, $(1,-1)$.

Βήμα με ελαχιστοποίηση

Σε αυτή την περίπτωση θα υπολογίσουμε το γ_k που ελαχιστοποιεί την $f(x_k + \gamma_k d_k)$ σε ένα προκαθορισμένο διάστημα αναζήτησης d . Για τον σκοπό αυτό θα χρησιμοποιήσουμε την μέθοδο Fibonacci. Πρόφανώς για να συγκλίνει η μέθοδος Fibonacci στο ελάχιστο θα πρέπει η συνάρτηση να είναι αυστηρά σχεδόν κυρτή στο d . Επίσης για το αρχικό σημείο $(1,-1)$, το d πρέπει να

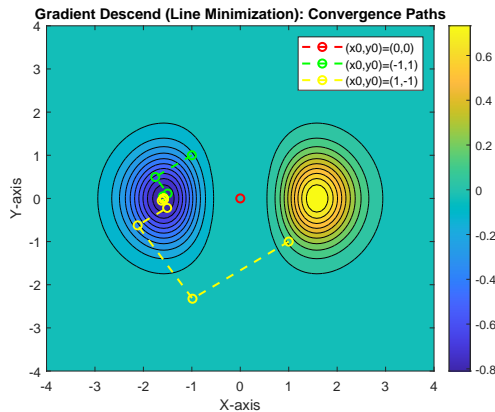


Σχήμα 1: Διαδοχικά σημεία υπολογισμού της μεθόδου Μέγιστης Καθόδου για σταθερό βήμα $\gamma_k = 0.001$

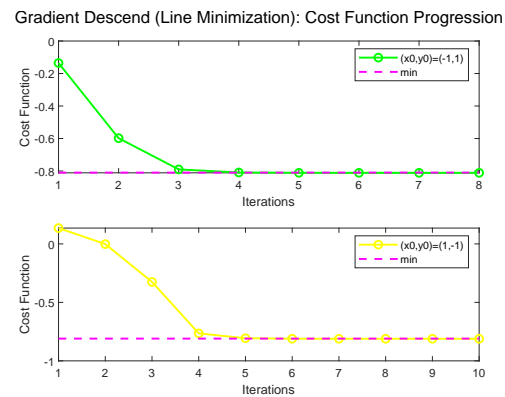


Σχήμα 2: Σύγκλιση της μεθόδου Μέγιστης Καθόδου για σταθερό βήμα $\gamma_k = 0.001$

είναι αρκούντως μεγάλο ώστε να ξεπεράσουμε την περιοχή όπου $\nabla f \approx 0$. Στο Σχήμα 3 παρατηρούμε ότι η μέθοδος συγκλίνει για τα αρχικά σημεία $(-1, 1)$, $(1, -1)$. Επίσης βλέπουμε ότι η πορεία προς το ελάχιστο είναι μία τεθλασμένη γραμμή, με τα διαδοχικά ευθύγραμμα τμήματα να είναι κάθετα μεταξύ τους, όπως αναμέναμε άλλωστε από την θεωρία. Από το Σχήμα 4 φαίνεται πως η σύγκλιση για αρχικό σημείο $(-1, 1)$ επιτεύχθηκε με πολύ μικρότερο αριθμό επαναλήψεων από ότι στην περίπτωση του σταθερού βήματος.



Σχήμα 3: Διαδοχικά σημεία υπολογισμού της μεθόδου Μέγιστης Καθόδου για βήμα μέσω ελαχιστοποίησης



Σχήμα 4: Σύγκλιση της μεθόδου Μέγιστης Καθόδου για βήμα μέσω ελαχιστοποίησης

Βήμα με τον κανόνα Armijo

Στο Σχήμα 5 παρουσιάζονται οι τρεις πρώτες επαναλήψεις της μεθόδου Armijo για τα αρχικά σημεία $(-1, 1)$ και $(1, -1)$, με παραμέτρους $s = 3$, $\beta = 0.5$, και $\alpha = 0.01$. Οι τιμές των α και s επιλέχθηκαν έτσι ώστε, στην πρώτη επανάληψη για το αρχικό σημείο $(1, -1)$, το γ_k να είναι αρκετά μεγάλο για να αποφύγουμε την περιοχή όπου $\nabla f(x_k) \approx 0$.

