Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρονικής και Υπολογιστών

2η Εργαστηριακή Άσκηση

Ελαχιστοποίηση με χρήση παραγώγων

Τσαμουρίδης Αναστάσιος Αθανάσιος

Δεκέμβριος 2022

Αυτή η σελίδα αφέθηκε σκοπίμως κενή

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή - Θέμα 1

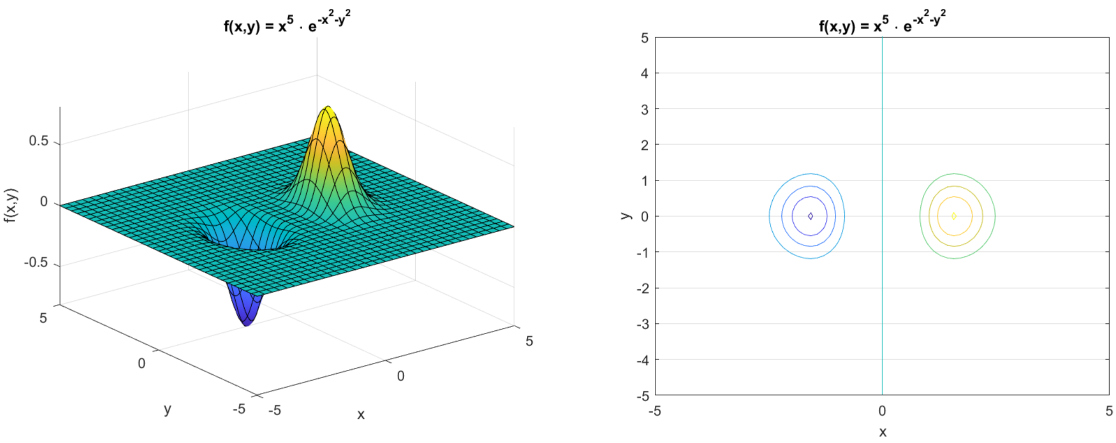
Στην παρούσα αναφορά γίνεται η παρουσίαση των μέθοδων και των αποτελεσμάτων κάποιων βασικών αλγορίθμων, οι οποίοι χρησιμοποιούνται σε προβλήματα ελαχιστοποίησης συναρτήσεων πολλών μεταβλητών, με χρήση παραγώγων και χωρίς την ύπαρξη περιορισμών. Συγκεκριμένα, οι αλγόριθμοι που παρουσιάζονται βασίζονται στις παρακάτω μεθόδους:

- Μέθοδος Μέγιστης Καθόδου - Steepest Descent
- Μέθοδος Newton
- Μέθοδος Levenberg - Marquardt ή Τροποποιημένη Μέθοδος Newton

Για την ανάλυση των αλγορίθμων έγινε η υλοποίηση τους σε MATLAB για την εφαρμογή τους σε συνάρτηση δύο μεταβλητών. Η ανάλυση που ακουλουθεί έγινε επί της συναρτήσεως:

$$f(x, y) = x^5 \cdot e^{-x^2-y^2} \tag{1.1}$$

Στο Σχήμα 1.1 μπορούμε να δούμε την γραφική παράσταση της συνάρτησης και τις ισοβαρείς της καμπύλες.



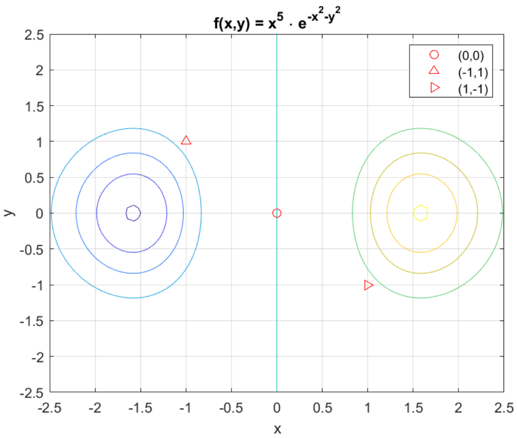
Σχήμα 1.1: Η αντικειμενική συνάρτηση $f(x, y)$.

Οι μέθοδοι που αναφέρθηκαν υλοποιούν την ιδέα της επαναληπτικής καθόδου, δηλαδή ξεκινώντας από ένα σημείο $x_1 = (X_1, Y_1)$ μέσω του αναδρομικού τύπου $x_{k+1} = x_k + \gamma_k d_k$, οδηγείται σε κάποιο x_k, το οποίο πλησιάζει στο πραγματικό ελάχιστο της συνάρτησης. Παραπάνω, ως k ορίζεται ο αριθμός επανάληψης, ως $\gamma_k > 0$ το βήμα που ικανοποιεί την σχέση $f(x_k + 1) < f(x_k)$ και ως d_k το διάνυσμα κατεύθυνσης που ικανοποιεί την συνθήκη $\nabla f^T(x_k)d_k < 0$. Η ειδοποιός διαφορά μεταξύ των μεθόδων αφορά την επιλογή του διανύσματος d_k.

Στην ανάλυση που ακολουθεί για την εφαρμογή των αλγορίθμων επιλέχτηκαν ως αρχικά σημεία τα:

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

Τα σημεία αυτά σημειώνονται και στο Σχήμα 1.2



Σχήμα 1.2: Τα σημεία εκκίνησης που επιλέχτηκαν για τους αλγόριθμους.

Το βήμα στην επανάληψη k επιλέγεται σε κάθε αλγόριθμο ως:

- γ_k σταθερό
- γ_k τέτοιο, ώστε να ελαχιστοποιεί την $\phi(\gamma_k) = f(x_k + \gamma_k d_k)$
- γ_k σύμφωνα με τον κανόνα του Armijo, οποίος περιγράφεται στην ενότητα 7.1

Ο τερματισμός των αλγορίθμων γίνεται όταν το μέτρο της κλίσης της συνάρτησης γίνει μικρότερο από ένα σφάλμα ϵ. Αν η ακρίβεια δεν επιτευχθεί, οι αλγόριθμοι τερματίζουν μετά από έναν μέγιστο αριθμό επαναλήψεων. Στον κώδικα που υλοποιήθηκε, ο μέγιστος αριθμός επαναλήψεων ορίστηκε ως 1000.

Κεφάλαιο 2

Μαθηματική Ανάλυση

2.1 Ανάλυση της Αντικειμενικής Συνάρτησης

Υπενθυμίζεται ότι η συνάρτηση που χρησιμοποιείται είναι η:

$$f(x, y) = x^5 \cdot e^{-x^2-y^2} \tag{2.1}$$

Η κλίση της $f(x, y)$ και ο Εσσιανός πίνακάς της υπολογίζονται ως:

$$\nabla f(x, y) = \begin{bmatrix} -x^4 \cdot e^{-x^2-y^2} \cdot (2x^2 - 5) \\ -2x^5 y \cdot e^{-x^2-y^2} \end{bmatrix}$$

$$\nabla^2 f(x, y) = \begin{bmatrix} 2x^3 \cdot e^{-x^2-y^2} \cdot (2x^4 - 11x^2 + 10) & 2x^4 y \cdot e^{-x^2-y^2} \cdot (2x^2 - 5) \\ 2x^4 y \cdot e^{-x^2-y^2} \cdot (2x^2 - 5) & 2x^5 \cdot e^{-x^2-y^2} \cdot (2y^2 - 1) \end{bmatrix}$$

Για την εύρεση των κρίσιμων σημείων παίρνουμε την κλίση της συνάρτησης ίση με το $\vec{0}$, οπότε προκύπτει:

$$\nabla f(x, y) = \vec{0} \Rightarrow \begin{cases} A = (0, k), k \in R \\ B = (\sqrt{\frac{5}{2}}, 0) \\ C = (-\sqrt{\frac{5}{2}}, 0) \end{cases}$$

Ο Εσσιανός πίνακας σε αυτά τα σημεία δίνει:

- Για το σημείο B: $H_B = \begin{bmatrix} -\frac{25 \cdot \sqrt{10} e^{-5/2}}{2} & 0 \\ 0 & -\frac{25 \cdot \sqrt{10} e^{-5/2}}{4} \end{bmatrix}$

  με τον H_B να έχει ορίζουσα διάφορη του μηδενός και αρνητικές ιδιοτιμές. Άρα το σημείο B είναι σημείο τοπικού μεγίστου, με $f(B) \approx 0.81$.

- Για το σημείο C: $H_C = \begin{bmatrix} \frac{25 \cdot \sqrt{10} e^{-5/2}}{2} & 0 \\ 0 & \frac{25 \cdot \sqrt{10} e^{-5/2}}{4} \end{bmatrix}$

  με τον H_C να έχει ορίζουσα διάφορη του μηδενός και θετικές ιδιοτιμές. Άρα το σημείο C είναι σημείο τοπικού ελαχίστου, με $f(C) \approx -0.81$.

- Για το σύνολο των σημείων A προκύπτει: $H_A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ και άρα δεν μπορούμε να εξάγουμε άμεσα συμπέρασμα για το σύνολο A. Τα σημεία αυτά μπορούν να αποτελούν σύνολο τοπικού μεγίστου ή ελαχίστου ή σαγματικά σημεία.

Η f στο σύνολο A παίρνει την τιμή 0, ενώ στο C έχουμε $f(C) \approx -0.81$, άρα το ολικό ελάχιστο βρίσκεται στο σημείο C, ανεξάρτητα από το αν τα σημεία A είναι τοπικά ελάχιστα ή όχι. Συνεπώς, οι αλγόριθμοι είναι επιθυμητό να επιστρέφουν ως αποτέλεσμα το σημείο C.

Ειδικά για το σημείο $A_1 = (0, 0)$, το οποίο επιλέχτηκε ως ένα από τα σημεία εκκίνησης των αλγορίθμων, έχουμε, όπως δείξαμε και πριν, τα εξής:

$f(0, 0) = 0$, $\nabla f(0, 0) = \vec{0}$, $H_{A1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

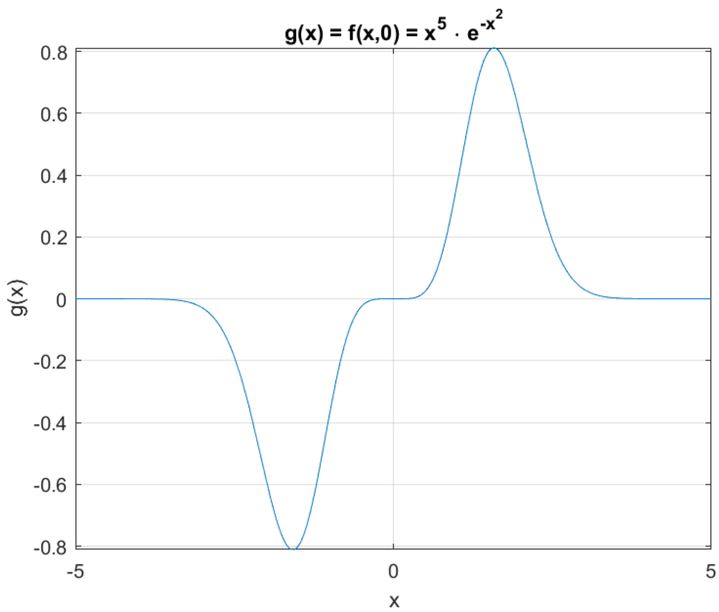
Δηλαδή είναι κρίσιμο σημείο με Εσσιανό πίνακα που δεν μας επιτρέπει την άμεση εξαγωγή συμπερασμάτων για την φύση του σημείου.

