Τεχνικές Βελτιστοποίησης

Ελαχιστοποίηση συναρτήσεων πολλών μεταβλητών χωρίς περιορισμούς με χρήση παραγώγων

2η Εργαστηριακή Άσκηση

Ονοματεπώνυμο: **Ρεπάνης Γεώργιος – Δημήτριος**

Νοέμβριος 2025

ΑΕΜ: 10588

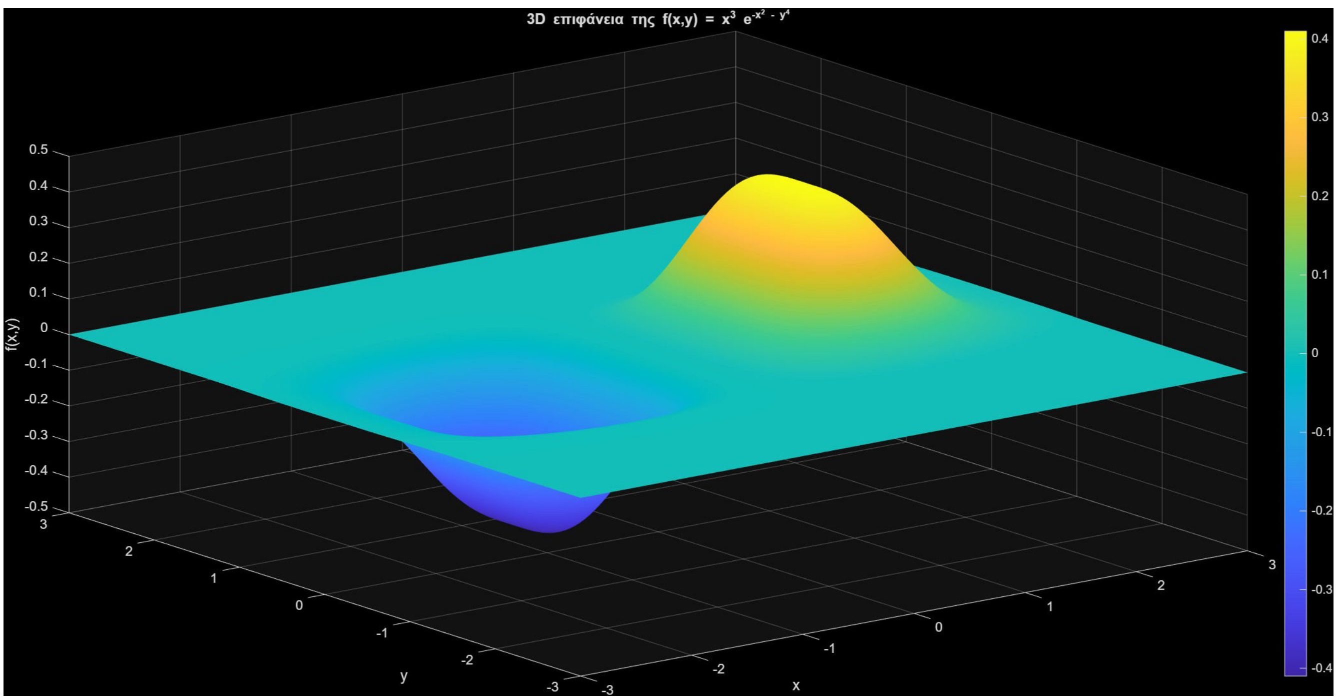
Εισαγωγή:
Στην παρούσα εργαστηριακή άσκηση ασχολούμαστε με το πρόβλημα ελαχιστοποίησης μιας συνεχούς συνάρτησης πολλών μεταβλητών χωρίς περιορισμούς, αξιοποιώντας πληροφορία παραγώγων. Συγκεκριμένα, μελετάμε την αντικειμενική συνάρτηση $f(x,y)=x^3 e^{-x^2-y^4}$ και εφαρμόζουμε επαναληπτικούς αλγορίθμους βελτιστοποίησης : μέθοδος Μέγιστης Καθόδου (Steepest Descent), μέθοδος Newton και μέθοδος Levenberg-Marquardt από διαφορετικά αρχικά σημεία και διαφορετικά βήματα γ.

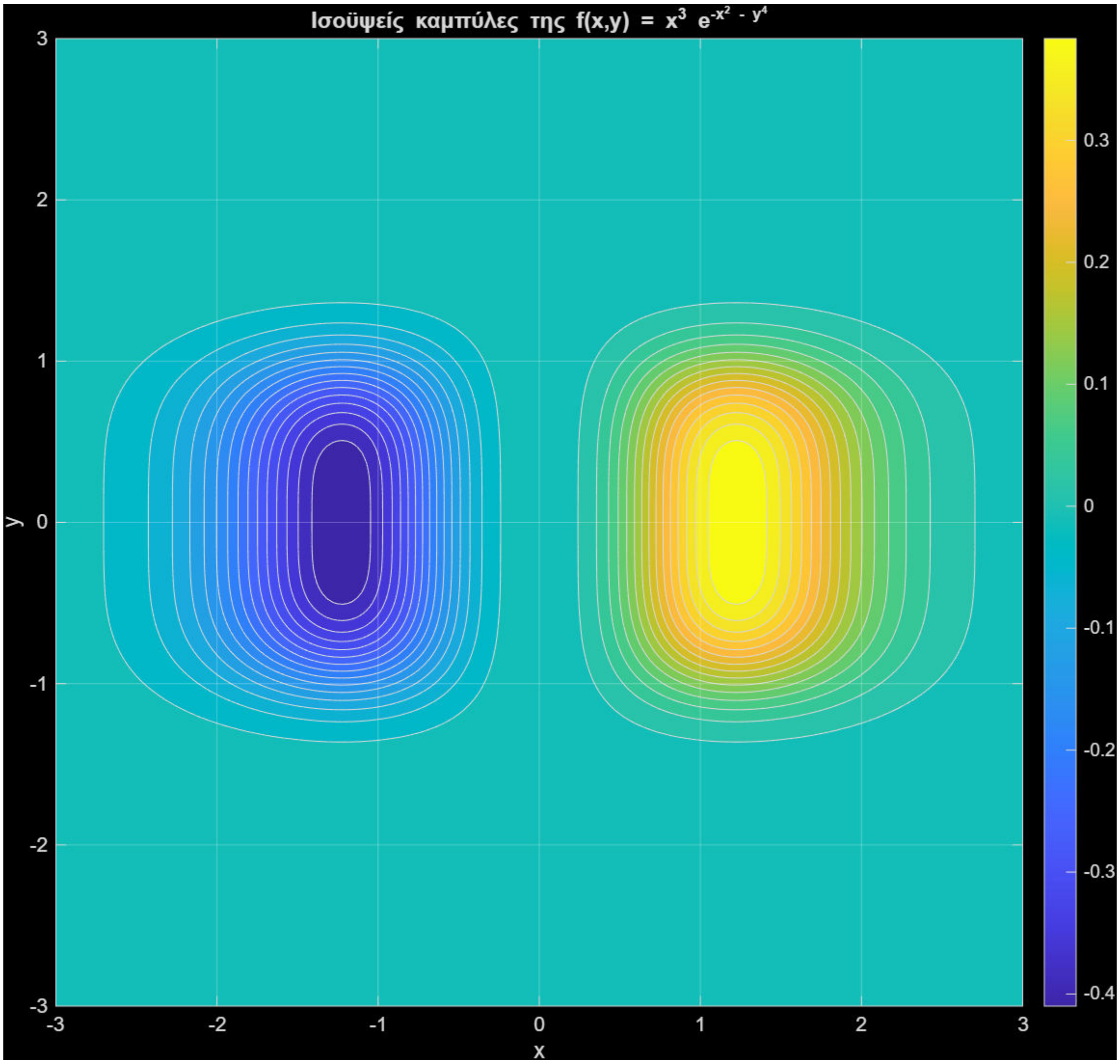
ΘΕΜΑ 1

Ξεκινάμε από την γραφική απεικόνιση της f:

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

Από την αλγεβρική της μορφή παρατηρούμε ότι η f είναι γινόμενο μιας πολυωνυμικής x^3 με έναν εκθετικό όρο $e^{-x^2-y^4}$ απόσβεσης.
Λόγω της ύπαρξης δύο μεταβλητών αναμένουμε να δούμε μια 3D απεικόνιση σε συντεταγμένες x,y,z.





Σχολιασμός:

Στην γραφική φαίνεται ότι για x>0 έχω έναν θετικό λόφο και για x<0 μια αρνητική κοιλάδα συμμετρικές ως προς y.
Μακριά από την περιοχή του λόφου και γύρο από το (-1,0) και (1,0) η επιφάνειά μας μηδενίζεται λόγω της εκθετικής απόσβεσης.
Έχουμε λοιπόν γρήγορη απόσβεση καθώς μεγαλώνουν τα $|x|,|y|$.
Οι συντεταγμένες των κορυφών της γραφικής μας, βρίσκονται αλγεβρικά με την μερική παραγώγιση ως προς x και y της f και στη συνέχεια θέτω της παράγωγο ίση με μηδέν:

Από $\frac{df}{dy}=0 \Rightarrow x=0 \acute{\eta}\ y=0$ και από $\frac{df}{dx}=0 \Rightarrow x=0 \acute{\eta}\ 2x^3-3=0 \Rightarrow x=\sqrt{\frac{3}{2}} \acute{\eta}\ x=-\sqrt{\frac{3}{2}}$

Άρα παίρνω δύο σημεία: (x*,y*)=$\left(+\sqrt{\frac{3}{2}},0\right) \approx (1.22,0)$ (x*,y*)=$\left(-\sqrt{\frac{3}{2}},0\right) \approx (-1.22,0)$

Όπου για το πρώτο έχω $f \approx 0.41$ και για το δεύτερο $f \approx -0.41$. Αυτά είναι τα κέντρα των ισοϋψών που φαίνονται στο δεύτερο διάγραμμα.

<u>ΘΕΜΑ 2</u>

Η Μέθοδος Καθόδου είναι μια μέθοδος ελαχιστοποίησης χωρίς περιορισμούς, ξεκινάμε δηλαδή από ένα αρχικό σημείο x_0 και σε κάθε βήμα κινούμαστε προς το ελάχιστο δηλαδή προς την μείωση της f. Αυτή η κατεύθυνση είναι το αρνητικό gradient.

Γενικά έχω :

$$x_{k+1} = x_k + \gamma_\kappa d_k \text{ , όπου } d_k = -\nabla f(x_k)$$

και $\gamma_k > 0$ είναι το μήκος του βήματος.