Η αύξηση του α συνεπάγεται μεγαλύτερη κλίση της ροζ ευθείας (κατά α -πόλυτη τιμή), γεγονός που σημαίνει ότι η απαιτούμενη βελτίωση σε κάθε επανάληψη αυξάνεται. Η πράσινη ευθεία ορίζει τη μέγιστη δυνατή βελτίωση που μπορεί να επιτευχθεί και αντιστοιχεί σε τιμή $\alpha = 1$, ισούται δε με την κλίση της αντικειμενικής συνάρτησης στο υπό εξέταση σημείο x_k .

Τέλος, η παράμετρος β καθορίζει τον ρυθμό με τον οποίο μειώνεται το βήμα σε κάθε επανάληψη, επηρεάζοντας τη σταδιακή προσέγγιση της λύσης.

Από το Σχήμα 7 βλέπουμε ότι η μέθοδος μέγιστης καθόδου με τον κανόνα Armijo απαιτεί περίπου τον ίδιο αριθμό επαναλήψεων με αυτόν της ελαχιστοποίησης συνάρτησης, συγκλίνει δε πολύ πιο γρήγορα από την περίπτωση σταθερού βήματος.

Μέθοδος Newton

Σε αυτή τη μέθοδο, το διάνυσμα κατεύθυνσης d_k επιλέγεται ως

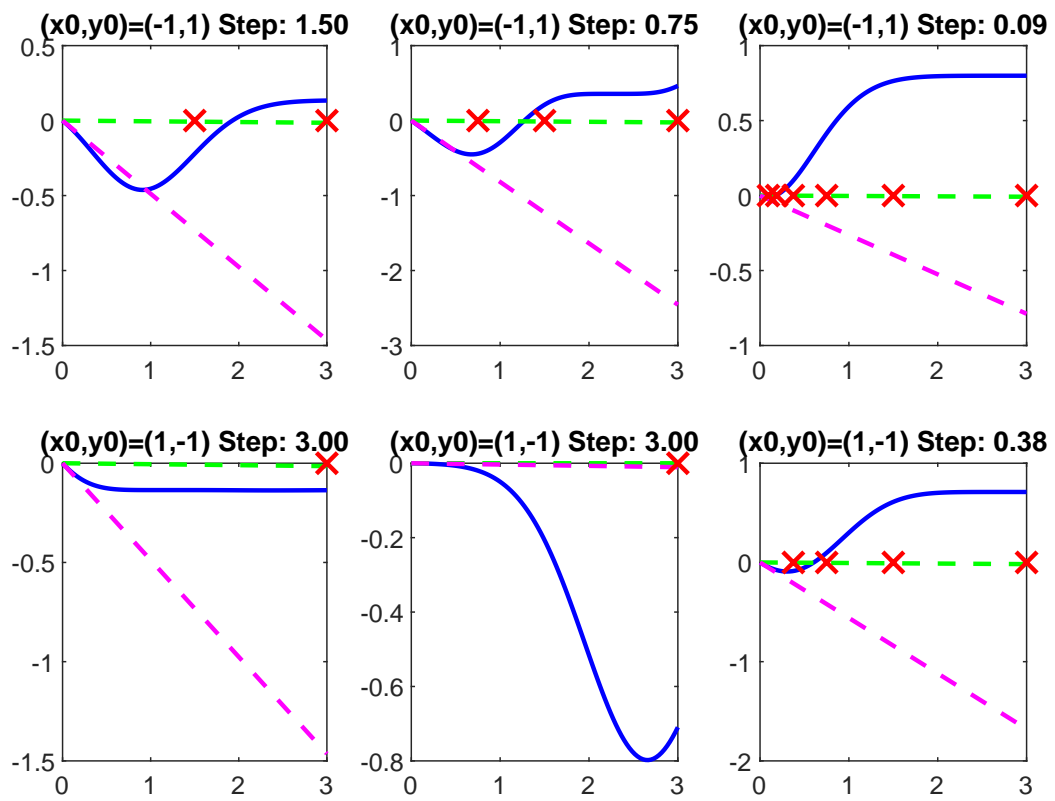
$$d_k = -\frac{[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)}{|[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)|}$$

όπου $\nabla^2 f(x_k)$ ο εσσιανός πίνακας της αντικειμενικής συνάρτησης. Η κανονικοποίηση του d_k γίνεται για τους λόγους που αναφέρθηκαν παραπάνω.