Για τον λόγο αυτό μελετούμε μια μονοδιάστατη καμπύλη της f, ώστε να βγάλουμε συμπεράσματα για το σημείο A_1. Ορίζουμε την συνάρτηση $g(x) = f(x, 0) = x^5 \cdot e^{-x^2}$, οπότε έχουμε:

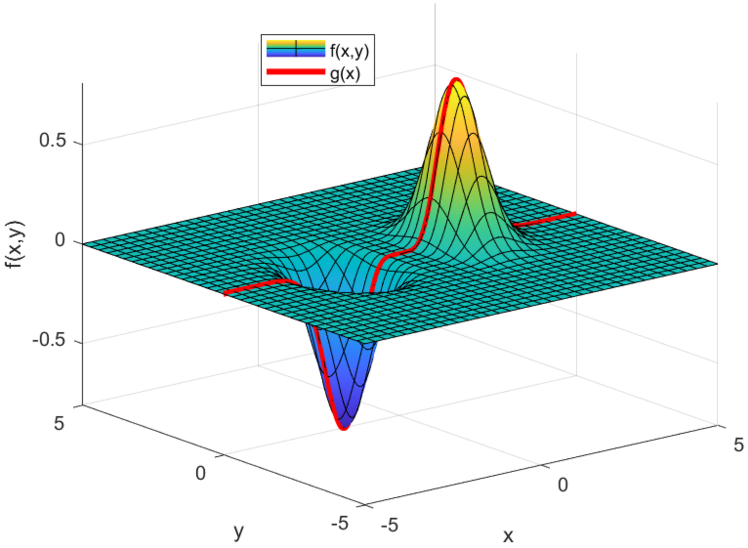
$$\frac{dg(x)}{dx} = -x^4 e^{-x^2} \cdot (2x^2 - 5)$$

Για $|x| \leq \sqrt{5/2}$ προκύπτει $\frac{dg(x)}{dx} \geq 0$, με την ισότητα να ισχύει στα άκρα του διαστήματος του x. Άρα η $g(x)$ γνησίως αύξουσα για τις παραπάνω τιμές του x και άρα το το σημείο $x = 0$ δεν αποτελεί σημείο μεγίστου ή ελαχίστου για την $g(x)$. Συμπερασματικά, το σημείο $A_1 = (0, 0)$ αποτελεί σαγματικό σημείο για την $f(x, y)$.

Η συνάρτηση $g(x)$ παριστάνεται και στο Σχήμα 2.1 και στο Σχήμα 2.2.



Σχήμα 2.1: Η συνάρτηση $g(x)$ σε σύστημα αξόνων δύο διαστάσεων.



Σχήμα 2.2: Η συνάρτηση $f(x, y)$ και η συνάρτηση $g(x) = f(x, 0)$

2.2 Εφαρμογή της ανάλυσης στους αλγορίθμους

Ως κριτήριο τερματισμού των αλγορίθμων που αναφέρθηκαν στην εισαγωγή ορίζεται η κλίση της συνάρτησης στο σημείο x_k της k-οστής επανάληψης να είναι μικρότερη απο μια τιμή ε, δηλαδή $|\nabla f(x_k)| < \epsilon$. Για τον παραπάνω λόγο, ο αλγόριθμος μπορεί να ¨εγκλωβιστεί¨ και να δώσει ως αποτέλεσμα του αλγορίθμου κάποιο τοπικό, αλλά μη ολικό, ελάχιστο ή κάποιο σαγματικό σημείο.

Με βάση τα παραπάνω, αν ο αλγόριθμος βρεθεί κοντά στα σημεία $(0, y)$, $(\sqrt{5/2}, 0)$ ο αλγόριθμος τερματίζει χωρίς να μας δώσει το σημείο ολικού ελαχίστου στο $(-\sqrt{5/2}, 0)$.

Συνεπώς, η εκκίνηση του αλγορίθμου από το (0,0) είναι αδύνατον να επιστρέψει το σημείο ολικού ελαχίστου, καθώς ο αλγόριθμος τερματίζει άμεσα.

Επιπρόσθετα, η εκκίνηση του αλγορίθμου από το σημείο (1,-1) είναι πολύ πιθανό να οδηγήσει στον εγκλωβισμό του αλγορίθμου σε κάποιο απο τα σημεία του συνόλου $A = (0, y)$, λόγω του μηδενισμού της κλίσης στα σημεία αυτά.

Κεφάλαιο 3

Αλγόριθμος Μεγίστης Καθόδου - Θέμα 2

Στο κεφάλαιο αυτό παρουσιάζεται και αναλύεται ο αλγόριθμος της Μεγίστης Καθόδου - Steepest Descent.

3.1 Ο αλγόριθμος

Η μέθοδος Μεγίστης Καθόδου ορίζει την επαναληπτική διαδικασία σύμφωνα με τον τύπο:

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k) \tag{3.1}$$

Δηλαδή επιλέγεται ως διάνυσμα κατεύθυνσης η αρνητική κλίση της συνάρτησης στο σημείο x_k.

3.2 Σταθερό βήμα

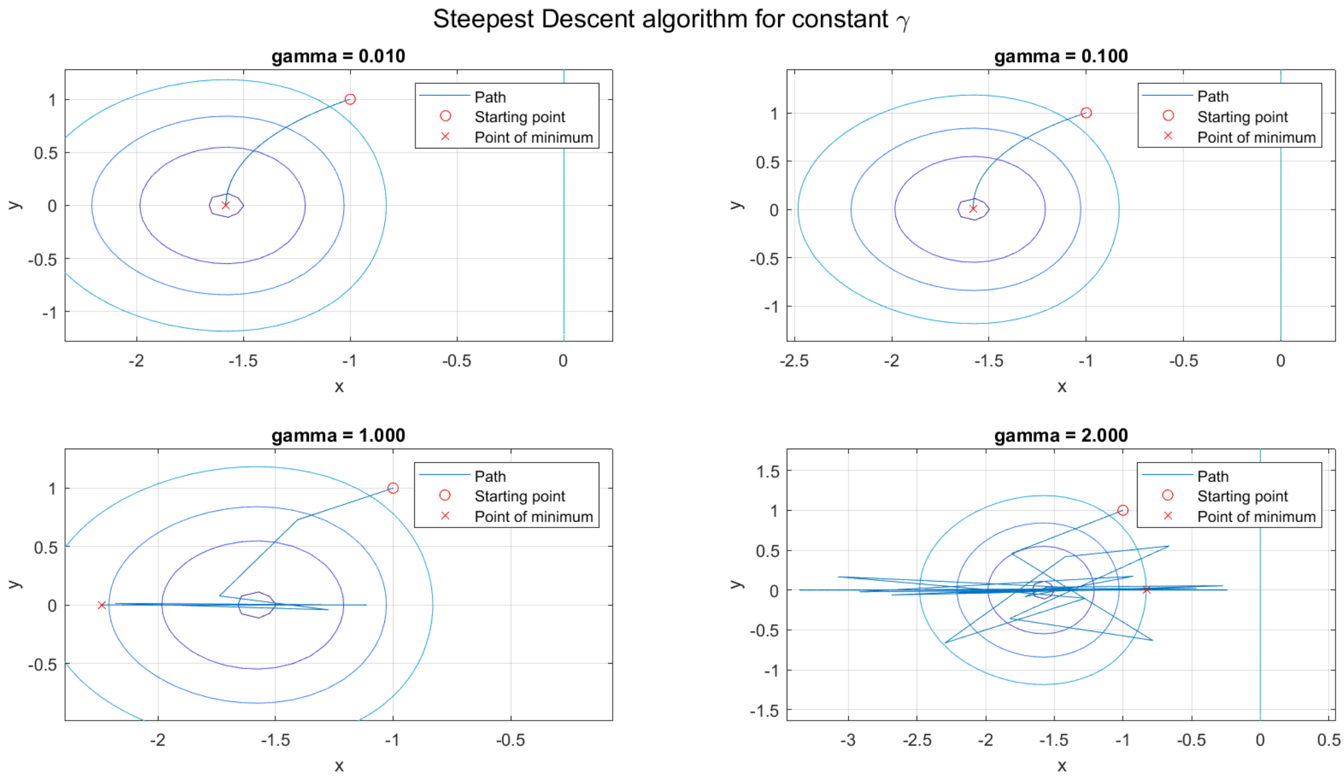
Γίνεται εφαρμογή του αλγορίθμου για διάφορες σταθερές τιμές του γ, και συγκεκριμένα για τις τιμές 0.01, 0.1, 1, 2.

3.2.1 Σημείο εκκίνησης (-1,1)

Η πορεία που ακολουθεί ο αλγόριθμος για την εύρεση ελαχίστου με σφάλμα $\epsilon = 0.001$ και αρχικό σημείο το (-1,1) σημειώνεται στο Σχήμα 3.1.

Παρατηρείται ότι για τα βήματα $\gamma_k = 0.1$ και $\gamma_k = 0.01$ ο αλγόριθμος τερματίζει επιτυχώς φτάνοντας ικανοποιητικά στο σημείο τοπικού ελαχίστου της συνάρτησης. Πρέπει να αναφερθεί ότι ο αλγόριθμος εκτελεί 520 επαναλήψεις για $\gamma_k = 0.01$, και 51 επαναλήψεις για $\gamma_k = 0.1$. Δηλαδή φτάνουμε στο ίδιο αποτέλεσμα με πολύ λιγότερες επαναλήψεις για μεγαλύτερο γ_k.

Από την άλλη μεριά, η σύγκλιση στο σημείο ελαχίστου δεν συμβαίνει για $\gamma_k = 1$ και $\gamma_k = 2$. Όπως φαίνεται, σε αυτές τις περιπτώσεις, ενώ αρχικά η σύγκλιση είναι ταχεία, ο αλγόριθμος οδηγείται σε ταλαντώσεις και τελικά καταλήγει σε σημείο που δεν ικανοποιεί το ελάχιστο σφάλμα. Ο αλγόριθμος τελειώνει με βάση τον μέγιστο αριθμό επαναλήψεων και όχι με βάση το ελάχιστο σφάλμα.



Σχήμα 3.1: Η πορεία του αλγορίθμου Μεγίστης καθόδου για διαφορετικές τιμές του γ, με αρχικό σημείο το (-1,1).

Συμπερασματικά, μπορεί να ειπωθεί ότι η επιλογή του γ_k, δεδομένου του αποδεκτού σφάλματος, καθορίζει την αποτελεσματικότητα και την ταχύτητα του αλγορίθμου. Μεγάλο γ_k οδηγεί σε ταχεία σύγκλιση με κίνδυνο, ωστόσο, να μην επιτευχθεί η εύρεση του ελαχίστου. Μικρό γ_k μπορεί να οδηγήσει σε πολύ αργό αλγόριθμο, λόγω της μικρής σύγκλισης προς το σημείο ελαχίστου σε κάθε επανάληψη.