Gradient της $f(x, y)$:

Η συνάρτηση είναι:

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

Μερική παράγωγος της f:

Ως προς x: $\dfrac{df}{dx} = 3x^2 e^{-x^2 - y^{-4}} + x^3 e^{-x^2 - y^4}(-2x) = e^{-x^2 - y^4}(3x^2 - 2x^4)$

Ως προς y: $\dfrac{df}{dy} = x^3 e^{-x^2 - y^4}(-4y^3) = -4x^3 y^3 e^{-x^2 - y^4}$

Άρα το gradient είναι

$$\nabla f(x, y) = \begin{bmatrix} x^2(3 - 2x^2) e^{-x^2 - y^4} \\ -4x^3 y^3 e^{-x^2 - y^4} \end{bmatrix}$$

Η κατεύθυνση μέγιστης καθόδου στο σημείο (x_k, y_k) είναι

$$d_k = -\nabla f(x_k, y_k) = \begin{bmatrix} -x_k^2(3 - 2x_k^2) e^{-x_k^2 - y_k^4} \\ 4x_k^3 y_k^3 e^{-x_k^2 - y_k^4} \end{bmatrix}$$

Τώρα για το διάνυσμα $x_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix}$ ο γενικός τύπος της μεθόδου είναι $x_{k+1} = x_k - \gamma_\kappa \nabla f(x_k)$, δηλαδή

$$\begin{cases} x_{k+1} = x_k - \gamma_k x_k^2 (3 - 2x_k^2) e^{-x_k^2 - y_k^4} \\ y_{k+1} = y_k + \gamma_k 4 x_k^3 y_k^3 e^{-x_k^2 - y_k^4} \end{cases}$$

Όσο το γ_k είναι κατάλληλα επιλεγμένο, η $f(x_k+1)<f(x)$ και η x_k κινείται προς στάσιμο σημείο της f. Προσπαθούμε να προσεγγίσουμε, όπου αυτό είναι εφικτό ανάλογα το βήμα και το αρχικό σημείο, το στάσιμο σημείο που υπολογίσαμε στο ΘΕΜΑ 1, δηλαδή $(-\sqrt{\frac{3}{2}},0)$ - Τοπικό Ελάχιστο.

Σχόλιο: Παρατηρούμε ότι η f δεν είναι κυρτή – έχει τουλάχιστον ένα τοπικό μέγιστο και ένα τοπικό ελάχιστο και μία γραμμή στάσιμων σημείων. Η Μέθοδος Καθόδου δεν μας εγγυάται απαραίτητα ότι από οποιοδήποτε αρχικό σημείο θα βρεθούμε στο ολικό ελάχιστο που είναι το ζητούμενο.

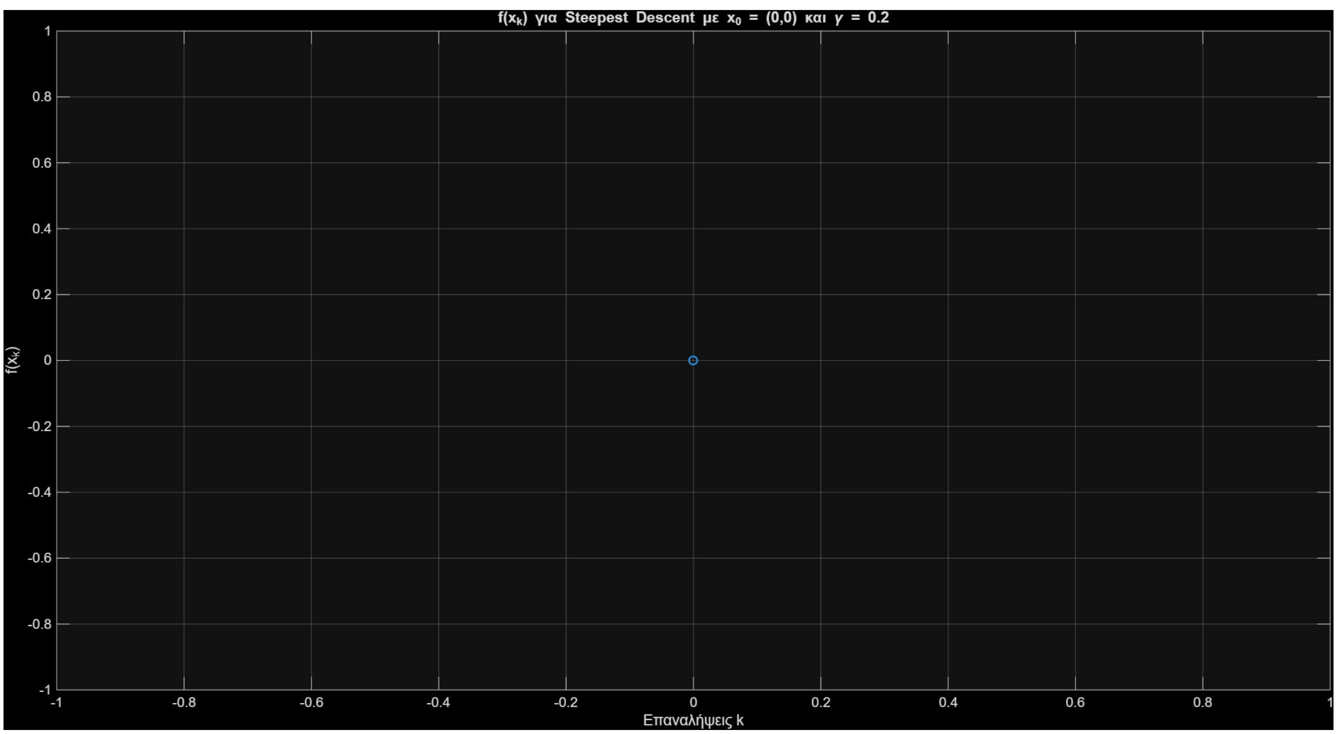
α)Επιλέγω σταθερό $\gamma_k=\gamma=0.2$**:**

i) Αρχικό σημείο το $x_0=(0,0)$

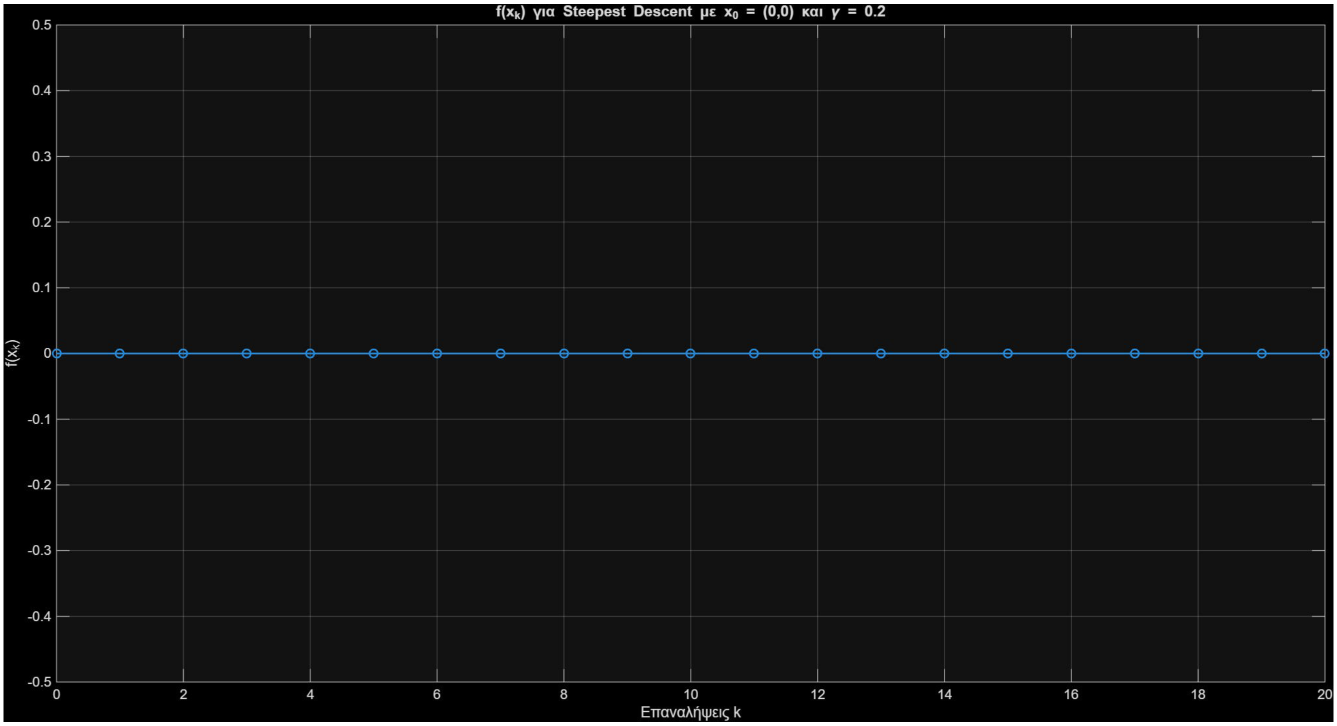
Η Μέθοδος Καθόδου είναι το $x_{k+1}=x_k+\gamma_\kappa \nabla f(x_k)$ με $f(x,y)=x^3 e^{-x^2-y^4}$.

Το gradient στο σημείο (0,0) θα είναι $\nabla f(0,0)=\begin{bmatrix} 0^2(3-0)e^0 \\ -4\,0^3 0^3 e^0 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Άρα: $x_1=x_0-\gamma \nabla f(x_0)=\begin{bmatrix} 0 \\ 0 \end{bmatrix}-\gamma\begin{bmatrix} 0 \\ 0 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Αν αυτό το κάνω για k, τότε $x_k=(0,0)\forall k$ και για οποιοδήποτε γ>0



Σχολιασμός:
Σε αυτό το διάγραμμα βλέπουμε το αναμενόμενο που υπολογίσαμε ότι δηλαδή υπάρχει μόνο ένα σημείο στο k=0 με $f(x_0)=0$. Ο αλγόριθμος τερματίζει αμέσως επειδή το gradient στο (0,0) είναι μηδενικό και δεν εκτελείται καμία επανάληψη βελτίωσης .



<u>Σχολιασμός:</u>
Εδώ βλέπουμε ξεκάθαρα την τιμή της συνάρτησης $f(x_k)=0$ για όλες τις επαναλήψεις. Η οριζόντια ευθεία στο 0 δείχνει ότι ξεκινώντας από το $x_0=(0,0)$,η Μέθοδος Καθόδου δεν πετυχαίνει καμία μείωση της αντικειμενικής συνάρτησης και ότι σε κάθε επανάληψη βρισκόμαστε στο ίδιο σημείο. Έτσι η κακή επιλογή αρχικού σημείου, κάνει την μέθοδο να μην φτάσει ποτέ στο ελάχιστο.

ii) Αρχικό σημείο το $x_0=(-1,-1)$.

Το σημείο αυτό βρίσκεται στο αριστερό βουναλάκι της συνάρτησης κοντά στο Τοπικό Ελάχιστο $(-\sqrt{\frac{3}{2}},0)$ οπότε αναμένω τον αλγόριθμο να πάει προς το συγκεκριμένο ελάχιστο.