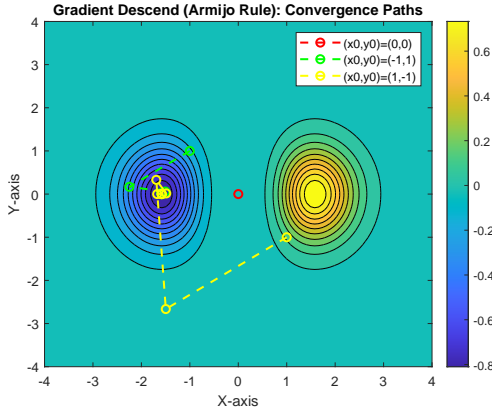
Επειδή η συγκεκριμένη μέθοδος λαμβάνει υπό όψιν την κυρτότητα της συνάρτησης επιτυγχάνει ταχύτερη σύγκλιση σε σχέση με την μέθοδο της μέγιστης καθόδου. Ωστόσο αυτή η επιτάχυνση έρχεται και με μεγαλύτερο υπολογιστικό κόστος λόγω του υπολογισμού του αντίστροφου του εσσιανού πίνακα. Ένα άλλο μειονέκτημα της μεθόδου είναι η απαίτηση ο εσσιανός πίνακας να είναι θετικά ορισμένος ώστε να μπορέσει να συγκλίνει στο ολικό ελάχιστο.

Αυτός είναι και ο λόγος που στην περίπτωσή μας ο αλγόριθμος αποτυγχάνει για τα αρχικά σημεία $(-1, 1)$, $(1, -1)$. Στο Σχήμα 8 έως Σχήμα 13 φαίνεται η απόκλιση του αλγορίθμου για κάθε περίπτωση υπολογισμού του βήματος.

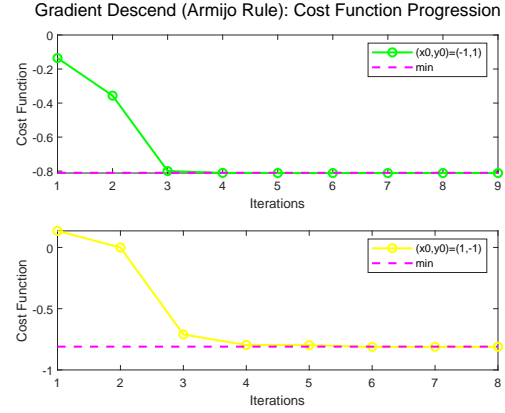
Gradient Descend (Armijo Rule)



Σχήμα 5: Μέθοδος Armijo για τις τρεις πρώτες επαναλήψεις της μεθόδου Μέγιστης Καθόδου



Σχήμα 6: Διαδοχικά σημεία υπολογισμού της μεθόδου Μέγιστης Καθόδου για βήμα μέσω του κανόνα Armijo



Σχήμα 7: Σύγκλιση της μεθόδου Μέγιστης Καθόδου για βήμα μέσω του κανόνα Armijo

Μέθοδος Levenberg-Marquardt

Σε αυτή τη μέθοδο, το διάνυσμα κατεύθυνσης d_k επιλέγεται ως

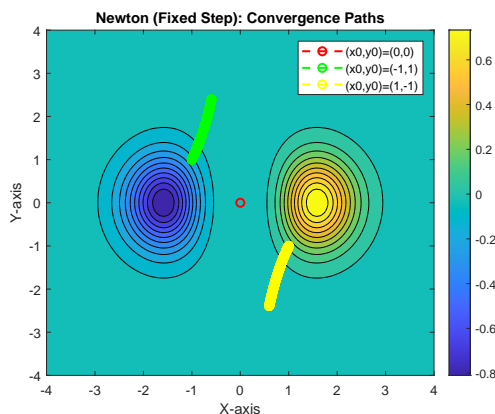
$$d_k = -\frac{[\nabla^2 f(x_k) + \mu_k I]^{-1} \nabla f(x_k)}{||[\nabla^2 f(x_k) + \mu_k I]^{-1} \nabla f(x_k)||}.$$

Η κανονικοποίηση του d_k γίνεται για λόγους που έχουν ήδη αναφερθεί. Αυτή η μέθοδος συνδυάζει την ευρωστία της μεθόδου μέγιστης καθόδου με την ταχύτητα σύγκλισης της μεθόδου Newton.