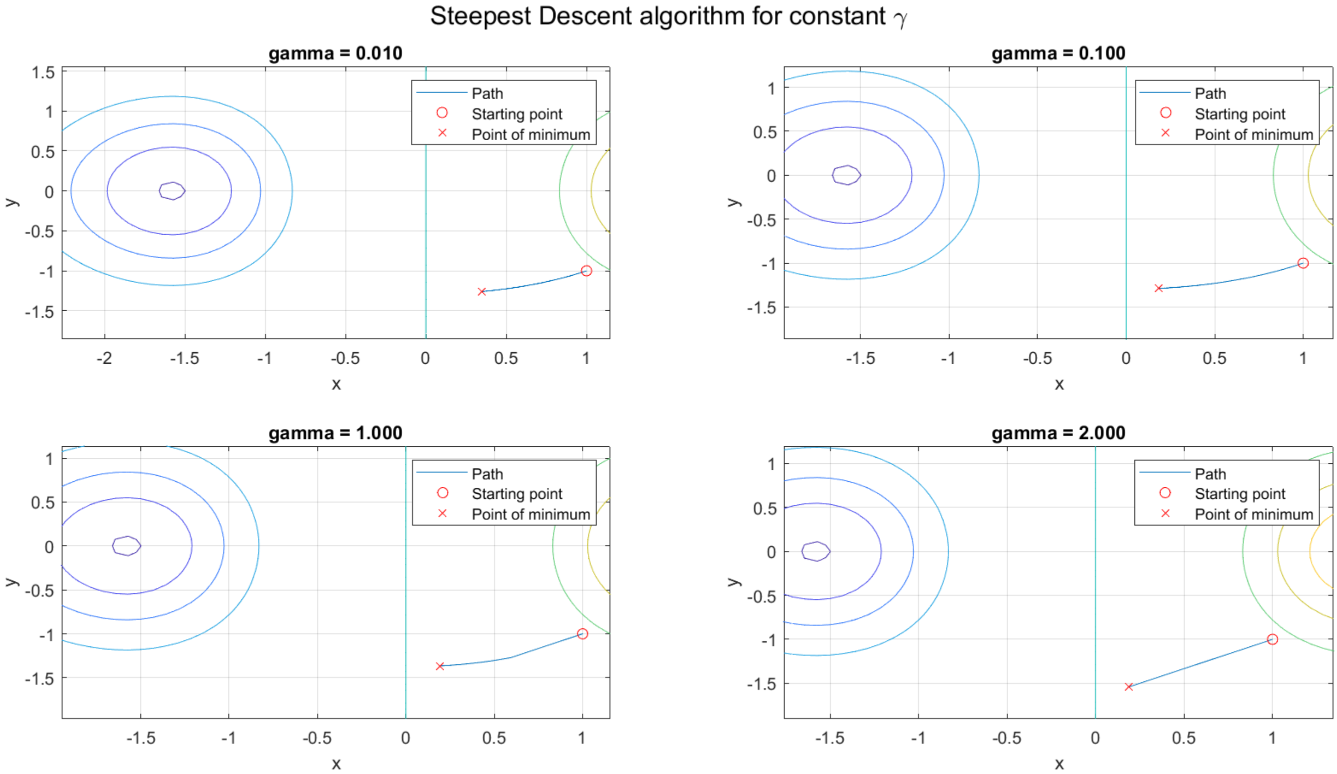
3.2.2 Σημείο εκκίνησης (0,0)

Για εκκίνηση από το (0,0) ο αλγόριθμος δεν πλησιάζει προς το ελάχιστο, αλλά παραμένει στο σημείο εκκίνησης και τερματίζει με την πρώτη επανάληψη, όπως περιγράφηκε στο κεφάλαιο 2.

3.2.3 Σημείο εκκίνησης (1,-1)

Για εκκίνηση από το σημείο (1,-1) ο αλγόριθμος δεν καταλήγει στο ελάχιστο, αλλά τείνει να πλησιάσει στην ευθεία $x = 0$. Τα αποτελέσματα της εκτέλεσης του αλγορίθμου φαίνονται στο Σχήμα 3.2. Για γ = 0.01 η μετακίνηση τους αλγορίθμου είναι πολύ αργή και τερματίζει εξαιτίας του ορίου των 1000 επαναλήψεων. Για τις άλλες τιμές του γ, ο αλγόριθμος τερματίζει νωρίτερα, διότι η κλίση της συνάρτησης πλησιάζει το 0 με αποτέλεσμα ο αλγόριθμος να ¨εγκλωβίζεται¨.

Μια ιδιαίτερη περίπτωση είναι η εφαρμογή πολύ μεγάλου γ έτσι, ώστε ο αλγόριθμος να υπερπηδήσει το εμπόδιο της ευθείας $x = 0$. Ωστόσο, όπως φαίνεται στο Σχήμα 3.3,



Σχήμα 3.2: Η πορεία του αλγορίθμου Μεγίστης Καθόδου για διαφορετικές τιμές του γ, με αρχικό σημείο το (1, -1).

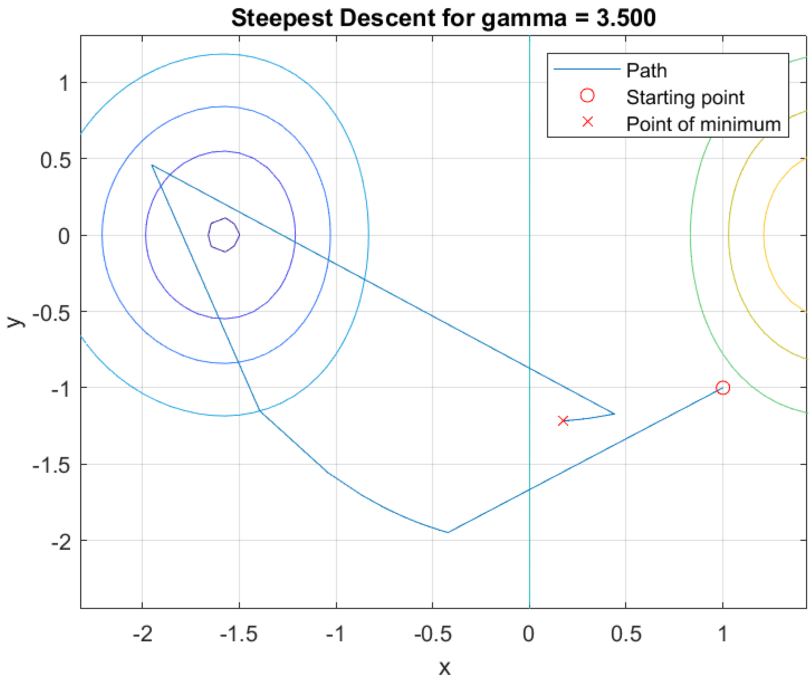
το μεγάλο βήμα γ αποτρέπει την σύγκλιση και ο αλγόριθμος καταλήγει ξανά σε λανθασμένο σημείο. Ίσως κάποια άλλη επιλογή του γ να οδηγούσε σε σύγκλιση κάτω από πολύ αυστηρές προϋποθέσεις (π.χ. το σημείο x_2 της δεύτερης επανάληψης να είναι σε σημείο που, παρά το μεγάλο βήμα, η επόμενη επανάληψη να δώσει x_3 ικανοποιητικά κόντα στο σημείο ελαχίστου). Αυτό, όμως, είναι δύσκολο να επιτευχθεί και δεν συμβαίνει για γ = 3.5.

3.3 Μεταβαλλόμενο βήμα

Όπως φάνηκε στην ανάλυση παραπάνω, η επιλογή σταθερού βήματος μπορεί να οδηγήσει σε αστάθεια του αλγορίθμου ή σε πολύ αργή σύγκλιση. Για τον λόγο αυτό, μπορεί να γίνεται επιλογή του γ_k σε κάθε επανάληψη τέτοια, ώστε να μειωθούν οι πιθανότητες να συμβούν τα παραπάνω. Για να γίνει η ανάλυση αυτή, επιλέγεται ανοχή σφάλματος ξανά $\epsilon = 0.001$ και για την επιλογή του γ_k ελαχιστοποιείται η συνάρτηση $\phi(\gamma) = f(x_k - \gamma\nabla f(x_k))$ ως προς γ, με βάση τις μεθόδους που χρησιμοποιούνται για την ελαχιστοποίηση συναρτήσεων μιας μεταβλητής.

3.3.1 Σημείο εκκίνησης (-1,1)

Για το σημείο εκκίνησης (-1,1) η σύγκλιση είναι επιτυχής, όπως φαίνεται στο Σχήμα 3.4. Η σύγκλιση επιτυγχάνεται σε μόλις 8 επαναλήψεις. Αν και οι επαναλήψεις είναι λιγότερες, ο αριθμός υπολογισμών της μεθόδου αυξάνεται ανά επανάληψη, διότι πλέον



Σχήμα 3.3: Ειδική περίπτωση για την σύγκλιση από το σημείο (1,-1), στην οποία ο αλγόριθμος ξεπερνά το ¨εμπόδιο¨ της ευθείας $x = 0$.

γίνεται και ελαχιστοποίηση της συνάρτησης ϕ που αναφέρθηκε προηγουμένως. Επομένως, η μέθοδος αυτή μπορεί να επιλεγεί ως καλύτερη εναλλακτική του σταθερού βήματος αν θέλουμε να εξασφαλίσουμε σύγκλιση χωρίς ταλαντώσεις και αν δεν μπορεί να γίνει καλή επιλογή ενός σταθερού γ που να εξασφαλίζει σύντομη σύγκλιση.

Όσον αφορά τις τιμές του βήματος, παρατηρείται ότι στα αρχικά βήματα το γ είναι μεγάλο και υπάρχει ταχεία σύγκλιση, ενώ καθώς ο αλγόριθμος πλησιάζει στο σημείο ελαχίστου, η τιμή του γ είναι μικρή και τέτοια, ώστε να μην οδηγείται σε μακρινά σημεία από το πραγματικό σημείο ελαχίστου.

3.3.2 Σημείο εκκίνησης (0,0)

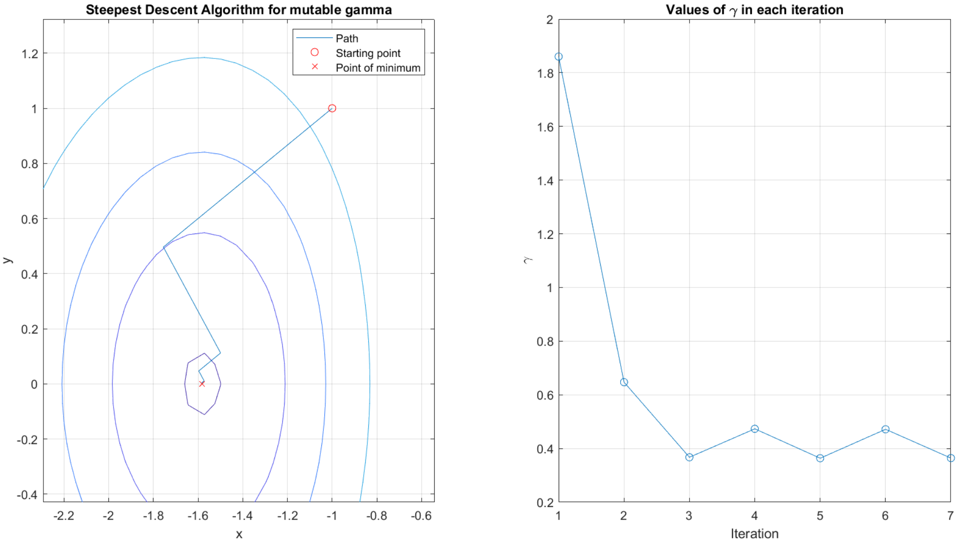
Ο αλγόριθμος τερματίζει άμεσα χωρίς να συγκλίνει.

3.3.3 Σημείο εκκίνησης (1,-1)

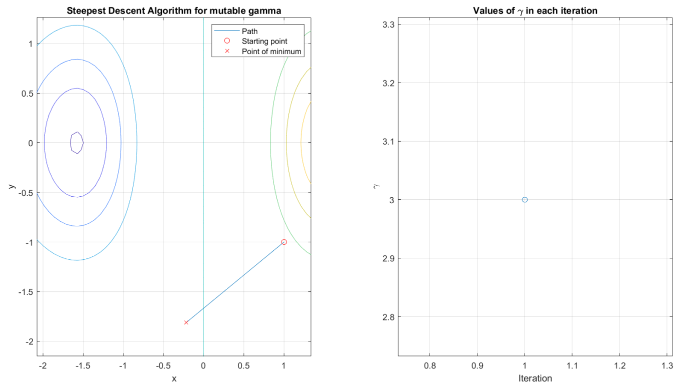
Για εκκίνηση από το σημείο (1,-1) ο αλγόριθμος δεν καταλήγει στο ελάχιστο, αλλά τείνει να πλησιάσει στην ευθεία $x = 0$, όπως ακριβώς και για σταθερό γ. Τα αποτελέσματα της εκτέλεσης του αλγορίθμου φαίνονται στο Σχήμα 3.5. Ο τερματισμός του αλγορίθμου είναι ταχύς αφού γίνεται με την δεύτερη επανάληψη(εδώ υπενθυμίζεται πως ως πρώτη επανάληψη θεωρήθηκε η εξέταση ως ελαχίστου του σημείου εκκίνησης).

3.4 Βήμα με βάση τον κανόνα του Armijo

Ο αλγόριθμος του Armijo και η επιλογή των παραμέτρων περιγράφονται με λεπτομέρεια στην ενότητα 7.1.