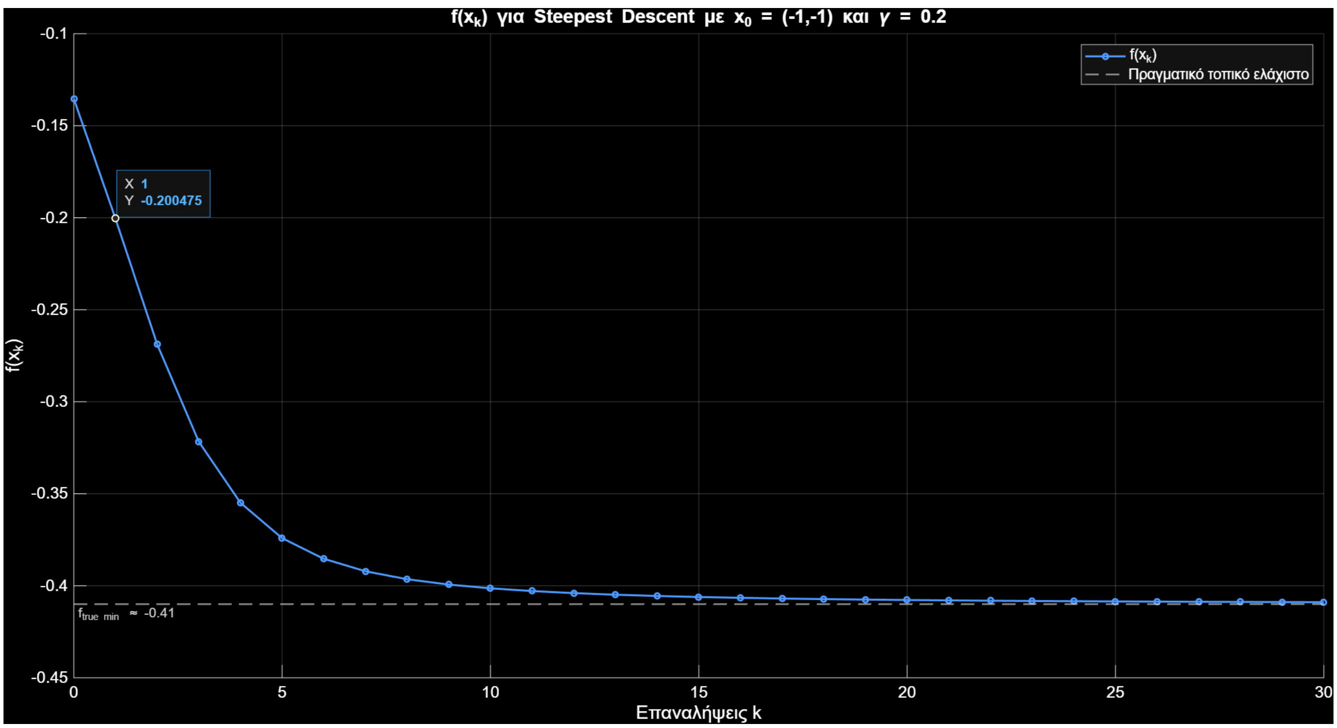
Υπολογισμός gradient $-\nabla f(-1,-1)=\begin{bmatrix} e^{-2} \\ -4e^{-2} \end{bmatrix} \approx \begin{bmatrix} -0.1353 \\ 0.5413 \end{bmatrix}$

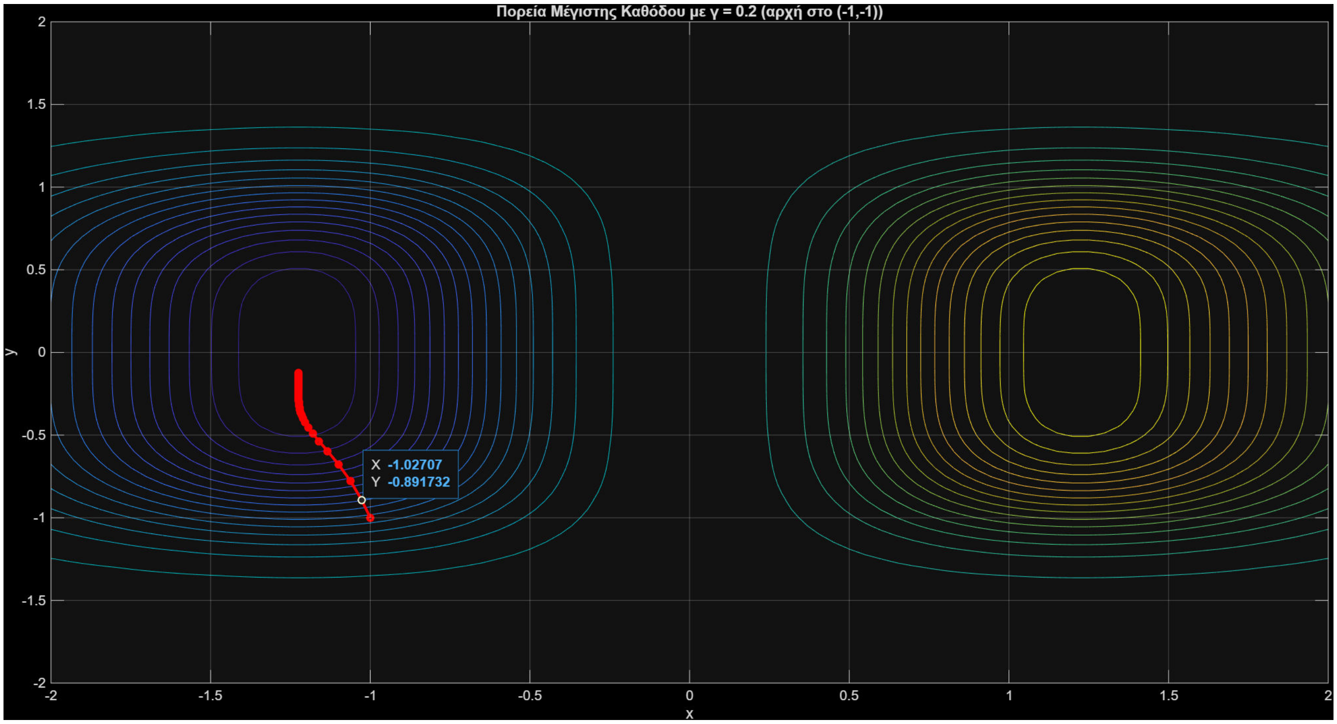
Με γ=0.1 το πρώτο βήμα είναι:

$$x_1=x_0-\gamma\nabla f(x_0)=\begin{bmatrix} -1 \\ -1 \end{bmatrix}-0.2\begin{bmatrix} 0.1353 \\ -0.5413 \end{bmatrix} \approx \begin{bmatrix} -1.0271 \\ -0.8917 \end{bmatrix}$$

Η τιμή της f στο νέο σημείο είναι $f(-1.0271,-0.8917)\approx -0.2005$

Επαναλαμβάνω την ίδια διαδικασία k επαναλήψεις μέχρι η ο αλγόριθμος να συγκλίνει και προκύπτουν τα διαγράμματα:





Σχολιασμός:
Η μπλε γραμμή δείχνει την τιμή της f σε κάθε επανάληψη
Η διακεκομμένη γραμμή δείχνει το θεωρητικό τοπικό ελάχιστο της f.
Στο k=0, $f(x_0) = -0.135$ ενώ στο τέλος των επαναλήψεων, οι τιμές συγκλίνουν στο ελάχιστο.

Τελικά η Μέθοδο Καθόδου πετυχαίνει τον σκοπό της με αρχικό σημείο το (-1,-1), συγκλίνει σωστά στο τοπικό ελάχιστο, αργά όμως με αρκετές επαναλήψεις να χρειάζονται μέχρι την σύγκλιση στο ελάχιστο (20-25 επαναλήψεις)

iii) Αρχικό σημείο $x_0 = (1,1)$ **.**

Το σημείο αυτό βρίσκεται στο δεξί βουναλάκι της συνάρτησης κοντά στο Τοπικό Μέγιστο $\left(+\sqrt{\frac{3}{2}}, 0\right)$.

Υπολογισμός gradient:

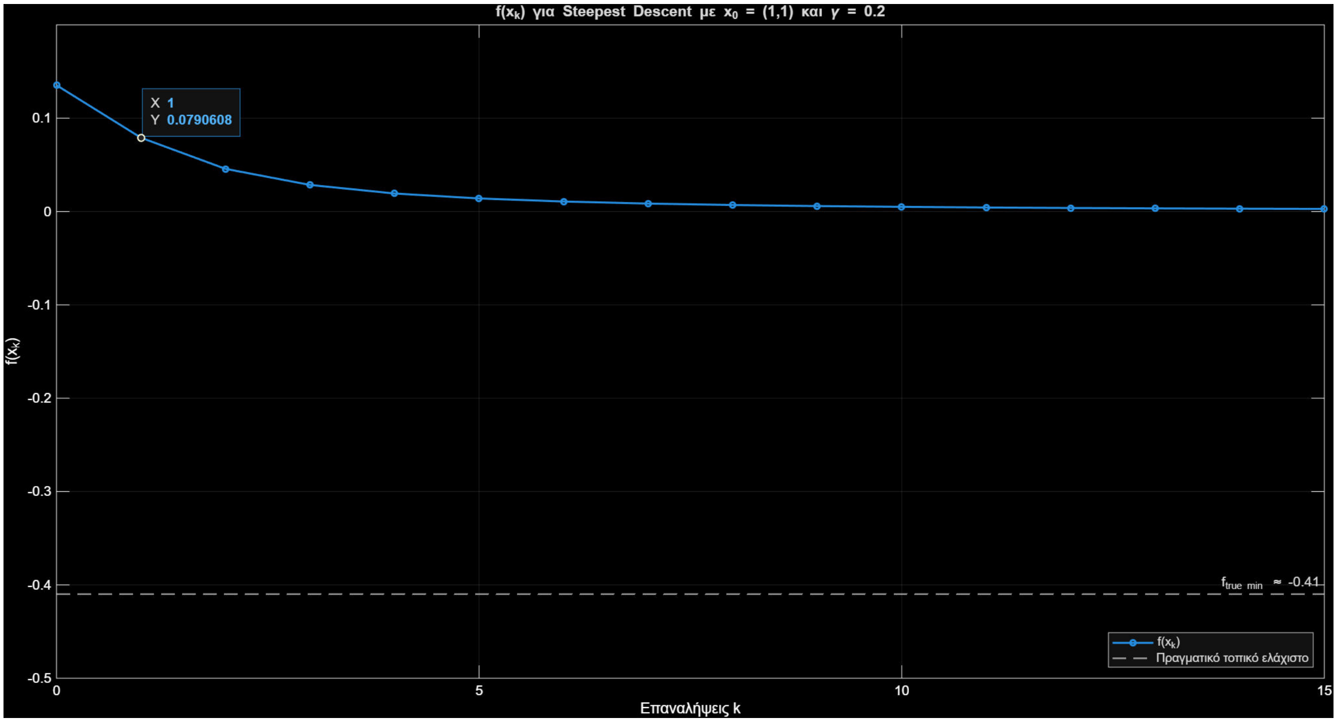
$$\nabla f(1,1) \approx \begin{bmatrix} 0.1353 \\ -0.5413 \end{bmatrix}$$