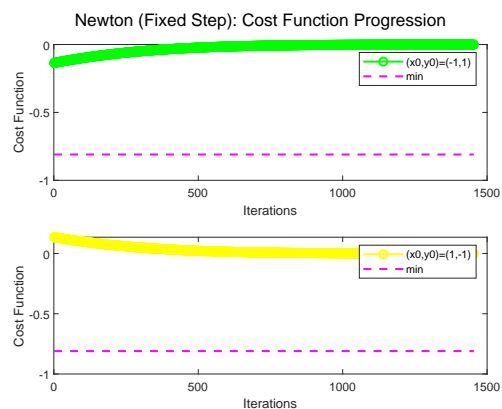
Το πρόβλημα της απαίτησης για θετικά ορισμένο εσσιανό πίνακα στην μέθοδο Newton αίρεται με την κατάλληλη επιλογή του μ_k . Πιο συγκεκριμένα όταν ο $\nabla^2 f(x_k)$ δεν είναι θετικά ορισμένος, υπολογίζουμε το μ_k έτσι ώστε ο πίνακας $\nabla^2 f(x_k) + \mu_k I$ να είναι θετικά ορισμένος. Αποδிகνύεται ότι αυτό ισχύει όταν το μ_k είναι μεγαλύτερο από την μέγιστη κατ' απόλυτη τιμή ιδιοτιμή του $\nabla^2 f(x_k)$.

Το μ_k διαδραματίζει καθοριστικό ρόλο στην συμπεριφορά του αλγορίθμου. Όταν $\mu_k \rightarrow 0$ ο όρος $\nabla^2 f(x_k)$ επικρατεί έναντι του $\mu_k I$ και η μέθοδος έχει παρόμοια συμπεριφορά με την μέθοδο Newton. Αντίστοιχα όταν $\mu_k \rightarrow \infty$ ο όρος $\mu_k I$ επικρατεί έναντι του $\nabla^2 f(x_k)$ και η μέθοδος έχει συμπεριφορά παρόμοια με αυτή της μέγιστης καθόδου.

Στην ανάλυσή μας το μ_k επιλέγεται να είναι ίσο με την μέγιστη κατ' απόλυτη τιμή ιδιοτιμή του $\nabla^2 f(x_k)$ πολλαπλασιασμένη με μία σταθερά ελαφρώς μεγαλύτερη της μονάδας. Με αυτόν τον τρόπο έχουμε εγγυημένη σύγκλιση, ενώ εξασφαλίζουμε τη βέλτιστη δυνατή ταχύτητα σύγκλισης.



Σχήμα 8: Διαδοχικά σημεία υπολογισμού της μεθόδου Newton για σταθερό βήμα



Σχήμα 9: Σύγκλιση της μεθόδου Newton για σταθερό βήμα

Σταθερό βήμα

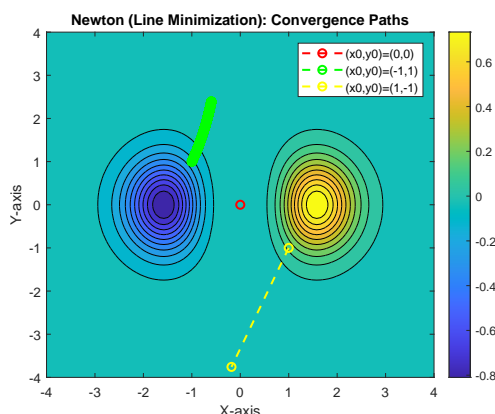
Στο Σχήμα 14 φαίνονται τα διαδοχικά σημεία που υπολογίζει ο αλγόριθμος σε κάθε επανάληψη για $\gamma_k = 0.001$. Παρόμοια με τη μέθοδο μέγιστης καθόδου, ο αλγόριθμος αποτυγχάνει να συγκλίνει όταν το αρχικό σημείο είναι $(1, -1)$.

Αυτό συμβαίνει επειδή, όταν ο αλγόριθμος βρίσκεται μακριά από το ελάχιστο, η συμπεριφορά του πλησιάζει εκείνη της μεθόδου μέγιστης καθόδου. Στη μέθοδο αυτή, η κατεύθυνση της κίνησης εξαρτάται αποκλειστικά από την κλίση της αντικειμενικής συνάρτησης, η οποία είναι περίπου μηδενική στην εν λόγω επιφάνεια, οδηγώντας σε αδυναμία βελτίωσης.