Σχήμα 3.4: Η πορεία του αλγορίθμου Μεγίστης καθόδου και οι τιμές του γ για μεταβαλλόμενο βήμα, με αρχικό σημείο το (1, -1).



Σχήμα 3.5: Η πορεία του αλγορίθμου Μεγίστης καθόδου και οι τιμές του γ για μεταβαλλόμενο βήμα, με αρχικό σημείο το (-1, 1).

3.4.1 Σημείο εκκίνησης (-1,1)

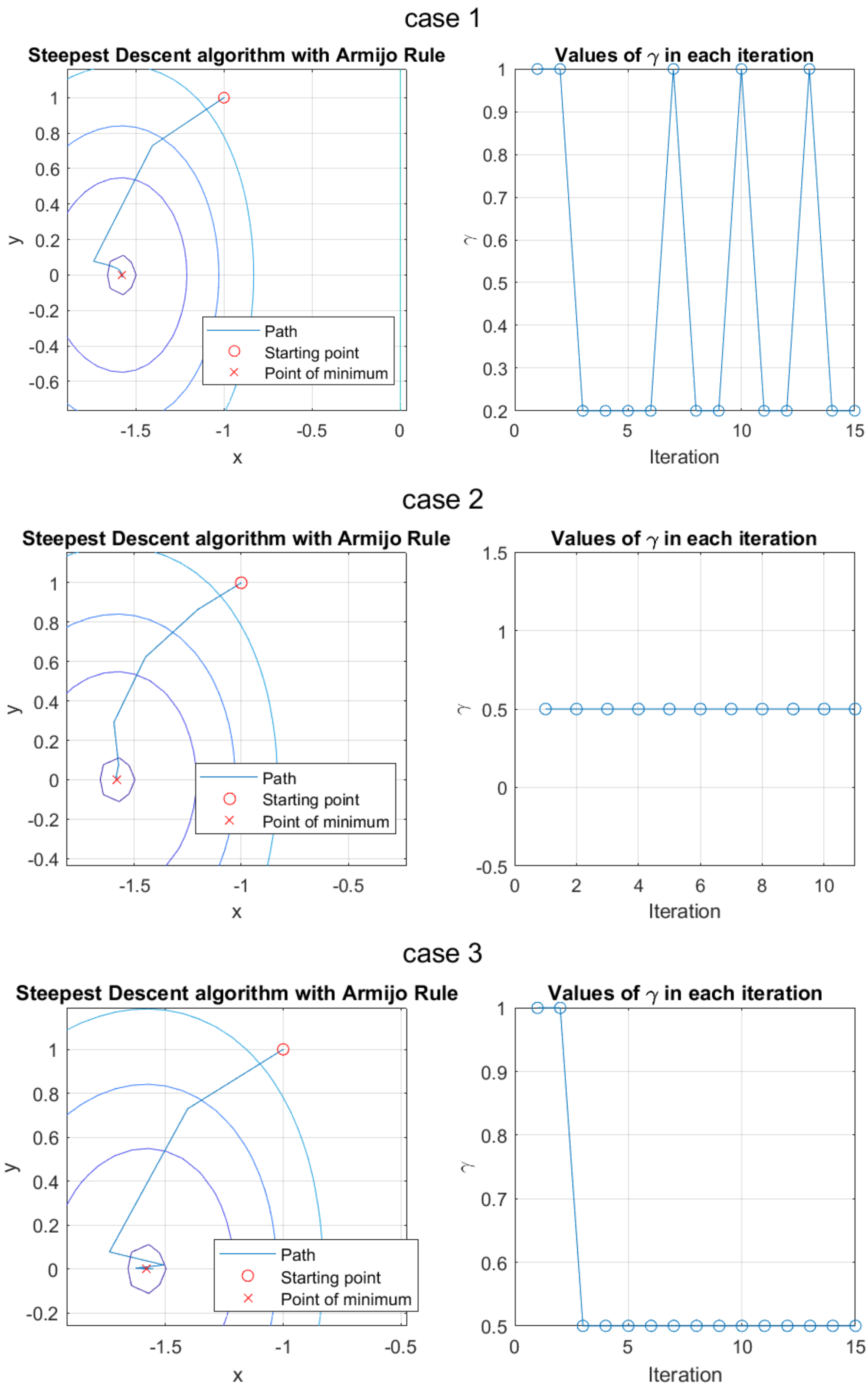
Για αρχικό σημείο το (-1,1), τα αποτέλεσμα φαίνονται στο Σχήμα 3.6. Φαίνεται πως οι επιλογές των παραμέτρων επηρεάζουν άμεσα την ταχύτητα του αλγορίθμου, τις τιμές του βήματος γ, καθώς και το αν θα εμφανίσει κάποια ταλάντωση ο αλγόριθμος.

3.4.2 Σημείο εκκίνησης (0,0)

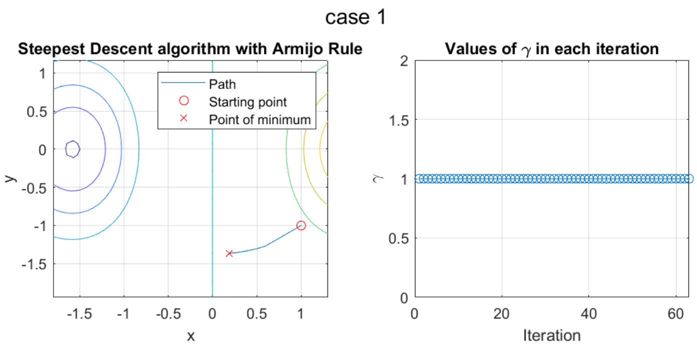
Ξανά ο αλγόριθμος μένει σταθερός και δεν συγκλίνει.

3.4.3 Σημείο εκκίνησης (1,-1)

Όπως και στην περίπτωση του γ που επιλέγεται από την ελαχιστοποίηση της ϕ, ο αλγόριθμος τερματίζει περίπου στο ίδιο σημείο. Ο τερματισμός, ωστόσο, δεν γίνεται άμεσα, αλλά μετά από έναν πεπερασμένο αριθμό επαναλήψεων (64, 126 και 64 για τις παραμέτρους των περιπτώσεων 1,2 και 3 αντίστοιχα). Ενδεικτικά, παρατίθεται το αποτέλεσμα για την περίπτωση 1 στο Σχήμα 3.7.



Σχήμα 3.6: Σύγκλιση της Μεθόδου Μεγίστης Καθόδου με την μέθοδο Armijo.



Σχήμα 3.7: Η πορεία του αλγορίθμου Μεγίστης καθόδου και οι τιμές του γ για μεταβαλλόμενο βήμα, με αρχικό σημείο το (1, -1) - Περίπτωση 1.

Κεφάλαιο 4

Αλγόριθμος Newton - Θέμα 3

Στο κεφάλαιο αυτό παρουσιάζεται και αναλύεται ο αλγόριθμος Newton.

4.1 Ο αλγόριθμος

Ο αλγόριθμος Newton ορίζει την επαναληπτική διαδικασία σύμφωνα με τον τύπο:

$$x_{k+1} = x_k - \gamma_k[\nabla^2 f(x_k)]^{-1}\nabla f(x_k) \tag{4.1}$$

Δηλαδή επιλέγεται ως διάνυσμα κατεύθυνσης το διάνυσμα $-[\nabla^2 f(x_k)]^{-1}\nabla f(x_k)$. Με την προϋπόθεση ο πίνακας $\nabla^2 f(x_k)$ να είναι θετικά ορισμένος, ώστε να ικανοποιείται η ιδιότητα της επαναληπτικής καθόδου. Αν η προϋπόθεση αυτή δεν ικανοποιείται, τότε η μέθοδος είτε δεν ορίζεται (π.χ. πίνακας μη αντιστρέψιμος) ή δίνει λύση μακριά από το σημείο ελαχίστου. Στην ειδική περίπτωση που ο πίνακας αυτός είναι αρνητικά ορισμένος, γίνεται αναζήτηση μεγίστου, και όχι ελαχίστου που είναι το ζητούμενο.

4.2 Σταθερό βήμα

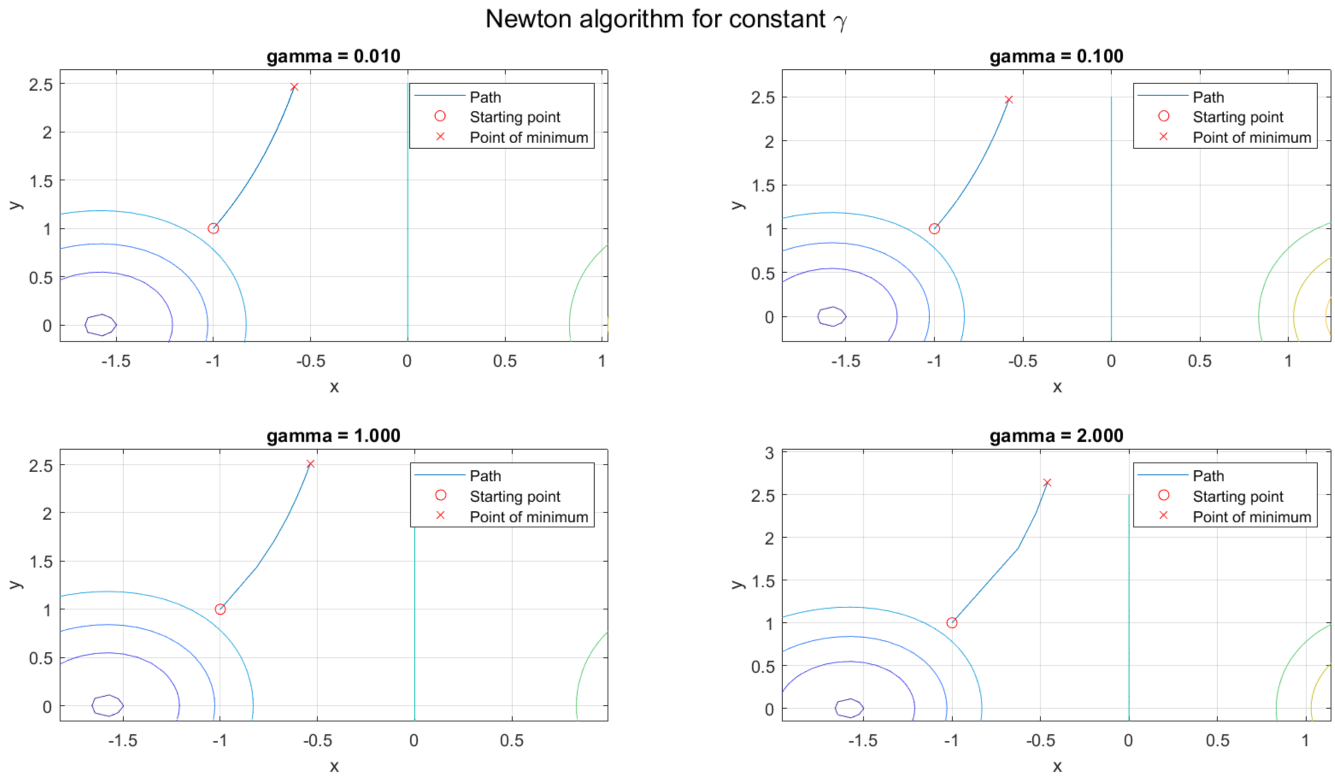
Γίνεται εφαρμογή του αλγορίθμου για διάφορες σταθερές τιμές του γ, και συγκεκριμένα για τις τιμές 0.01, 0.1, 1, 2. Στην υλοποίηση του κώδικα η παράμετρος *check_hessian* δίνει την δυνατότητα στον χρήστη να επιλέξει αν ο κώδικας θα ελέγξει ή όχι τον Εσσιανό πίνακα, σχετικά με το αν είναι θετικά ορισμένος. Σε περίπτωση που ο χρήστης επιλέξει να γίνεται έλεγχος και ο Εσσιανός προκύψει μη θετικά ορισμένος, ο αλγόριθμος τερματίζει με αντίστοιχο μήνυμα λάθους (error).