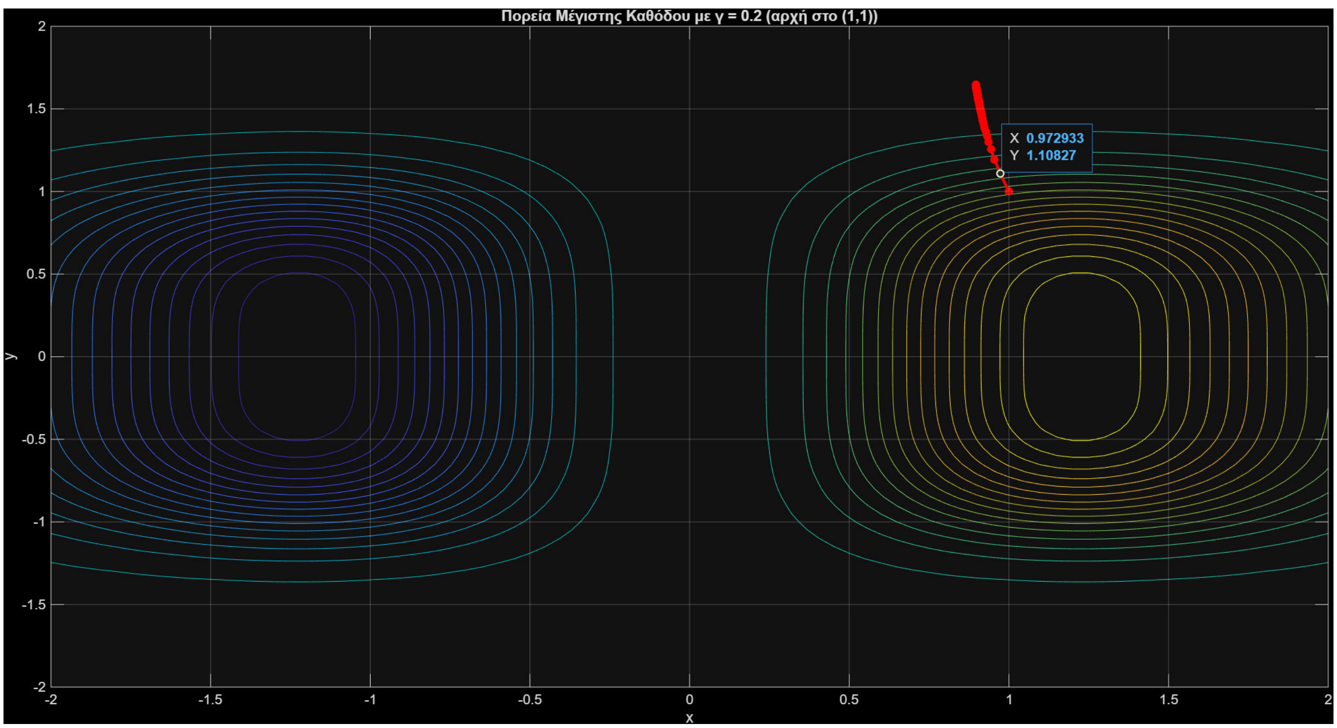
Με γ=0.1 το πρώτο βήμα δίνει:

$$x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0.2 \begin{bmatrix} 0.1353 \\ -0.5413 \end{bmatrix} = \begin{bmatrix} 0.9729 \\ 1.1083 \end{bmatrix}$$

Η τιμή της f στο πρώτο βήμα θα είναι $f(0.9729, 1.1083) = 0.079$

Επαναλαμβάνω την ίδια διαδικασία k φορές ώσπου ο αλγόριθμος να συγκλίνει σε μια τιμή όπως φαίνεται στο διάγραμμα παρακάτω:





<u>Σχολιασμός:</u>

Το διάγραμμα της f μας δείχνει την ανικανότητα του συστήματος να συγκλίνει στο τοπικό ελάχιστο. Αντίθετα συγκλίνει στο 0.

Ακόμη και για διαφορετικά γ (πχ 0.3) η Μέθοδος Καθόδου μειώνει την f αλλά συγκλίνει κοντά στο 0 και όχι στο πραγματικό τοπικό ελάχιστο. Αυτό συμβαίνει επειδή το σημείο (1,1) βρίσκεται στο δεξιό

θετικό λοβό της συνάρτησης και ο αλγόριθμος οδηγείται προς τη γραμμή στάσιμων σημείων x=0-επιφάνεια όπου f=0 χωρίς ποτέ να περάσει στην αριστερή αρνητική λεκάνη όπου βρίσκεται το ελάχιστο.
Επομένως γιαυτό το αρχικό σημείο, η μέθοδος δεν μπορεί να βρει το επιθυμητό ελάχιστο και βαλτώνει σε στάσιμο σημείο.

β)Επιλέγω γ_k τέτοιο ώστε να ελαχιστοποιείται η $f(x_k+\gamma_k d_k)$

i) Με αρχικό σημείο το $x_0=(0,0)$ δεν μπορούμε να βρούμε κάποιο γ ώστε να προκαλέσει τελικά την ελαχιστοποίηση, επειδή το gradient μηδενίζεται και η κατεύθυνση καθόδου γίνεται $d_0=0$. Έτσι ο αλγόριθμος εγκλωβίζεται σε ένα στάσιμο σημείο.

ii) Αρχικό σημείο $x_0=(-1,-1)$

έχω gradient της f να είναι:

$$\nabla f(-1,-1)=\begin{bmatrix} e^{-2} \\ -4e^{-2} \end{bmatrix} \text{ και } d_0=-\nabla f(-1,-1)=\begin{bmatrix} -e^{-2} \\ 4e^{-2} \end{bmatrix} \text{ η διεύθυνση καθόδου.}$$

Για να κάνω το σύστημα να συγκλίνει θα το εκπαιδεύσω μέσα από exact line search όπου το γ_k δεν θα είναι σταθερό αλλά σε κάθε επανάληψη θα αποκτά μια τιμή έτσι ώστε στο τέλος να ελαχιστοποιεί την f κατά μήκος της διεύθυνσης καθόδου.

Για γνωστό λοιπόν x_k και $d_k=-\nabla f(x_k)$ ορίζω τη συνάρτηση:

$$\varphi_k(\gamma)=f(x_k+\gamma d_k) \text{ με } \gamma>0.$$

Τότε το γ_k ορίζεται θεωρητικά ως:

$$\gamma_k=arg\,min\,\varphi_k(\gamma)=arg\,min\,f(x_k+\gamma d_k), \text{ με } \gamma>0$$

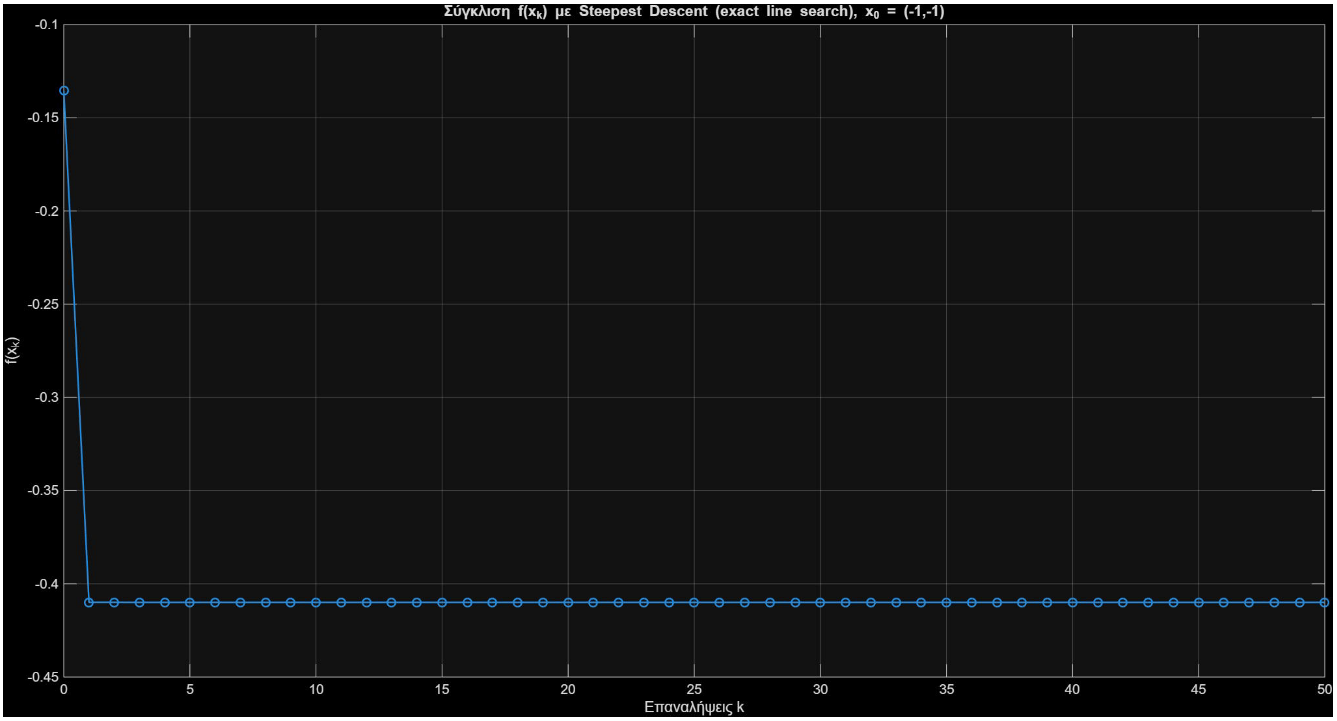
Διαλέγουμε δηλαδή το γ_k που δίνει τη μικρότερη δυνατή τιμή της f πάνω στην γραμμή που ορίζεται από το σημείο x_k και τη διεύθυνση d_k.