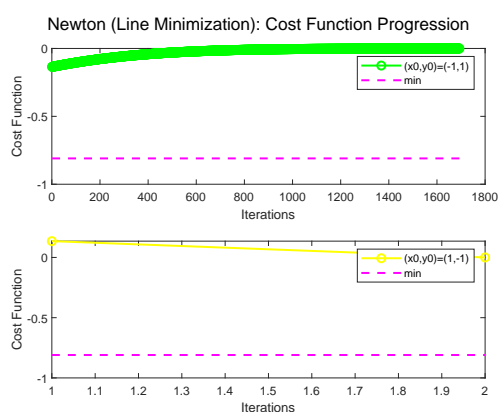
Εάν επιχειρήσουμε να αυξήσουμε το μέγεθος του βήματος ώστε να αποφύγουμε την επίπεδη περιοχή, αυτό οδηγεί σε αδυναμία σύγκλισης για τα δύο αρχικά σημεία λόγω ταλαντώσεων γύρω από το σημείο ισορροπίας. Επιπλέον, παρατηρούμε ότι, μετά από έναν αριθμό επαναλήψεων, η πορεία του σημείου $(-1, 1)$ προς το ελάχιστο παύει να είναι κάθετη στις ισοδυναμικές επιφάνειες.

Αυτό συμβαίνει διότι, καθώς πλησιάζουμε στο ελάχιστο, η συμπεριφορά του αλγορίθμου προσεγγίζει εκείνη της μεθόδου Newton, όπου η κυρτότητα της συνάρτησης καθοδηγεί τη λύση, αντί της κλίσης, επιτρέποντας την εύρεση συντομότερης διαδρομής προς το ελάχιστο.

Στο Σχήμα 15 φαίνεται το κόστος συναρτήσεων των επαναλήψεων. Παρατηρούμε ότι η μέθοδος Levenberg-Marquardt συγκλίνει ταχύτερα από την μέθοδο μέγιστης καθόδου όταν το βήμα είναι σταθερό, για τους λόγους που αναφέρθηκαν.



Σχήμα 10: Διαδοχικά σημεία υπολογισμού της μεθόδου Newton για βήμα μέσω ελαχιστοποίησης



Σχήμα 11: Σύγκλιση της μεθόδου Newton για βήμα μέσω ελαχιστοποίησης

Βήμα με ελαχιστοποίηση

Στο Σχήμα 16 βλέπουμε τα διαδοχικά σημεία που υπολογίζει ο αλγόριθμος Levenberg-Marquardt με βήμα γ_k που να ελαχιστοποιεί την $f(x_k + \gamma_k d_k)$. Το διάστημα αναζήτησης ελαχίστου, όπως και στην μέθοδο μέγιστης καθόδου, είναι $d = 3$ ώστε να αποφύγουμε την επίπεδη περιοχή. Η ελαχιστοποίηση γίνεται και πάλι με την μέθοδο Fibonacci.

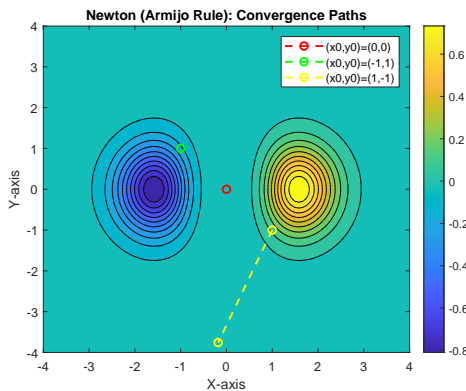
Στο Σχήμα 17 απεικονίζεται η σύγκλιση της μεθόδου σε κάθε επανάληψη για την περίπτωση ελαχιστοποίησης συνάρτησης. Παρατηρούμε ότι η μέθοδος κατέληξε στη λύση μέσα σε μόλις τέσσερις επαναλήψεις για τα αρχικά σημεία $(-1, 1)$ και $(1, -1)$. Σε σύγκριση, η μέθοδος μέγιστης καθόδου απαιτούσε οκτώ επαναλήψεις για το σημείο $(-1, 1)$ και δέκα επαναλήψεις για το σημείο $(1, -1)$, αναδεικνύοντας την ταχύτερη σύγκλιση της παρούσας μεθόδου.