4.2.1 Σημείο εκκίνησης (-1,1)

Η εφαρμογή του αλγορίθμου δίνει το αποτέλεσμα που φαίνεται στο Σχήμα 4.1. Η μέθοδος αποκλίνει, αφού προκύπτει Εσσιανός σε κάποιο σημείο μη θετικά ορισμένος.

4.2.2 Σημείο εκκίνησης (0,0)

Για εκκίνηση από το (0,0) ο αλγόριθμος τερματίζει άμεσα και παραμένει στο σημείο εκκίνησης, όπως έγινε και με την μέθοδο Μεγίστης Καθόδου.



Σχήμα 4.1: Η πορεία του αλγορίθμου Newton για διαφορετικές τιμές του γ, με αρχικό σημείο το (-1,1). Η μέθοδος αποκλίνει.

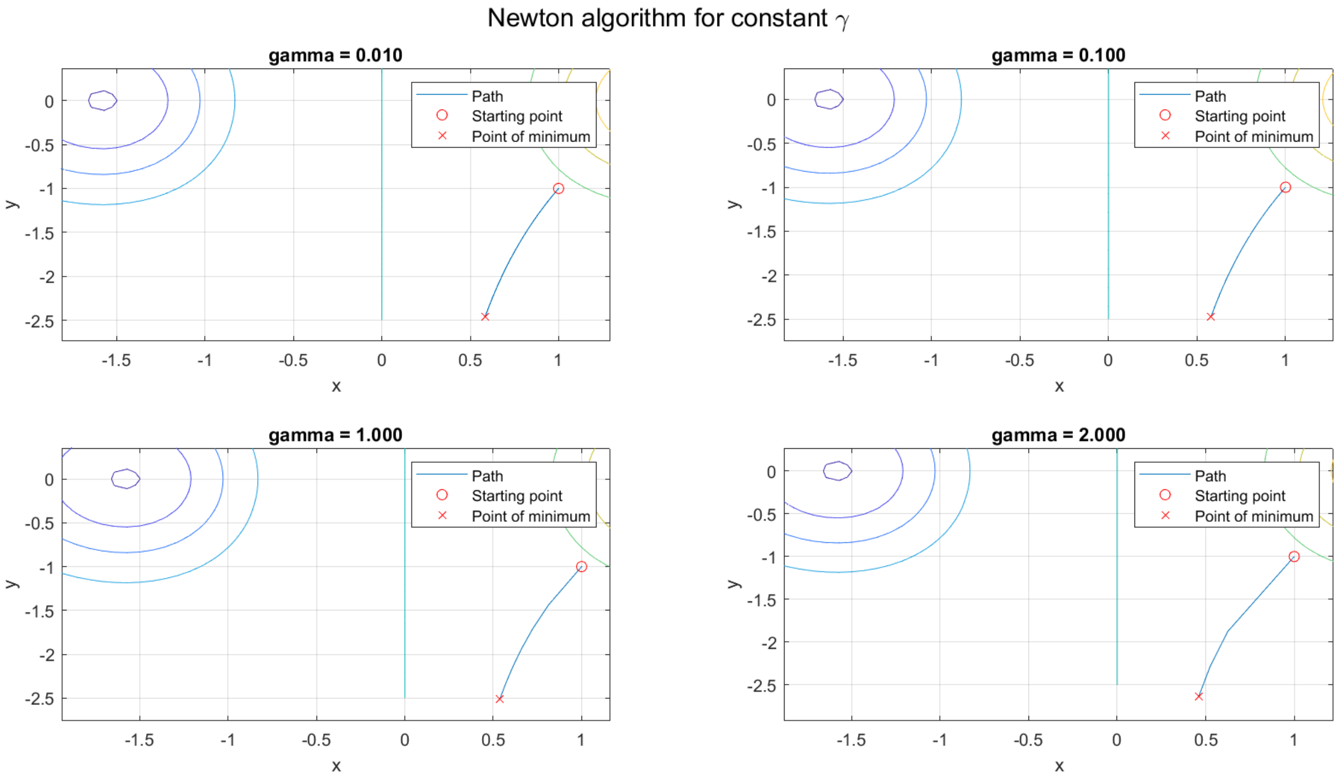
4.2.3 Σημείο εκκίνησης (1,-1)

Και σε αυτή την περίπτωση ο αλγόριθμος εμφανίζει μη επιθυμητή συμπεριφορά (έχοντας ξανά μη θετικό Εσσιανό πίνακα σε κάποια επανάληψη). Τα αποτελέσματα φαίνονται στο Σχήμα 4.2.

4.3 Μεταβαλλόμενο βήμα και βήμα με βάση τον κανόνα του Armijo

Υλοποιήθηκε η μέθοδος για γ προσδιοριζόμενο από την ελαχιστοποίηση της συνάρτησης $\phi(\gamma) = f(x_k + \gamma d_k)$, καθώς και για γ που προκύπτει μέσω του κανόνα του Armijo. Λόγω του μη θετικού Εσσιανού που προκύπτει ωστόσο, οι μέθοδοι δεν μπορούν να δώσουν λογικό αποτέλεσμα (π.χ: καθυστερούν υπερβολικά, αποκλίνουν, οι παράμετροι όπως το μ_k του Armijo αυξάνονται αυθαίρετα) και τερματίζουν με αντίστοιχο μήνυμα λάθους (error). Εδώ πρέπει να σημειωθεί πως για το σημείο (0,0) η μέθοδος τερματίζει άμεσα και επιστρέφει ως απάντηση το σημείο εκκίνησης, όπως και στις προηγούμενες περιπτώσεις.

Οι διάφορες εκδοχές του αλγορίθμου Newton εμφανίζουν επιθυμητή συμπεριφορά σε συναρτήσεις που τηρούν το κριτήριο του Εσσιανού πίνακα, όπως φαίνεται στο κεφάλαιο 7.



Σχήμα 4.2: Η πορεία του αλγορίθμου Newton για διαφορετικές τιμές του γ, με αρχικό σημείο το (1,-1). Η μέθοδος αποκλίνει.

Κεφάλαιο 5

Μέθοδος Levenberg - Marquardt ή Τροποποιημένη Μέθοδος Newton - Θέμα 4

Στο κεφάλαιο αυτό παρουσιάζεται και αναλύεται ο τροποιημένος αλγόριθμος Newton, γνωστός και ως Levenberg-Marquardt.

5.1 Ο αλγόριθμος

Η ιδέα του αλγορίθμου του Newton είναι τέτοια ώστε να ελαχιστοποιεί την τετραγωνική μορφή της αντικειμενικής συνάρτησης και άρα να πετυχαίνει ταχύτερη σύγκλιση από την μέθοδο Μεγίστης Καθόδου. Ωστόσο, όπως φάνηκε από το κεφάλαιο 4, η προϋπόθεση για τον θετικά ορισμένο Εσσιανό πίνακα, δεν επιτρέπει την καθολική εφαρμογή της μεθόδου. Το πρόβλημα αυτό λύνεται με την μέθοδο Levenberg-Mrquardt. Στην μέθοδο αυτή, ορίζεται ο συντελεστής μ_k ώστε ο πίνακας

$D = \nabla^2 f(x_k) + \mu_k I$

να είναι θετικά ορισμένος. Το διάνυσμα κατεύθυνσης ορίζεται ως:

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

Η αντιστροφή του πίνακα D για τον καθορισμό του d_k είναι δυνατή, αφού ο πίνακας είναι θετικά ορισμένος και άρα αντιστρέψιμος.

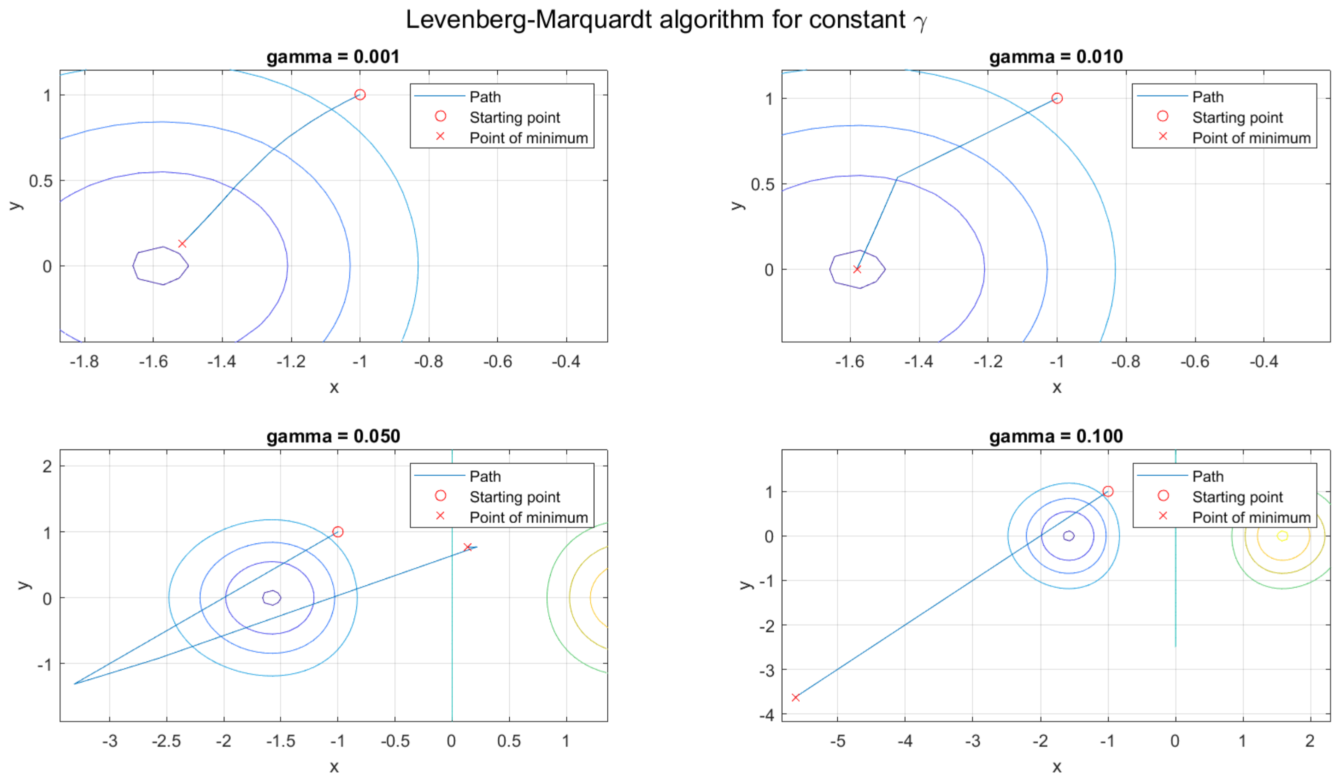
Εδώ σημειώνεται ότι για $\mu_k = 0$, προκύπτει η μέθοδος του Newton, ενώ για μεγάλο μ_k, η μέθοδος τείνει στην μέθοδο Μεγίστης Καθόδου.

5.2 Σταθερό βήμα

Γίνεται εφαρμογή του αλγορίθμου για σταθερό βήμα σε κάθε επανάληψη και συγκεκριμένα για τις τιμές του γ: 0.001, 0.01, 0.05, 0.1 και επιτρεπτό σφάλμα $\epsilon = 0.001$. Εδώ πρέπει να σημειωθεί ότι ο αλγόριθμος Levenberg-Marquardt απαιτεί μικρότερες τιμές του γ από την μέθοδο της Μεγίστης Καθόδου για να λειτουργήσει και να μην αρχίσει να αποκλίνει, όπως θα φανεί εν συνεχεία.