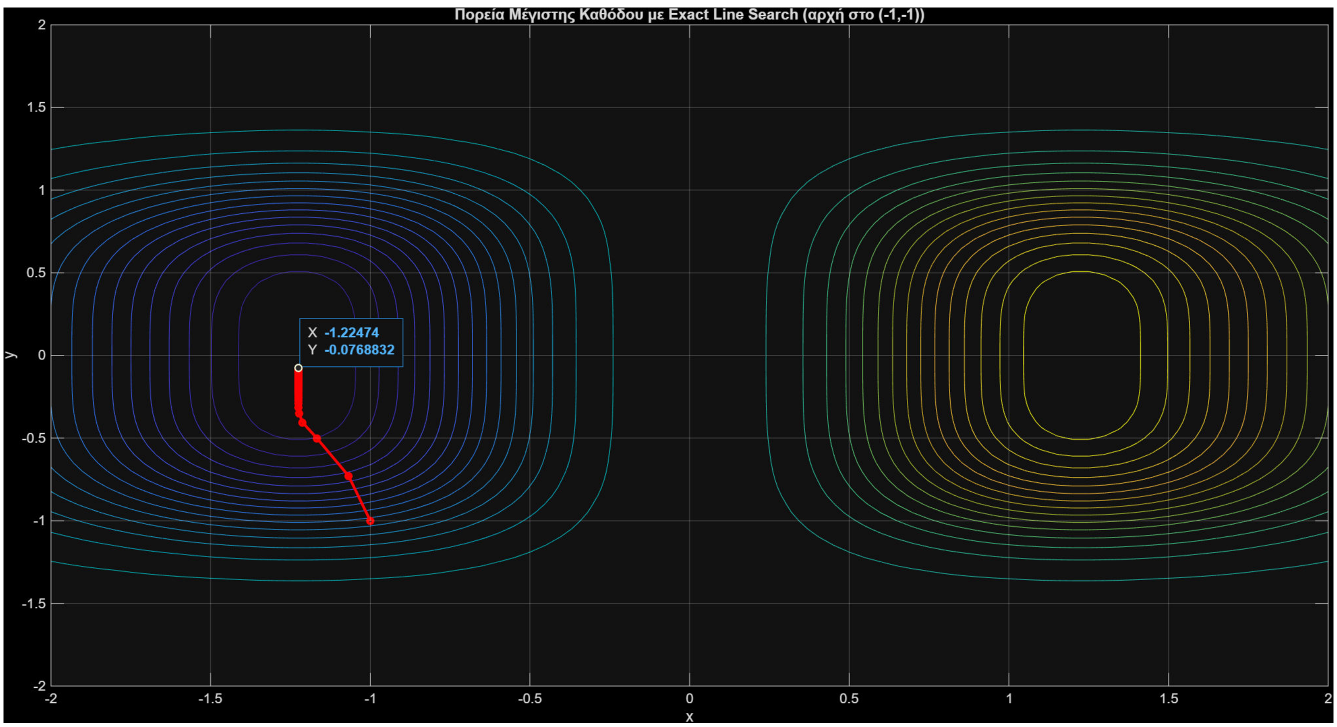
Στον κώδικα MATLAB, η εκπαίδευση του γ_k γίνεται με την ελαχιστοποίηση του $\varphi_k(\gamma)$ σε κάθε επανάληψη με την εντολή : phi = @(gamma) f(xk + gamma * dk);

Στη συνέχεια καλώ τον ελαχιστοποιητή fminbnd στο διάστημα [0,5].

Οπότε ο αλγόριθμος θα είναι για k=0,1,…. :

$$\begin{gathered} \gamma_k=\nabla f(x_k) \\ d_k=-\gamma_k \\ \gamma_k=arg\,min\,f(x_k+\gamma d_k) \\ x_{k+1}=x_k+\gamma_k d_k \end{gathered}$$





<u>Σχολιασμός</u>:
Στο k=0: $f(x_0) \approx -0.135$
Στο πρώτο βήμα (k=1) πέφτει στο -0.41 δηλαδή στο τοπικό ελάχιστο που ζητάμε.

Από εκεί και πέρα η καμπύλη είναι σχεδόν οριζόντια, οι επόμενες επαναλήψεις κάνουν μικρές βελτιώσεις γύρω από την ίδια τιμή.

Τελικά με αρχικό σημείο το $x_0=(-1,-1)$ με την διαδικασία εκπαίδευσης που επιλέξαμε ο αλγόριθμος οδηγείται πολύ γρήγορα στο τοπικό ελάχιστο του πρώτου λοβού (-0.41). Το πρώτο βήμα βελτιώνει έντονα την f και στη συνέχεια βελτιώνεται ελάχιστα γύρω από το σημείο . Έτσι η επιλογή μη σταθερού γ_k κάνει την σύγκλιση πιο αποδοτική.

iii) Αρχικό σημείο το $x_0=(1,1)$

Το gradient στο x_0:

$$\nabla f(1,1)=\begin{bmatrix} e^{-2} \\ -4e^{-2} \end{bmatrix} \text{ και το } d_0=-\nabla f(1,1)=\begin{bmatrix} -e^{-2} \\ 4e^{-2} \end{bmatrix}$$

Αυτό σημαίνει ότι το x μειώνεται (κινείται προς το 0 και μετά αρνητικός) και το y αυξάνεται (πάει προς τα θετικά.

Γραμμή αναζήτησης $\varphi_0(\gamma)$:

Το επόμενο σημείο είναι:

$$x_1(\gamma)=x_0+\gamma d_0=\begin{bmatrix} 1-0.1353\gamma \\ 1+0.5413\gamma \end{bmatrix}$$

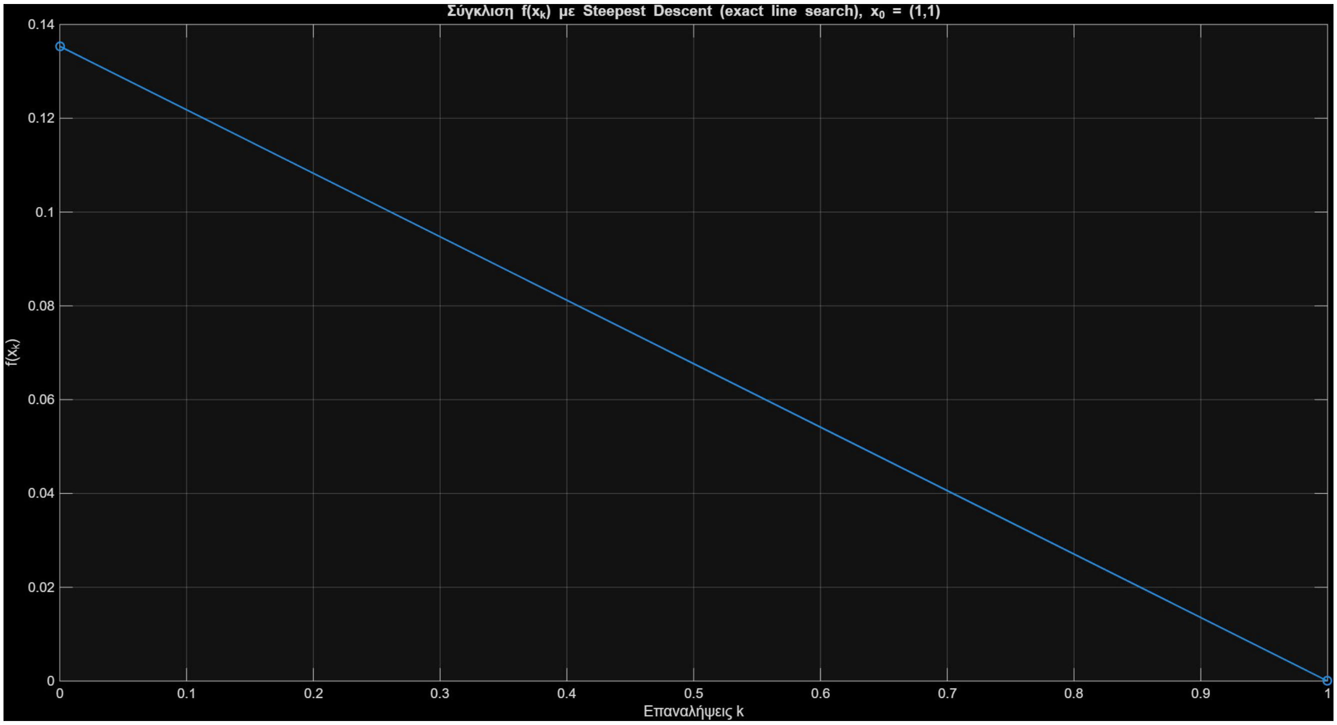
Ορίζω συνάρτηση :

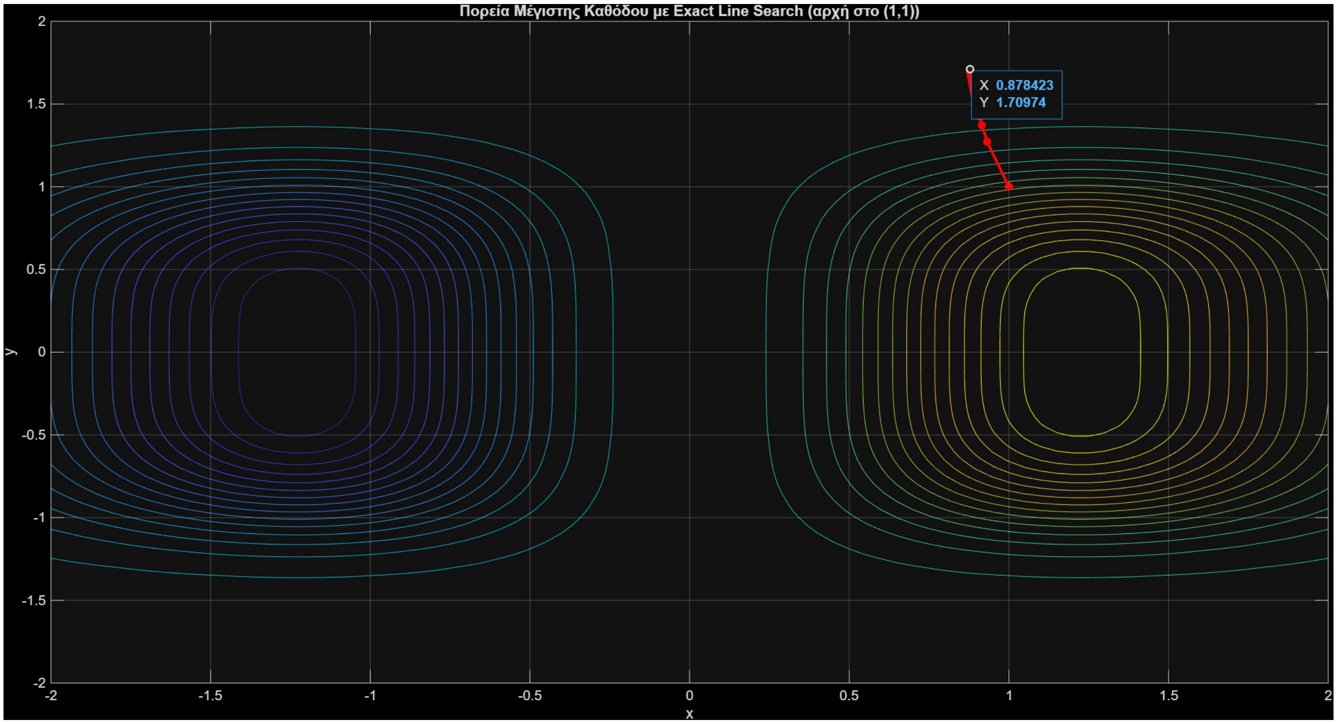
$$\varphi_0(\gamma)=f(x_1(\gamma))=f(1-0.1353\gamma, 1+0.5413\gamma), \gamma>0$$

Η μέθοδος εκπαίδευσης επιλέγει

$$\gamma_0=arg\,min\,\varphi_0(\gamma), \gamma>0$$

Παρατηρούμε ότι για μικρές τιμές του γ, η f πέφτει πέφτει από $f(1,1)\approx 0.135$ σε μικρότερες θετικές τιμές.
Όσο αυξάνει η γ η $f(\gamma)\rightarrow 0$ από την θετική πλευρά της συνάρτησης.
Η $\varphi_0(\gamma)$ κατά μήκος της διεύθυνσης δεν βλέπει το ελάχιστο 0.41 αλλά όσο μεγαλώνει το γ βλέπει το 0.