Όπως έχουμε ήδη αναφέρει, καθώς ο αλγόριθμος πλησιάζει στο ελάχιστο, η συμπεριφορά του αρχίζει να προσομοιάζει αυτήν της μεθόδου Newton. Σε αυτό το στάδιο, η πορεία προς τη λύση καθορίζεται κυρίως από την κυρτότητα της συνάρτησης αντί για την κλίση, επιτρέποντας την επιλογή μιας πιο αποτελεσματικής και άμεσης διαδρομής προς το ελάχιστο.

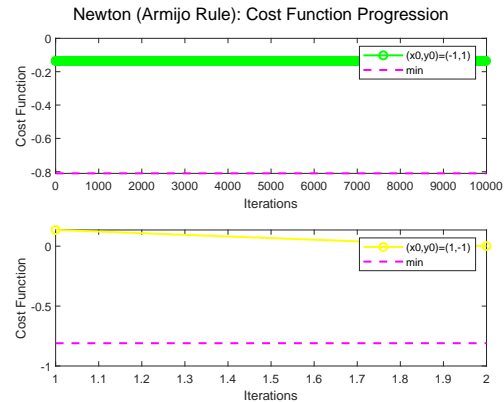
Βήμα με τον κανόνα Armijo

Στο Σχήμα 18 παρουσιάζονται οι τρεις πρώτες επαναλήψεις της μεθόδου Armijo για τα αρχικά σημεία $(-1, 1)$ και $(1, -1)$, με παραμέτρους $s = 3$, $\beta = 0.5$, και $\alpha = 0.01$. Οι τιμές των παραμέτρων επιλέχθηκαν όπως και στην μέθοδο μέγιστης καθόδου για τους λόγους που έχουν είδη αναφερθεί.

Από το Σχήμα 17 και το Σχήμα 20 παρατηρούμε ότι η μέθοδος Levenberg-



Σχήμα 12: Διαδοχικά σημεία υπολογισμού της μεθόδου Newton για βήμα μέσω του κανόνα Armijo



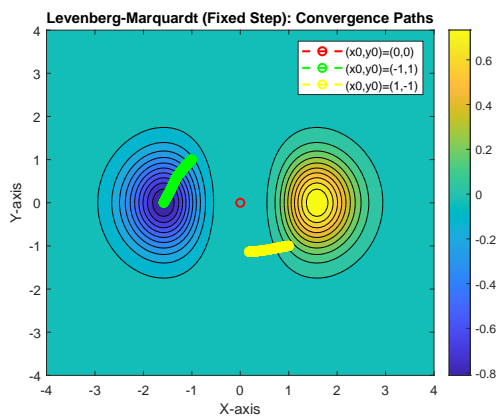
Σχήμα 13: Σύγκλιση της μεθόδου Newton για βήμα μέσω του κανόνα Armijo

Marquardt συγχλίνει πιο αργά όταν χρησιμοποιείται ο κανόνας Armijo, σε σχέση με την περίπτωση της ελαχιστοποίησης της συνάρτησης $f(x_k + \gamma_k d_k)$. Αυτό μπορεί να εξηγηθεί, καθώς στην περίπτωση της ελαχιστοποίησης της συνάρτησης επιτυγχάνεται η μέγιστη δυνατή βελτίωση σε κάθε βήμα.

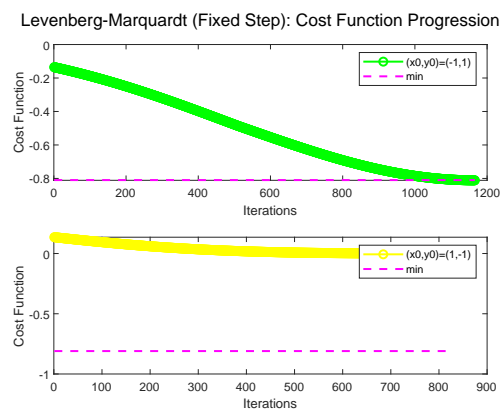
Επίσης, από το Σχήμα 7 και το Σχήμα 20 βλέπουμε ότι για αρχικό σημείο το $(-1, 1)$ οι δύο μέθοδοι απαιτούν τον ίδιο αριθμό επαναλήψεων ενώ για αρχικό σημείο το $(1, -1)$ η μέθοδος Levenberg-Marquardt παρουσιάζει πιο αργή σύγκλιση σε σχέση με την μέθοδο μέγιστης καθόδου.