5.2.1 Σημείο εκκίνησης (-1,1)

Όπως φαίνεται για $\gamma \geq 0.05$ ο αλγόριθμος αποκλίνει έντονα και δεν επιστρέφει σωστή απάντηση. Μάλιστα, η εύρεση ελαχίστου για μικρότερες τιμές του γ είναι αργή, αφού η σύκλιση δεν είναι γρήγορη. Αυτό φαίνεται και για γ = 0.001, στην οποία περίπτωση ο αλγόριθμος συγκλίνει σωστά, αλλά τερματίζει πρόωρα λόγω του ορίου επαναλήψεων. Συμπερασματικά, μπορεί να ειπωθεί ότι η επιλογή του γ στον συγκεκριμένο αλγόριθμο επηρεάζει κατά πολύ το τελικό αποτέλεσμα. Τα αποτελέσματα για το σημείο (-1,1) φαίνονται στο Σχήμα 5.1.



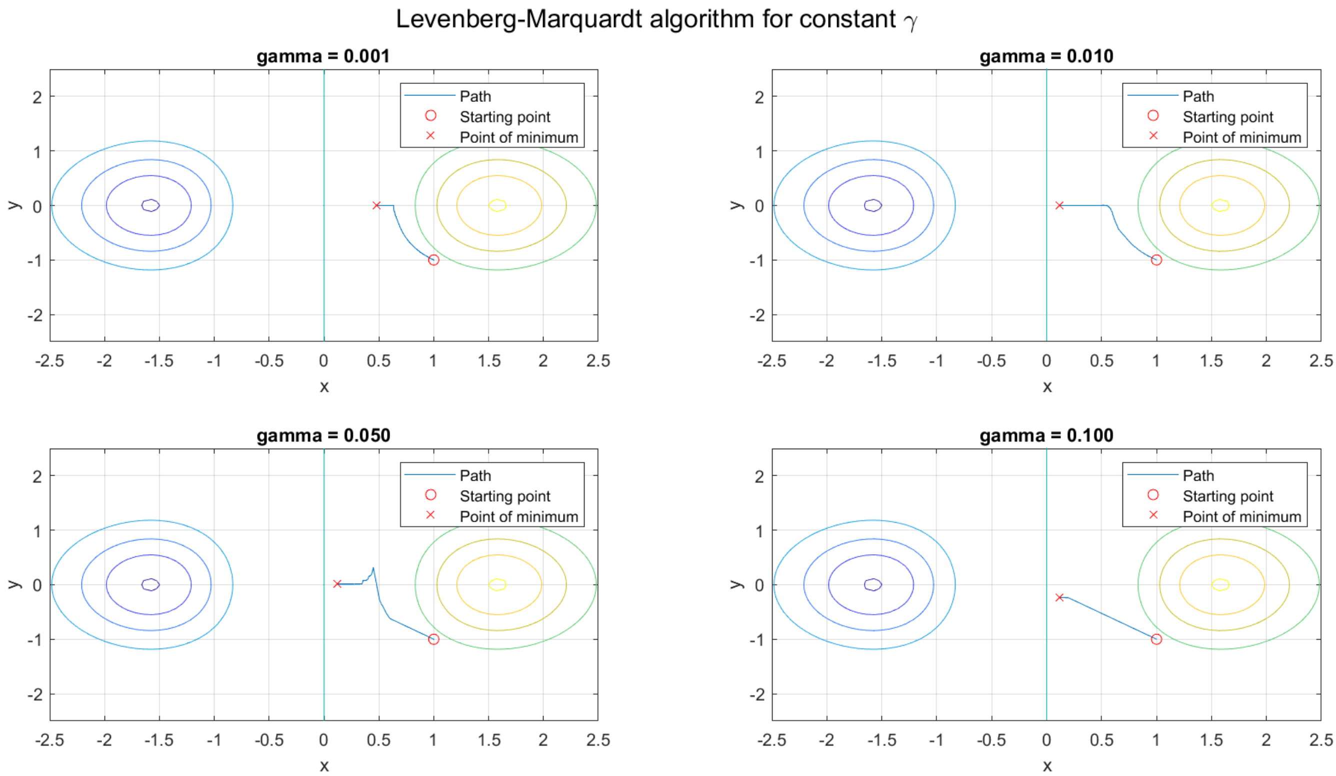
Σχήμα 5.1: Η πορεία του αλγορίθμου Levenberg-Marquardt για διαφορετικές τιμές του γ, με αρχικό σημείο το (-1,1)

5.2.2 Σημείο εκκίνησης (0,0)

Όπως και στην έως τώρα ανάλυση ο αλγόριθμος τερματίζει πριν ξεκινήσει η επαναληπτική διαδικασία.

5.2.3 Σημείο εκκίνησης (1, -1)

Σε αυτή την περίπτωση ο αλγόριθμος δεν καταλήγει στο ελάχιστο, αλλά εγκλωβίζεται κοντά στο (0,0), ωστόσο, η σύγκλιση είναι συγκριτικά καλύτερη από τον αλγόριθμο της Μεγίστης Καθόδου. Τα αποτελέσματα φαίνονται στο Σχήμα 5.2. Πειραματικά, μπορεί να βρεθεί ότι για κάποιες τιμές του γ ο αλγόριθμος ξεπερνά την ευθεία $x = 0$ και εμφανίζει καλύτερες πιθανότητες σύγκλισης. Για γ = 0.4 παραδείγματος χάριν, ο αλγόριθμος χαράσει την πορεία που φαίνεται στο Σχήμα 5.3.



Σχήμα 5.2: Η πορεία του αλγορίθμου Levenberg-Marquardt για διαφορετικές τιμές του γ, με αρχικό σημείο το (1,-1).

5.3 Μεταβαλλόμενο βήμα

Σε κάθε επανάληψη επιλέγεται βήμα από την ελαχιστοποίηση της συνάρτησης $\phi(\gamma) = f(x_k + \gamma d_k)$.

5.3.1 Σημείο εκκίνησης (-1,1)

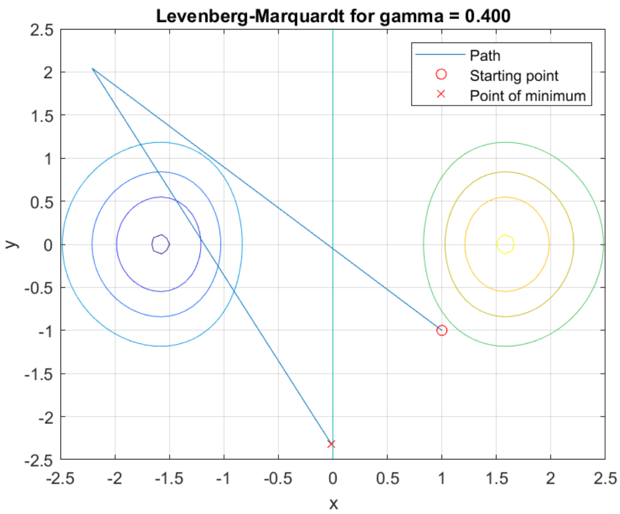
Για εκκίνηση από το (-1,1) ο αλγόριθμος συγκλίνει σε μόλις 4 επαναλήψεις. Μάλιστα το γ αυξάνεται σε κάθε επανάληψη σε αντίθεση με την μέθοδο Μεγίστης Καθόδου, στην οποία το γ είχε μειούμενη τάση. Το γεγονός αυτό αφείλεται στο μεγάλο μέτρο του διανύσματος κατεύθυνσης στις αρχικές επαναλήψεις του αλγορίθμου. Τα αποτελέσματα φαίνονται στο Σχήμα 5.4.

5.3.2 Σημείο εκκίνησης (0,0)

Ο αλγόριθμος τερματίζει πριν ξεκινήσει η επαναληπτική διαδικασία.

5.3.3 Σημείο εκκίνησης (1,-1)

Η σύγκλιση με αρχικό σημείο το (1,-1) είναι επιτυχής με την μέθοδο αυτή, παρόλο που με τις έως τώρα μεθόδους αυτό ήταν αδύνατο. Η επιτυχία οφείλεται στην πολύ καλή μέθοδο για την επιλογή του διανύσματος κλίσης και στην επιλογή του γ μέσω της $\phi(\gamma)$



Σχήμα 5.3: Ειδική περίπτωση του αλγορίθμου Levenberg-Marquardt.

που οδηγεί μακριά από την περιοχή της ευθείας $x = 0$. Τα αποτελέσματα φαίνονται στο Σχήμα 5.5. Και εδώ το γ έχει αυξητική τάση.

5.4 Βήμα με βάση τον κανόνα του Armijo

Υλοποιείται ο αλγόριθμος και με βάση τον κανόνα του Armijo με τις παραμέτρους που ορίζονται στο κεφάλαιο 7.

5.4.1 Σημείο εκκίνησης (-1,1)

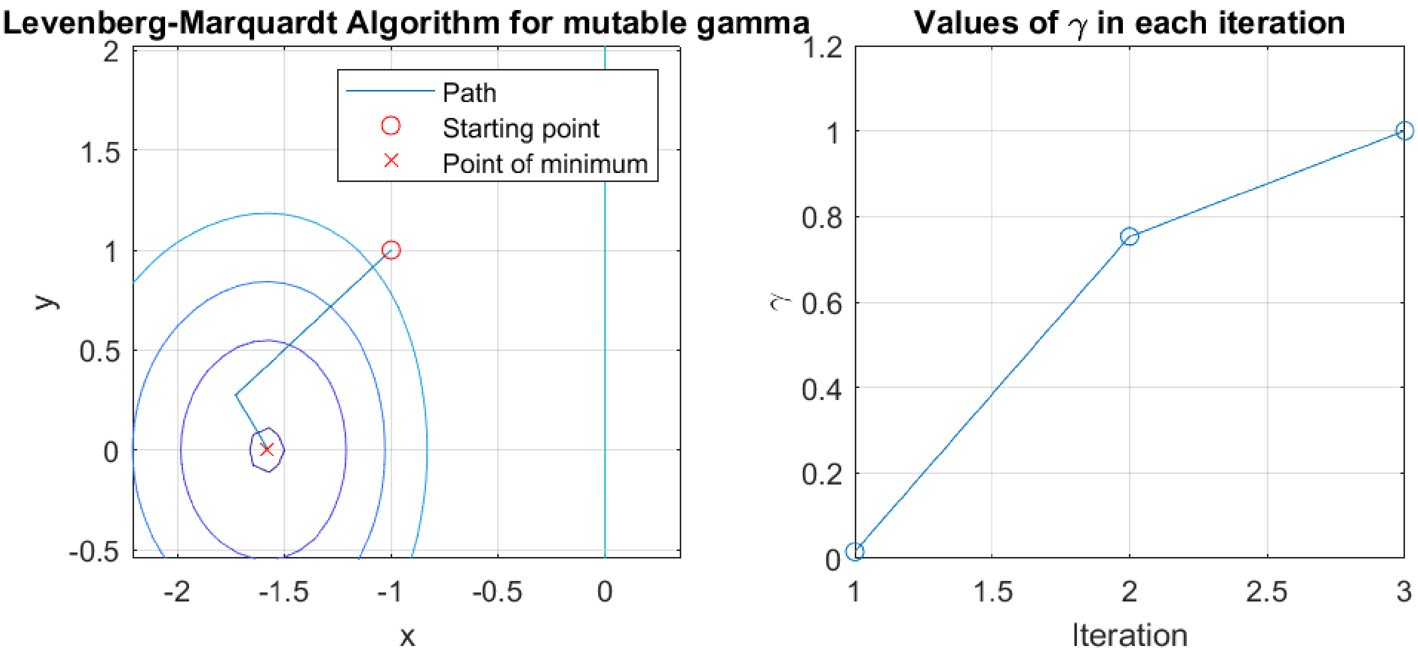
Τα αποτελέσματα φαίνονται στο Σχήμα 5.6. Και εδώ φαίνεται η μεγάλη εξάρτηση από τις παραμέτρους του κανόνα Armijo, σχετικά με την ταχύτητα σύγκλισης και την ύπαρξη ταλαντώσεων. Ο αριθμός επαναλήψεων είναι μεταξύ 5 και 20, συνεπώς, μπορεί να υπάρξει ταχύτερη σύγκλιση σε σχέση με το γ που επιλέγεται μέσω της ϕ αν λάβουμε υπόψιν ότι ο αριθμός υπολογισμών σε κάθε επανάληψη είναι μειωμένος με το κανόνα Armijo.