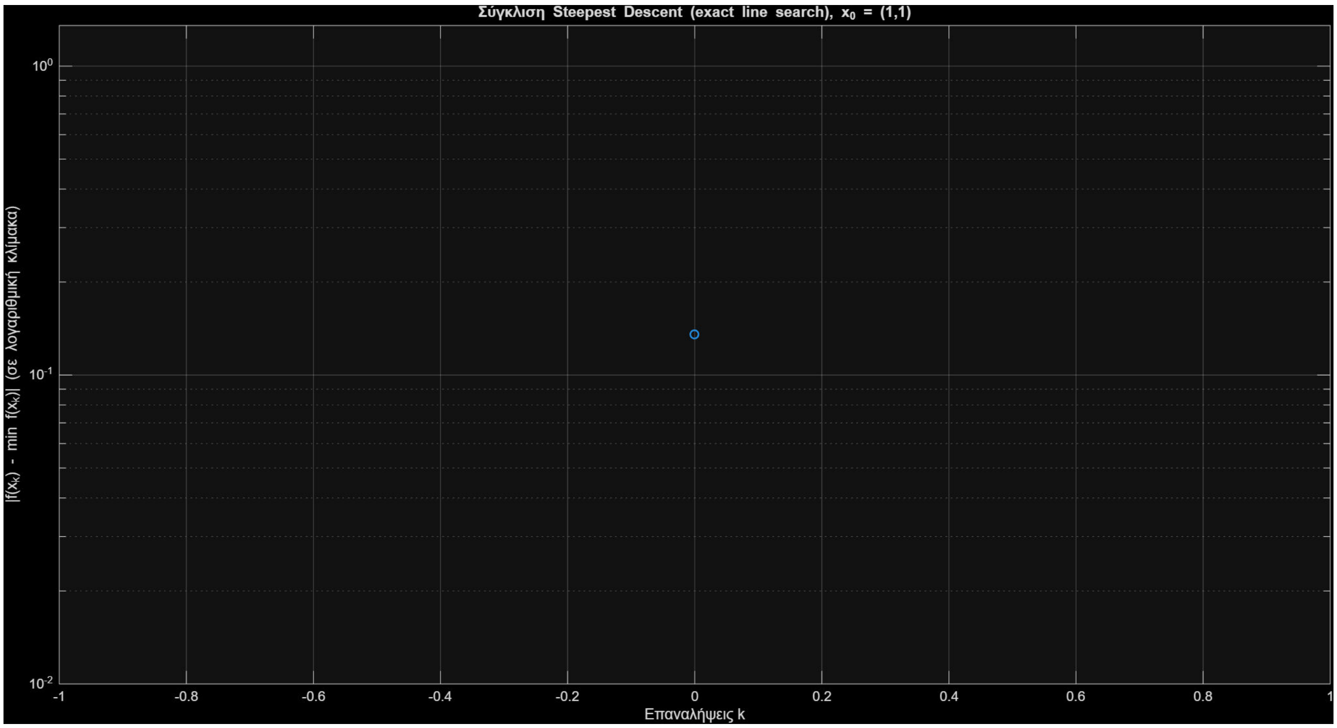


<u>Σχολιασμός:</u>
Στο k=0 βρισκόμαστε στο αρχικό σημείο $x_0=(1,1)$ με $f(x_0)\approx 0.135$

Στο k=1 έχει γίνει ένα και μόνο βήμα Steepest Descent με exact line search. Το line search βρίσκει ένα μεγάλο γ_0 που ελαχιστοποιεί την $f(x_0+\gamma d_0)$ στην διεύθυνση καθόδου . Ο αλγόριθμος έτσι πηγαίνει σε σημείο όπου $f(x_1)=0$. Στο γράφημα φαίνεται η ευθεία πτώσης από 0.135 στο 0.

Μετά το πρώτο βήμα ο αλγόριθμος σταματά διότι το gradient στο νέο σημείο είναι πολύ μικρό καθώς βρισκόμαστε στην κοιλάδα γύρω από τον άξονα x=0, όπου $f(x,y)=0$ και $\|\nabla f(x_k)\|=0$ άρα τερματίζει η μέθοδος.

Τελικά ναι μεν η $f(x_k)$ μειώνεται (καθοδικό γράφημα), αλλά η μέθοδος δεν συγκλίνει στο τοπικό ελάχιστο καθώς εγκλωβίζεται σε στάσιμο σημείο της κοιλάδας x=0, και από εκεί και πέρα ο αλγόριθμος της μέγιστης καθόδους δεν την βελτιώνει.



<u>Σχολιασμός:</u>
Η $min f(x_k)$ είναι η καλύτερη τιμή που μπορεί να πετύχει ο αλγόριθμος σε όλες τις επαναλήψεις.

Για το $x_0=(1,1)$ με exact line search $f(x_0)\approx 0.135$ και στο επόμενο βήμα x_1 έχω $f(x_1)\approx 0$.

Άρα το σφάλμα είναι για k=0: $|f(x_0)-min f|\approx|0.135-0|=0.135$
ενώ για k=1 : $|f(x_1)-min f|\approx|0-0|=0$

Γιαυτό βλέπουμε μόνο ένα σημείο αυτό του k=0.

Αυτό μας δείχνει ότι σε μία επανάληψη το σφάλμα πέφτει από 0.135 στο 0 δηλαδή ο αλγόριθμος βρίσκει αμέσως τη δική του καλύτερη τιμή που είναι το f=0 και σταματά. Τελειώνει σε λάθος στόχο αλλά τουλάχιστον τελειώνει!

γ) το βήμα γ_k θα επιλέγει βάσει του κανόνα Armijo

Στη μέθοδο Μέγιστης Καθόδου έχω

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

Ο κανόνας Armijo διαλέγει με backtracking έτσι ώστε να ικανοποιείται η ανισότητα:(5.2.39 βιβλίο)

$$f(x_k + \gamma_k d_k) \leq f(x_k) + \sigma \gamma_k \nabla f(x_k)^T d_k \text{, όπου } 0 < \sigma < 1$$

<u>Πρακτικά:</u>

Ξεκινάμε με ένα αρχικό βήμα $\gamma = \gamma_0$
Αν δεν ισχύει η παραπάνω ανισότητα Armijo, δηλαδή

$$f(x_k + \gamma_k d_k) > f(x_k) + \sigma \gamma_k \nabla f(x_k)^T d_k,$$

τότε μειώνω το γ
Επαναλαμβάνω μέχρι να ικανοποιηθεί η ανισότητα
Το τελικό γ_k είναι το βήμα της επανάληψης
Έτσι εξασφαλίζεται ότι κάθε βήμα δίνει επαρκή μείωση της f.

i) Αρχικό σημείο $x_0 = (0,0)$

Το gradient $\nabla f(0,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ και συνεπώς $d_0 = -\nabla f(0,0) = (0,0)$
Η διεύθυνση καθόδου είναι μηδενική . Η Armijo κοιτάει για σημεία της μορφής $x_0 + \gamma d_0$ τα οποία εδώ είναι (0,0) για κάθε γ.

Δεξί μέλος : $f(x_0) + \sigma \gamma \nabla \varphi(x_0)^T d_0 = f(0,0) + \sigma \gamma 0 = f(0,0)$

Αριστερό μέλος: $f(x_0 + \gamma d_0) = f(0,0)$

Άρα η ανισότητα Armijo γίνεται $f(0,0) \leq f(0,0)$ που ισχύει για οποιοδήποτε γ.

Συμπέρασμα: Ο κανόνας Armijo αποδέχεται οποιαδήποτε τιμή για το γ. Αλλά επειδή η διεύθυνση d_0 είναι μηδέν το επόμενο σημείο είναι και αυτό μηδέν