Συμπέρασμα

Συμπερασματικά, η επιλογή του σημείου εκκίνησης και του βήματος γ_k επηρεάζει σημαντικά τη συμπεριφορά των μεθόδων βελτιστοποίησης. Στη μέθοδο μέγιστης καθόδου, ένα κακό σημείο εκκίνησης μπορεί να οδηγήσει σε αργή σύγκλιση, ιδίως αν βρίσκεται σε επίπεδες περιοχές. Επιπλέον, η χρήση σταθερού γ_k ενδέχεται να μην προσαρμόζεται στις αλλαγές της γεωμετρίας του προβλήματος, καθυστερώντας τη σύγκλιση, ενώ η προσαρμοστική επιλογή γ_k , π.χ. μέσω ελαχιστοποίησης της κατεύθυνσης ή κανόνα Armijo, επιταχύνει τη διαδικασία. Στη μέθοδο Newton, το σημείο εκκίνησης επηρεάζει την εγγύτητα προς τη σύγκλιση, ενώ οι κακές αρχικές επιλογές μπορεί να οδηγήσουν σε μη θετικούς ορισμένους εσσιανούς ή σε αποκλίσεις. Η μέθοδος Levenberg-Marquardt ξεπερνά τέτοια προβλήματα με την προσθήκη όρου απόσβεσης, καθιστώντας τη πιο ανεκτική σε απομακρυσμένα σημεία εκκίνησης. Συνολικά, οι μέθοδοι που ρυθμίζουν δυναμικά το γ_k αποδείχθηκαν πιο αποτελεσματικές, εξισορροπώντας την ταχύτητα και την ακρίβεια της σύγκλισης, ιδιαίτερα όταν

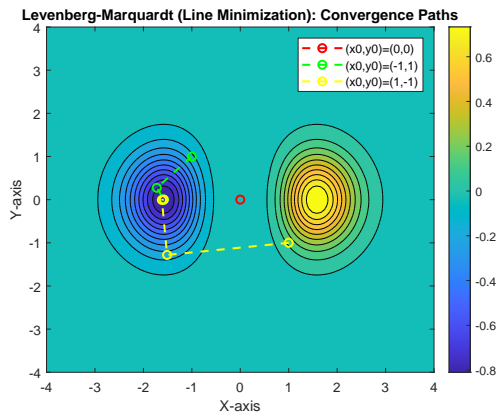


Σχήμα 14: Διαδοχικά σημεία υπολογισμού της μεθόδου Levenberg-Marquardt για σταθερό βήμα $\gamma_k = 0.001$

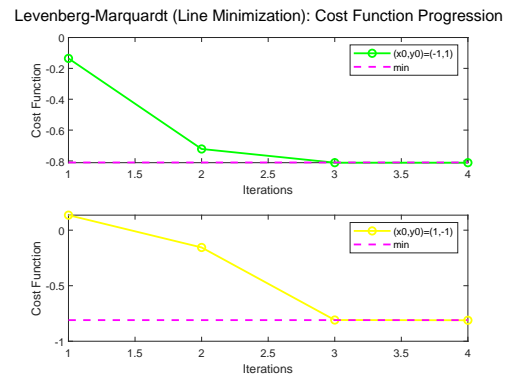


Σχήμα 15: Σύγκλιση της μεθόδου Levenberg-Marquardt για σταθερό βήμα $\gamma_k = 0.001$

το σημείο εκκίνησης βρίσκεται μακριά από το ολικό ελάχιστο.