5.4.2 Σημείο εκκίνησης (0,0)

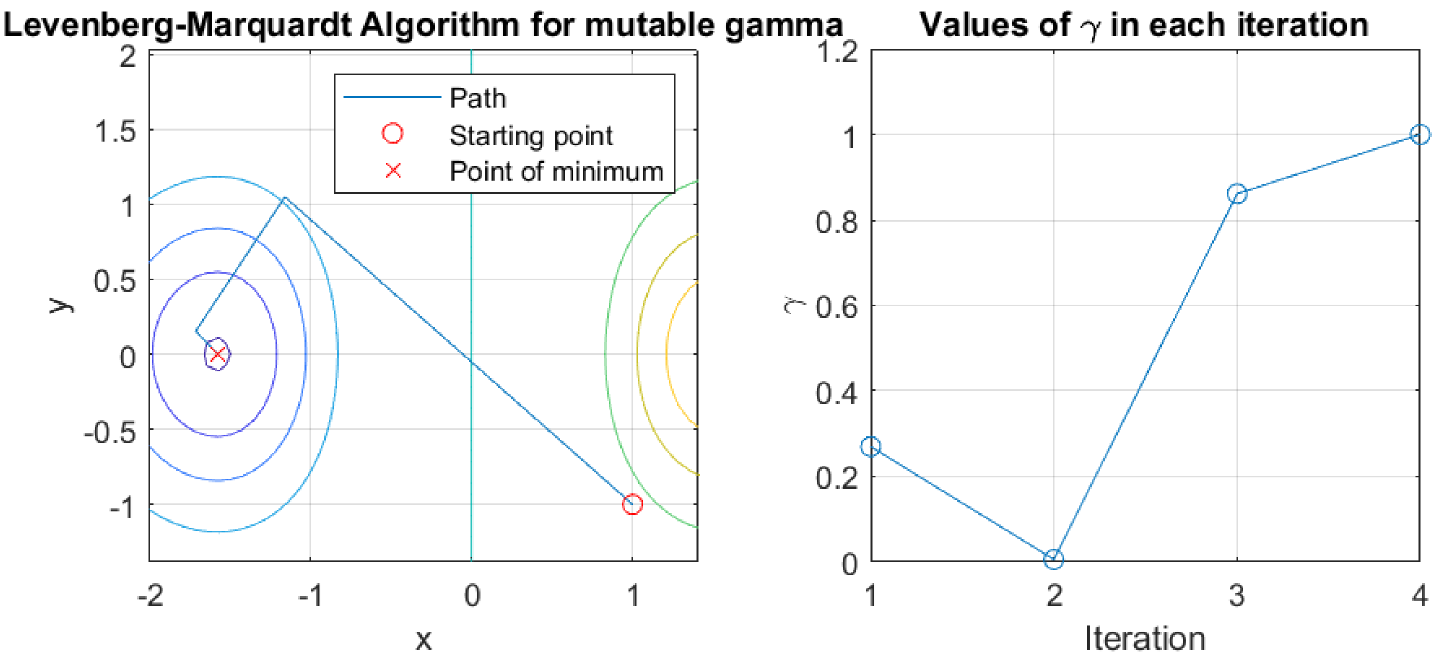
Ο αλγόριθμος τερματίζει πριν ξεκινήσει η επαναληπτική διαδικασία.

5.4.3 Σημείο εκκίνησης (1,-1)

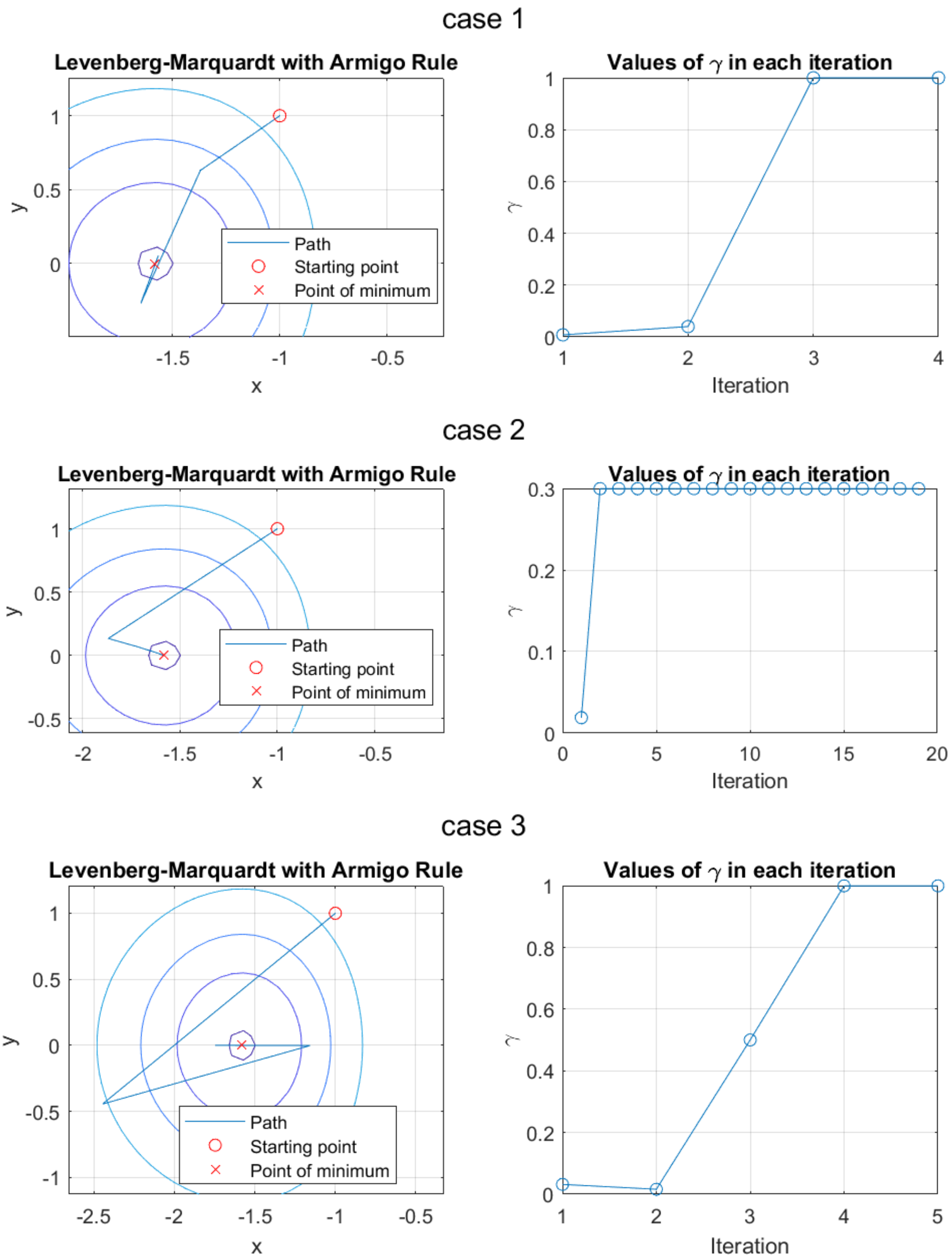
Η μέθοδος πετυχαίνει σύγκλιση μόνο για την 2η περίπτωση των παραμέτρων, αν και με ταλαντώσεις, ενώ στις άλλες αποκλίνει. Τα αποτελέσματα της πετυχημένης σύγκλισης φαίνονται στο Σχήμα 5.7. Η εξάρτηση από την επιλογή των παραμέτρων στην μέθοδο Levenberg-Marquardt είναι έντονη και εδώ.



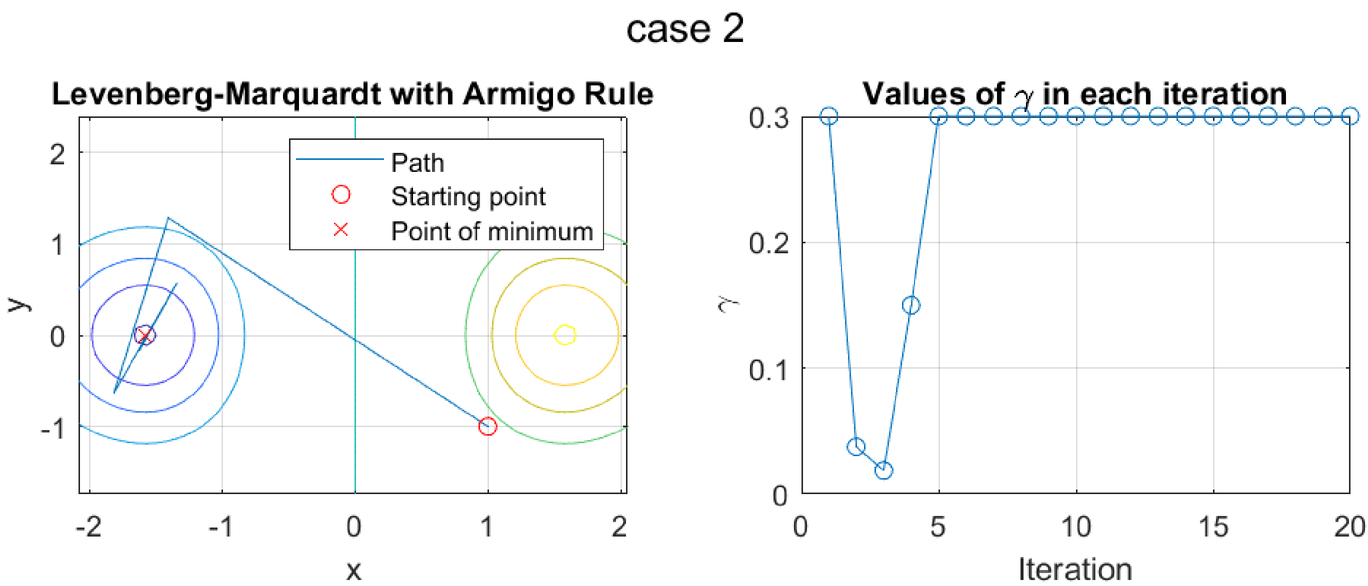
Σχήμα 5.4: Η πορεία του αλγορίθμου Levenberg-Marquardt για μεταβαλλόμενες τιμές του γ, με αρχικό σημείο το (-1,1)



Σχήμα 5.5: Η πορεία του αλγορίθμου Newton για μεταβαλλόμενες τιμές του γ, με αρχικό σημείο το (1,-1)



Σχήμα 5.6: Σύγκλιση της Μεθόδου Levenberg-Marquardt με την μέθοδο Armijo.



Σχήμα 5.7: Η πορεία του αλγορίθμου Levenberg -Marquardt με τον κανόνα Armijo, με αρχικό σημείο το (-1,1)

Κεφάλαιο 6

Συμπεράσματα

6.1 Αλγόριθμοι Υπολογισμού Ελαχίστου

Η ανάλυση που προηγήθηκε μας δίνει τα εξής:

- Η μέθοδος Μεγίστης Καθόδου είναι μια σχετικά σταθερή μέθοδος, η οποία ωστόσο, μπορεί να καταστεί αργή λόγω της μικρής προόδου σε κάθε επανάληψη. Η επιλογή του γ επηρεάζει έντονα την ταχύτητα του αλγορίθμου, καθώς επίσης και την ικανότητα επιστροφής σωστής απάντησης.

- Η μέθοδος Newton είναι γενικά ταχύτερη από την μέθοδο Μεγίστης Καθόδου λόγω της τετραγωνικής προσέγγισης. Παρ'όλα αυτά, απαιτείται ο Εσσιανός πίανακας σε κάθε σημείο αναζήτησης να είναι θετικά ορισμένος για να επιτυγχάνεται η επαναληπτική διαδικασία. Αυτό, ωστόσο, δεν είναι πάντα εφικτό και όπως και στο παράδειγμα της αναφοράς οδηγεί σε σφάλμα-απόκλιση από την πραγματική λύση του προβλήματος βελτιστοποίησης.