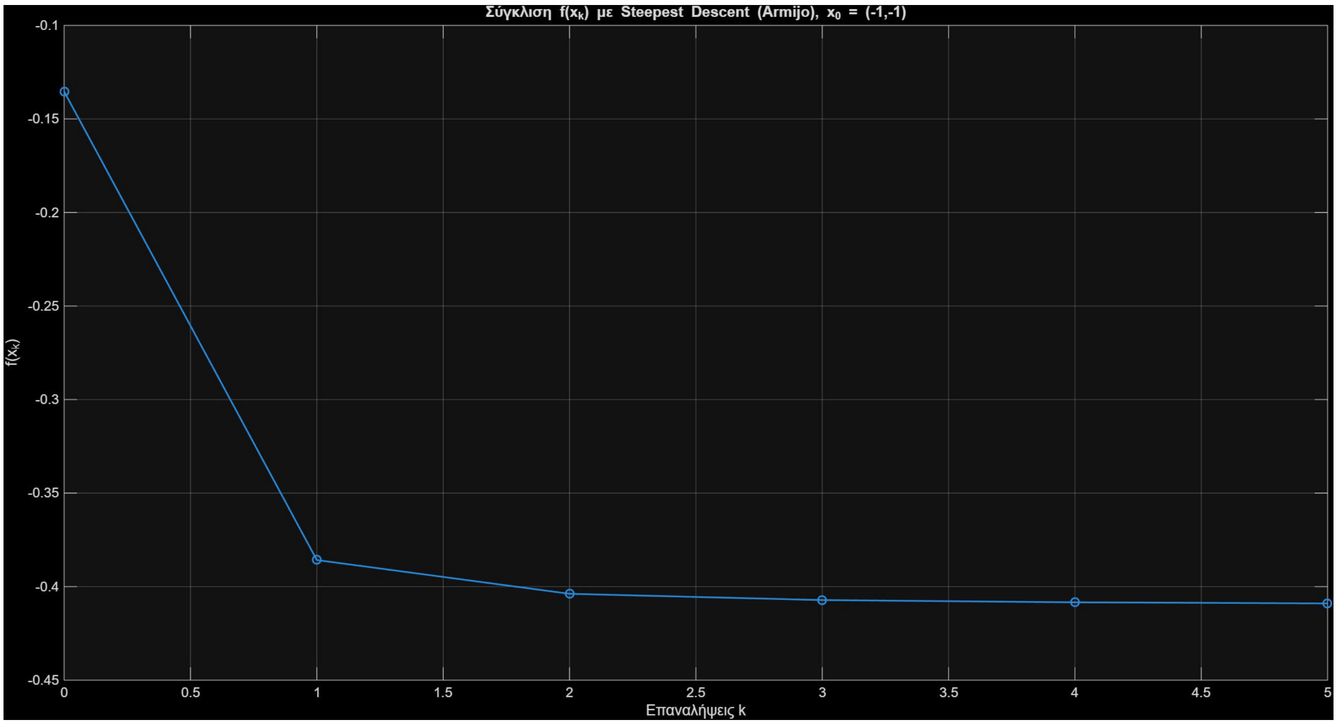
$$x_1 = x_0 + \gamma_0 d_0 = (0,0)$$

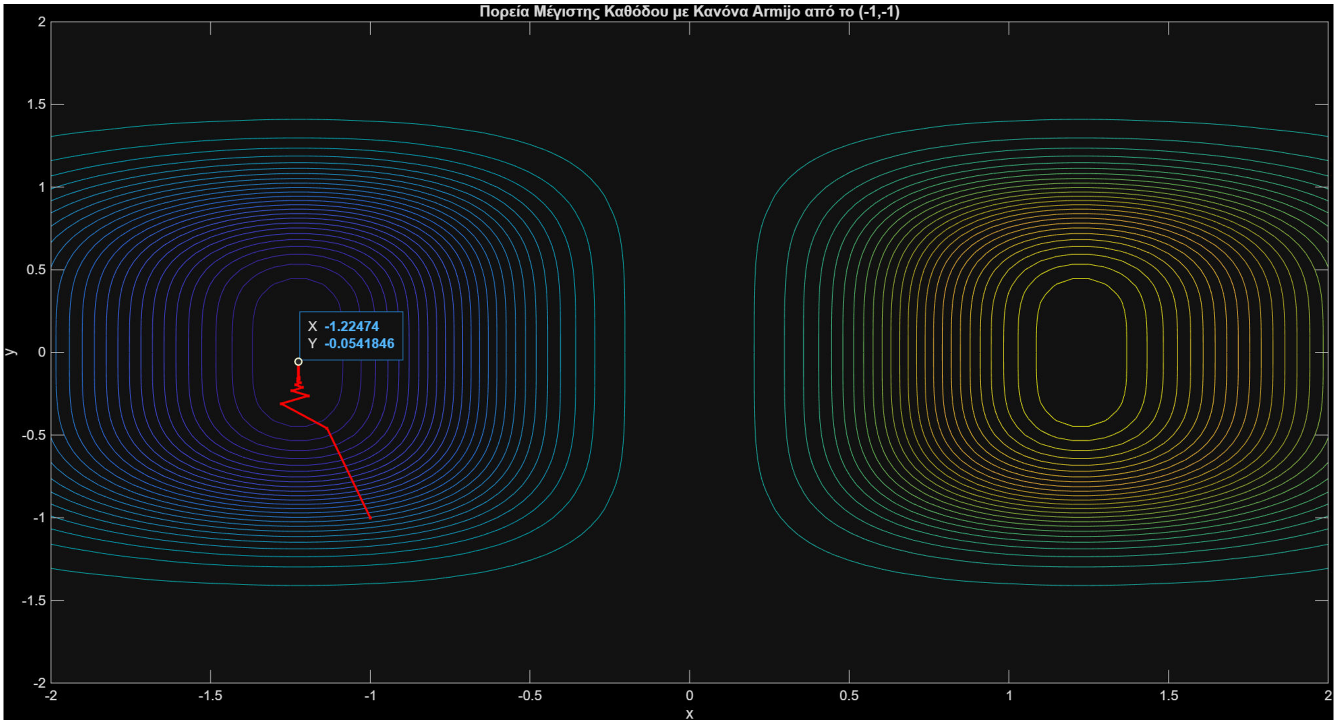
Άρα η Μέθοδος Καθόδου με αρχικό σημείο $x_0 = (0,0)$ και μέθοδο Armijo μένει κολλημένη στο (0,0) το οποίο είναι στάσιμο σημείο της f(x,y) με τιμή μηδέν αλλά δεν είναι το τοπικό ελάχιστο της συνάρτησης $(f_{min} \approx -0.41)$. Η αποτυχία σύγκλισης δεν οφείλεται στον κανόνα Armijo αλλά στο γεγονός ότι το αρχικό σημείο έχει μηδενικό gradient.

ii) Αρχικό σημείο $x_0 = (-1, -1)$

διεύθυνση καθόδου $d_0 = -\nabla f(-1, -1) = \begin{bmatrix} -0.135 \\ 0.541 \end{bmatrix}$ δηλαδή το πρώτο βήμα πηγαίνει τον x πιο αρνητικό (προς τα αριστερά) τον y προς το 0 (από το -1). Δηλαδή είμαστε μέσα στην λεκάνη του αριστερού τοπικού ελαχίστου.

Ο κανόνας Armijo εγγυάται ότι θα παίρνουμε καθοδικά βήματα.





<u>Σχολιασμός:</u>

Στο k=0 ξεκινάμε από $f(x_0) \approx -0.135$

Στο k=1 βλέπουμε μια τεράστια πτώση και η τιμή να πάει στο -0.38 Αυτό σημαίνει ότι ο κανόνας Armijo επέλεξε ένα σχετικά μεγάλο γ_0 και το φέρνει πολύ κοντά στο τοπικό ελάχιστο.

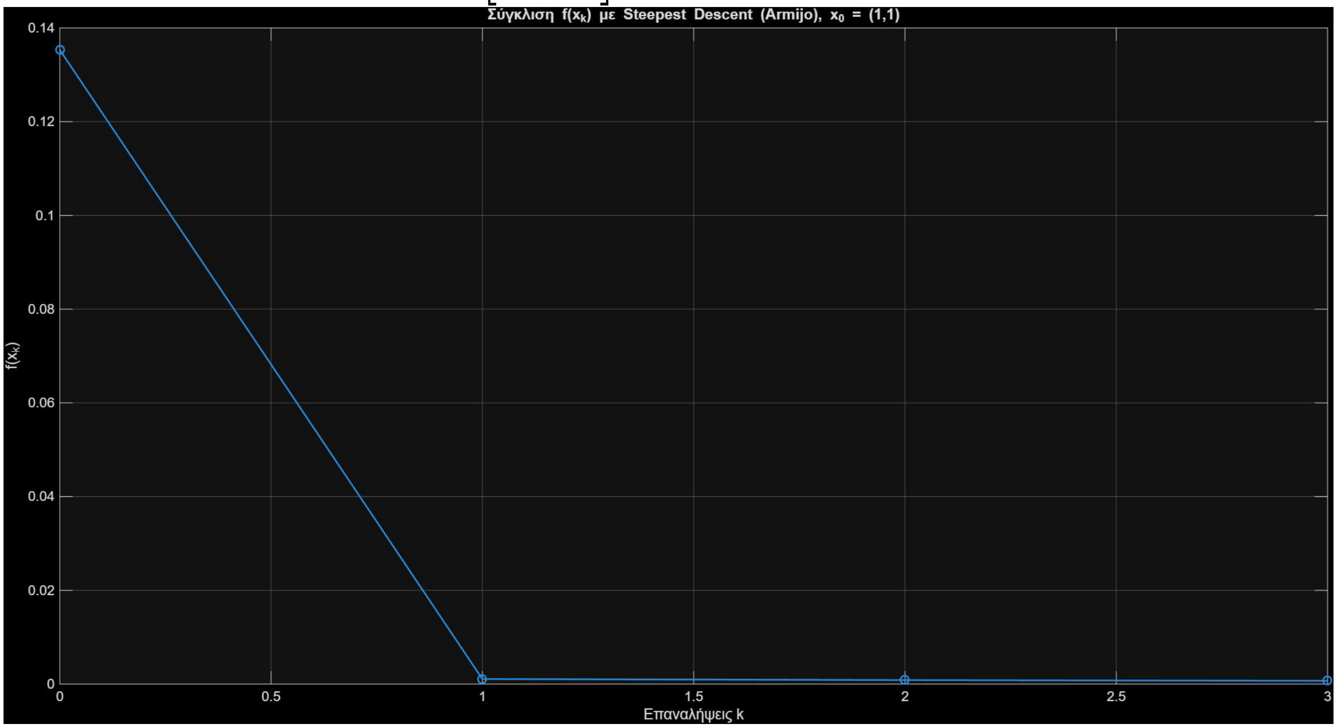
Στις επόμενες επαναλήψεις βελτιώνεται και ουσιαστικά ισορροπεί στο τοπικό ελάχιστο.

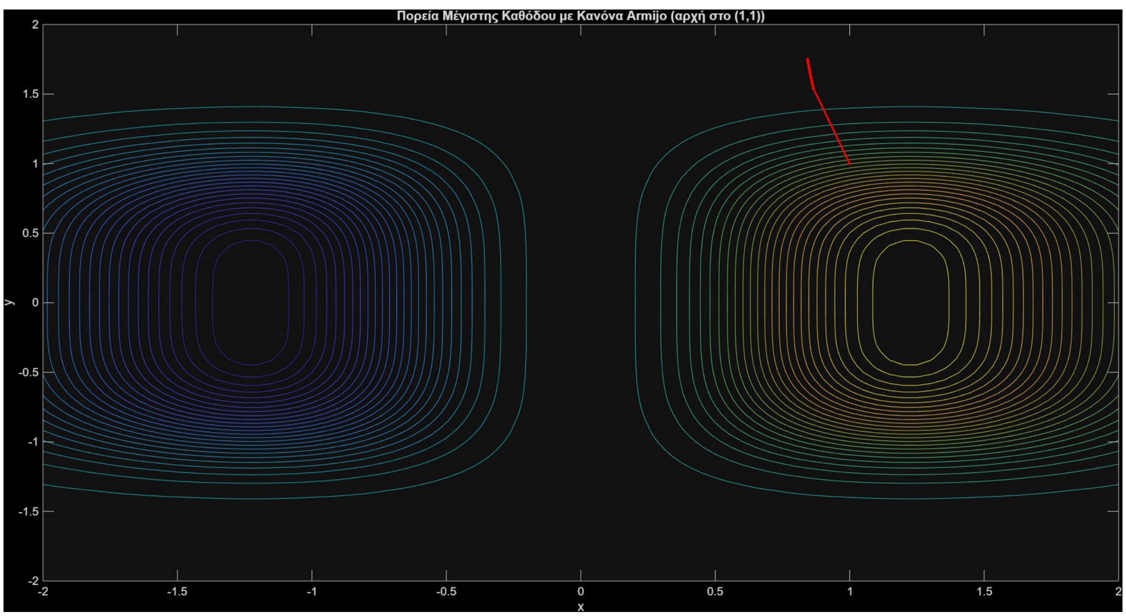
Σε σχέση με το σταθερό γ εδώ ο προσαρμοστικός Armijo διαλέγει έξυπνα τα βήματα: μεγάλο στην αρχή για να κάνει το μεγάλο άλμα, και μετά μικρότερα για να ισορροπίσει στο -0.41.