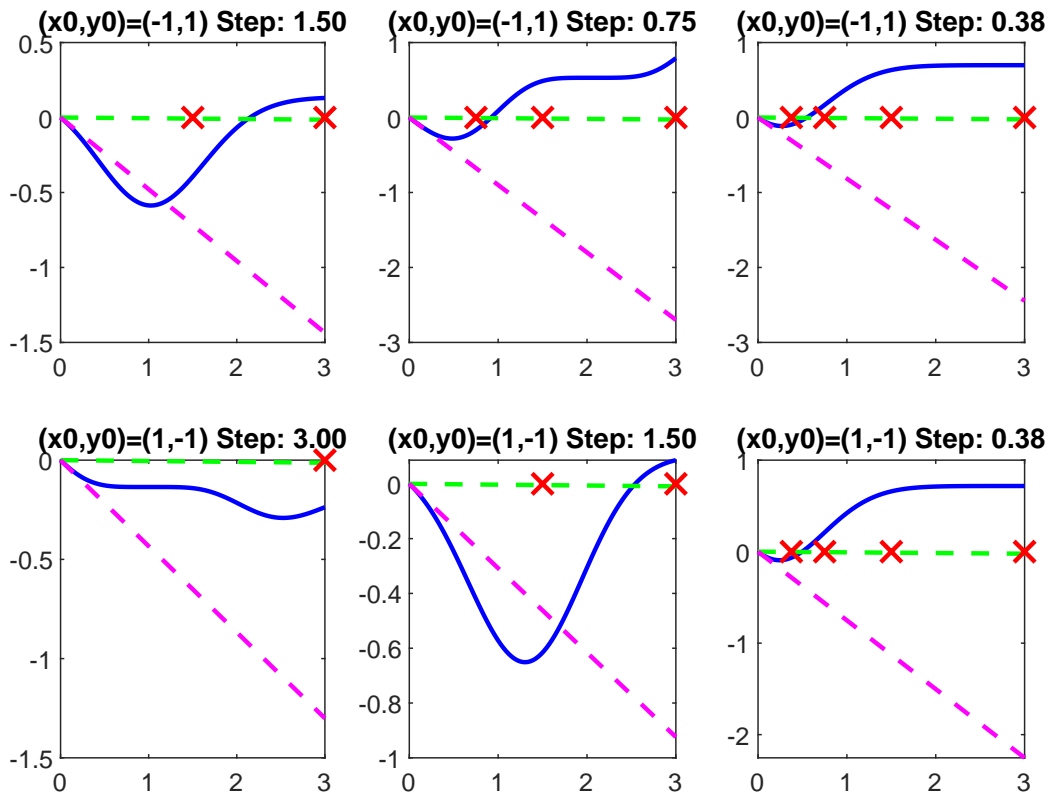


Σχήμα 16: Διαδοχικά σημεία υπολογισμού της μεθόδου Levenberg-Marquardt για βήμα μέσω ελαχιστοποίησης

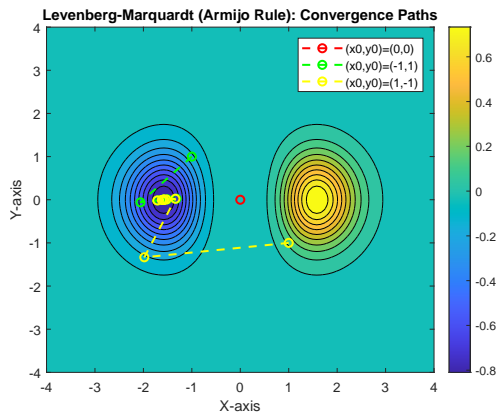


Σχήμα 17: Σύγκλιση της μεθόδου Levenberg-Marquardt για βήμα μέσω ελαχιστοποίησης

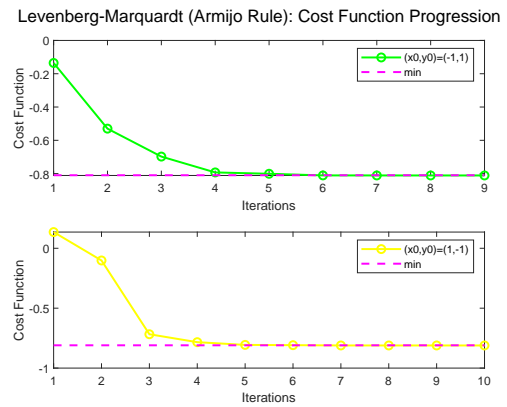
Levenberg-Marquardt (Armijo Rule)



Σχήμα 18: Μέθοδος Armijo για τις τρεις πρώτες επαναλήψεις της μεθόδου Levenberg-Marquardt



Σχήμα 19: Διαδοχικά σημεία υπολογισμού της μεθόδου Levenberg-Marquardt για βήμα μέσω του κανόνα Armijo



Σχήμα 20: Σύγκλιση της μεθόδου Levenberg-Marquardt για βήμα μέσω του κανόνα Armijo