- Η λύση για τα προβλήματα της μεθόδου Newton δίνεται από την τροποποιημένη μέθοδο, Levenberg-Marquardt, η οποία επιλέγει το διάνυσμα κατεύθυνσης με βάση τον τροποποιημένο Εσσιανό πίνακα της αντικειμενικής συνάρτησης, ώστε η σύγκλιση να επιτυγχάνεται ακόμα και αν ο Εσσιανός δεν είναι θετικά ορισμένος. Μάλιστα, ο αλγόριθμος αυτός φαίνεται να είναι πιο αποτελεσματικός από τους δύο προηγούμενους, αφού συνδιάζει τα θετικά στοιχεία των δύο μεθόδων. Μια αρνητική πτυχή του αλγορίθμου είναι η μεγάλη εξάρτησή του από τις αρχικές παραμέτρους, οι οποίες μπορεί να τον καταστήσουν αργό ή να τον κάνουν να αποκλίνει. Για τον λόγο αυτό είναι προτιμότερη η χρήση του με τον καθορισμό του γ μέσω της συνάρτησης ϕ, παρά με σταθερό γ ή με γ μέσω του κανόνα Armijo.

6.2 Επιλογή Βήματος γ

Όσον αφορά την επιλογή του βήματος γ:

- Η χρήση σταθερού βήματος είναι μια πολύ απλή διαδικασία που μειώνει το πλήθος των αριθμητικών υπολογισμών ανά επανάληψη, αλλά η επιλογή του μπορεί να καταστεί δύσκολη. Μάλιστα, η επιλογή μεγάλου ή μικρού βήματος οδηγεί σε αργή σύγκλιση ή σε απόκλιση από την πραγματική λύση του προβλήματος.

- Η χρήση μεταβαλλόμενου βήματος μέσω της ελαχιστοποίησης της συνάρτησης $\phi(\gamma)$ οδηγεί στα καλύτερα αποτελέσματα, ωστόσο επιβαρύνει τον αλγόριθμο με αυξημένο αριθμό υπολογισμών. Ακόμα και αν ο αλγόριθμος εμφανίζει λίγες επαναλήψεις, κάθε επανάληψη φέρει μεγάλο κόστος σε υπολογιστική ισχύ. Για αυτόν τον λόγο, η μέθοδος μπορεί να οδηγήσει τον αλγόριθμο σε χρονική καθυστέρηση.
- Η επιλογή του βήματος μέσω του κανόνα του Armijo είναι μια ενδιάμεση λύση. Προσφέρει ταχύτερη επιλογή γ σε σχέση με την ελαχιστοποίηση της ϕ, και δεν αφήνει την επιλογή του γ στην κρίση του χρήστη της μεθόδου. Ωστόσο, οι επιλογές του γ δεν είναι τόσο αξιόπιστες όσο με την μέθοδο της συναρτήσεως ϕ, ούτε έχουν μηδενικό κόστος υπολογισμού, όπως συμβαίνει για σταθερό βήμα.

6.3 Επιλογή Αρχικού Σημείου

Η επιλογή αρχικού σημείου είναι καθοριστική για την αποτελεσματικότητα του αλγορίθμου. Όπως φάνηκε. η επιλογή του σημείου (-1,1) ήταν η ιδανικότερη και οδηγεί τον αλγόριθμο σε σύγκλιση (δεδομένου ότι τηρούνται και άλλες προϋποθέσεις, όπως καλή επιλογή του βήματος γ). Η επιλογή του (0,0), όπου η κλίση της συνάρτησης ήταν μηδενική, εγκλωβίζει τον αλγόριθμο στο σημείο έναρξης και δεν του επιτρέπει να κινηθεί προς το ολικό ελάχιστο. Τέλος, το σημείο (1,-1) είναι μια ενδιάμεση λύση, διότι ο αλγόριθμος δεν εγκλωβίζεται μεν στην εκκίνηση, αλλά είναι σχετικά δύσκολο να πετύχει σωστή σύγκλιση και με λίγες επαναλήψεις. Συγκεκριμένα, θα πρέπει να ξεπεραστεί το εμπόδιο της ευθείας $x = 0$, όπου μηδενίζει η παράγωγος και εν συνεχεία ο αλγόριθμος να κινηθεί έτσι, ώστε να μην αποκλίνει ή ταλαντωθεί μακριά από το ζητούμενο σημείο ελαχίστου.

Η βέλτιστη επιλογή σημείου εκκίνησης, συμπερασματικά, πρέπει να γίνει όσο το δυνατόν εγγύτερα στο σημείο ελαχίστου. Επιπρόσθετα, το σημείο εκκίνησης δεν πρέπει να έχει μηδενική κλίση (εκτός αν είναι το ζητούμενο σημείο ελαχίστου) ή να βρίσκεται κοντά σε σημεία με μηδενική παράγωγο, στα οποία μπορεί να εγκλωβιστεί ο αλγόριθμος (όπως στα σημεία της ευθείας $x = 0$ της αναφοράς).

Κεφάλαιο 7

Παράρτημα

7.1 Η μέθοδος του Armijo

Ο κανόνας του Armijo αποτελεί μια μέθοδο καθορισμού του βήματος σε κάθε επανάληψη. Με την μέθοδο αυτή αποφεύγεται η χρήση σταθερού βήματος, το οποίο μπορεί να επιφέρει αστάθεια στον αλγόριθμο, αλλά και η μεγάλη καθυστέρηση εξαιτίας του υπολογισμού ελαχίστου της συνάρτησης $\phi(\gamma_k) = f(x_k - \gamma\nabla f(x_k))$. Ο Κανόνας επιλέγει ως βήμα:

$$\gamma_k = s \cdot \beta^{m_k} \tag{7.1}$$

Με m_k ο μικρότερος μη αρνητικός ακέραιος για τον οποίο ισχύει:

$$f(x_k) - f(x_{k+1}) \geq -a\beta^{m_k} d_k^T \cdot \nabla f(x_k) \tag{7.2}$$

Για την ανάλυση των αλγορίθμων με την χρήση του κανόνα αυτού επιλέχτηκαν 3 διαφορετικές περιπτώσεις για τις παραμέτρους:

- **Περίπτωση 1:** α = 0.01, β = 0.2, s = 1
- **Περίπτωση 2:** α = 0.01, β = 0.5, s = 0.5
- **Περίπτωση 3:** α = 0.001, β = 0.5, s = 1

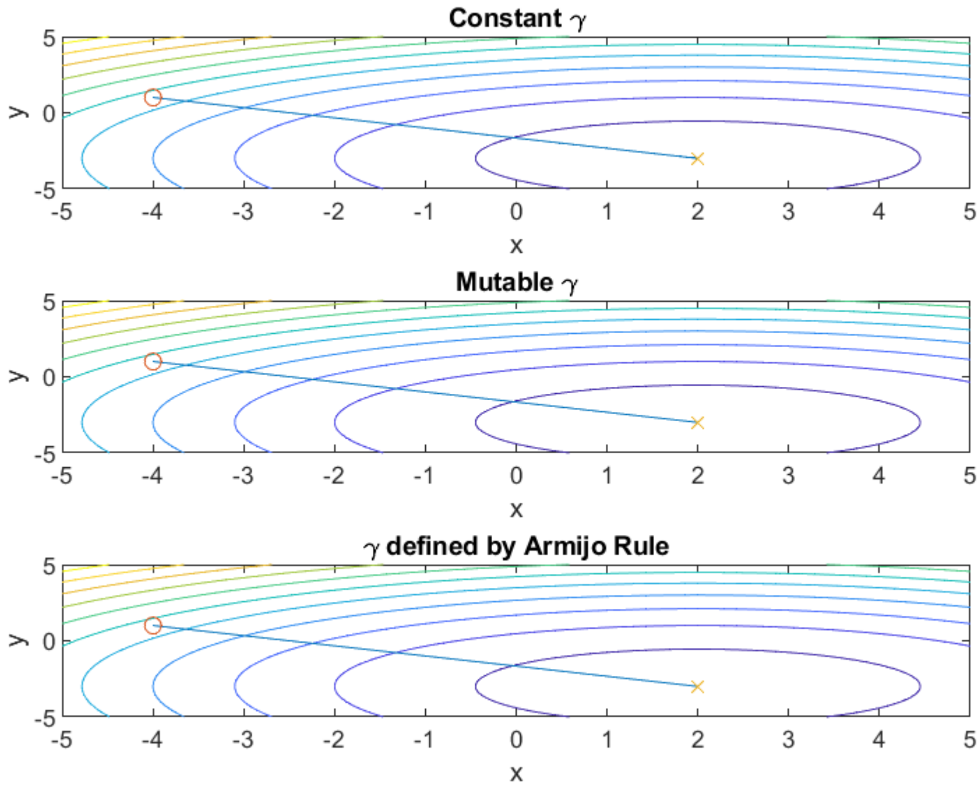
Κάθε περίπτωση επισημαίνεται στα γραφήματα ως *case i*, όπου *i* ο αριθμός της περίπτωσης.

7.2 Εφαρμογή του αλγορίθμου Newton

Γίνεται εφαρμογή του αλγορίθμου Newton στην συνάρτηση $h(x, y) = (x-2)^2 + (y+3)^2 + 4$ για να επιβεβαιωθεί η σωστή λειτουργία. Η συνάρτηση h έχει θετικά ορισμένο Εσσιανό και έχει ελάχιστο στο σημείο (2,-3), όπως δείχνεται παρακάτω. Η σύγκλιση είναι επιτυχής όπως φαίνεται στο Σχήμα 7.1

$$\nabla h(x, y) = \vec{0} \Rightarrow \begin{bmatrix} 2x-4 \\ 2y+6 \end{bmatrix} = \vec{0} \Rightarrow \begin{cases} x = 2 \\ y = -3 \end{cases} \qquad \nabla^2 h(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Ο Εσσιανός είναι θετικά ορισμένος, αφού έχει μη μηδενική ορίζουσα και θετικές ιδιοτιμές. Άρα το (2,-3) σημείο ολικού ελαχίστου.



Σχήμα 7.1: Η επιτυχημένη σύγκλιση του αλγορίθμου Newton για θετικά οριμένο Εσσιανό.

7.3 Παρατήρηση για τον Αριθμό Επαναλήψεων και το επιτρεπτό σφάλμα στους Αλγορίθμους

Ως πρώτη επανάληψη στους αλγορίθμους προσμετράται ο έλεγχος του αρχικού σημείου $(x1, y1)$ ως υποψήφιο σημείο ελαχίστου. Επομένως αν κάποιος αλγόριθμος ξεκινήσει από το (x_1, y_1) και βρει ελάχιστο στο επόμενο βήμα, δηλαδή στο (x_2, y_2), ως πλήθος επανάληψης επιστρέφεται το $k = 2$.

Επίσης, πρέπει να αναφερθεί ότι στην εφαρμογή των αλγορίθμων, όπου δεν αναφέρεται ρητά, ως επιτρεπτό σφάλμα ορίστηκε το $\epsilon = 0.001$.

7.4 Ελαχιστοποίηση Μονοδιάστατης Συνάρτησης

Στην δεύτερη μέθοδο επιλογής του βήματος γ που χρησιμοποιείται από τους αλγορίθμους, έγινε χρήση της μεθόδου Χρυσού Τομέα. Η μέθοδος είναι σχετικά απλή, δεν απαιτεί πολλούς υπολογισμούς και δεν προϋποθέτει από την αντικειμενική συνάρτηση να είναι παραγωγίσιμη, με αποτέλεσμα να είναι μια αξιόπιστη επιλογή για την ελαχιστοποίηση της $\phi(\gamma)$. Φυσικά μπορούν να χρησιμοποιηθούν και άλλες μέθοδοι, όπως αυτές που περιγράφηκαν στην πρώτη εργαστηριακή άσκηση.

Κατάλογος Σχημάτων