Από το (-1,-1) με Μέθοδο Καθόδου και Armijo έχω μείωση της f και πολύ γρήγορη προσέγγιση του τοπικού ελαχίστου.

iii) Αρχικό σημείο $x_0=(1,1)$

διεύθυνση καθόδου $d_0=-\nabla f(1,1)\approx\begin{bmatrix}-0.135\\0.541\end{bmatrix}$ όμοια με πρίν.





<u>Σχολιασμός</u>:
Στο k=0 ξεκινάμε από το $x_0=(1,1)$ με $f(x_0)\approx 0.135$

Στο k=1 βρίσκει η Armijo ένα βήμα που ρίχνει την τιμή σχεδόν στο 0.

Από εκεί και μετά η καμπύλη $f(x_k)$ μένει στην επίπεδη επιφάνεια 0. Δηλαδή ο αλγόριθμος μπαίνει γρήγορα σε μια επίπεδη κοιλάδα γύρω από τον άξονα x=0 όπου η f(x,y)=0 και το gradient μηδενίζει οπότε ικανοποιείται το κριτήριο στάσης.
Δεν φτάνει στο τοπικό ελάχιστο $f_{min} \approx -0.41$ γιατί από το αρχικό σημείο (1,1) και κατά μήκος των διευθύνσεων καθόδου η συνάρτηση βλέπει ως καλύτερη περιοχή την κοιλάδα f=0 και όχι την λεκάνη ελαχίστου στην αριστερή της πλευρά έτσι σταματά εσφαλμένα.

<u>ΘΕΜΑ 2</u>

Κανόνας Newton:

$$x_{k+1} = x_k + \gamma\, p_k,\ \text{ με } p_k = H(x_k)^{-1} \nabla f(x_k)$$

όπου $\nabla f(x_k)$ είναι το gradient, $H(x_k)$ είναι ο Hessian (πίνακας δεύτερων παραγώγων) και γ>0 σταθερό

κοντά σε ένα σημείο x_k προσεγγίζουμε με (5.2.17 βιβλίο)

$$f(x) \approx f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T H(x_k)(x - x_k)$$

ελαχιστοποιούμε ως προς x το παραπάνω και παίρνουμε την διεύθυνση Newton p_k.
Αν βρισκόμαστε κοντά σε ένα τοπικό ελάχιστο και ο Hessian εκεί είναι θετικός και αντιστρέψιμος η μέθοδος έχει πολύ γρήγορη σύγκλιση.

<u>Παρατήρηση:</u> Ο αλγόριθμος λειτουργεί σωστά μόνο όταν το gradient είναι διάφορο του μηδενός και όταν ο Hessian είναι αντιστρέψιμος αυτό θα το παρατηρούμε έντονα όταν αναπτύξουμε τον αλγόριθμο με αρχικό σημείο το $x_0 = (0,0)$.

Θυμίζω:

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

Gradient:

$$\nabla f(x, y) = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} x^2(3 - 2x^2) e^{-x^2 - y^4} \\ -4x^3 y^3 e^{-x^2 - y^4} \end{bmatrix}$$

Hessian:

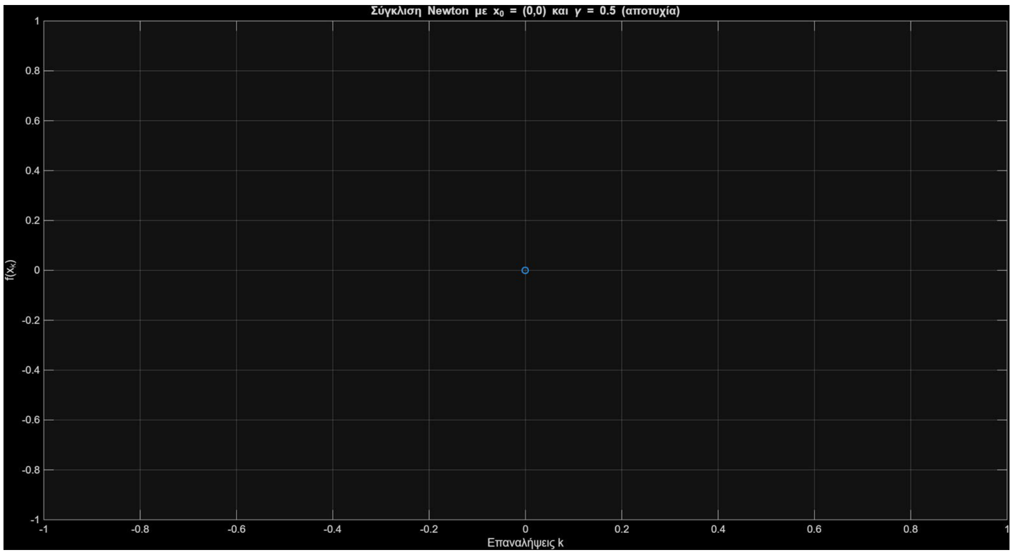
$$H(x, y) = e^{-(x^2 + y^4)} \begin{bmatrix} 4x^5 - 14x^3 + 6x & x^2 y^3 (8x^2 - 12) \\ x^2 y^3 (8x^2 - 12) & x^3 y^2 (16y^4 - 12) \end{bmatrix}$$

α) Σταθερό βήμα $\gamma_k = \gamma = 0.5$

i) Αρχικό Σημείο $x_0=(0,0)$

gradient $\nabla f(0,0)=\begin{bmatrix}0\\0\end{bmatrix}$ και Hessian $H(0,0)=\begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$ μηδενικός και μη αντριστρέψιμος.
Newton $p_0=-H(0,0)^{-1}\nabla f(0,0)$ δεν ορίζεται αφού δεν υπάρχει η H^{-1} και το gradient είναι ήδη μηδέν άρα ο αλγόριθμος σταματά αμέσως στο (0,0) χωρίς να γίνει βήμα.

Η τιμή της f(0,0)=0 που δεν είναι το τοπικό ελάχιστο.



Σχολιασμός:

Το διάγραμμα δείχνει μόνο ένα σημείο στο k=0 με τιμή $f(x_0)=0$ και καμία εξέλιξη όπως περιμέναμε. Η μέθοδος Newton βλέπει μηδενικό gradient και δεν μπορεί να υπολογίσει την διεύθυνση p_k οπότε σταματά αμέσως.
Η ακολουθία των x_k μένει στο (0,0) και η $f(x_k)$ είναι συνέχεια στο μηδέν χωρίς καμία προσέγγιση προς το ελάχιστο.
ii) Αρχικό Σημείο $x_0=(-1,-1)$

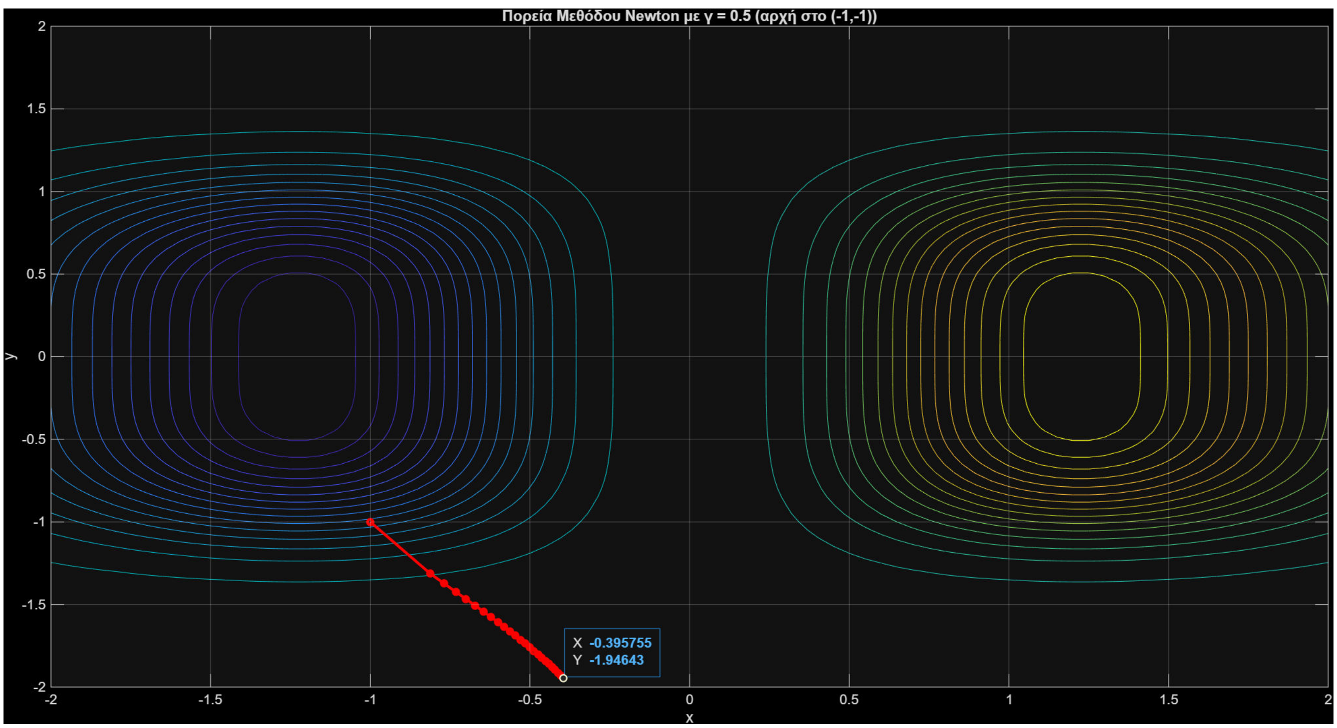
gradient $\nabla f(-1,-1)=\begin{bmatrix}e^{-2}\\-4e^{-2}\end{bmatrix}$, η τιμή της f στο αρχικό σημείο $f(-1,-1)\approx-0.135$ και Hessian $H(-1,-1)=e^{-2}\begin{bmatrix}4 & 4\\4 & -4\end{bmatrix}$ οι ιδιοτιμές του είναι $\pm4e^{-2}\sqrt{2}$ άρα ο Hessian είναι αόριστος μη θετικά ορισμένος και ο Newton δεν εγγυάται βήμα καθόδου.

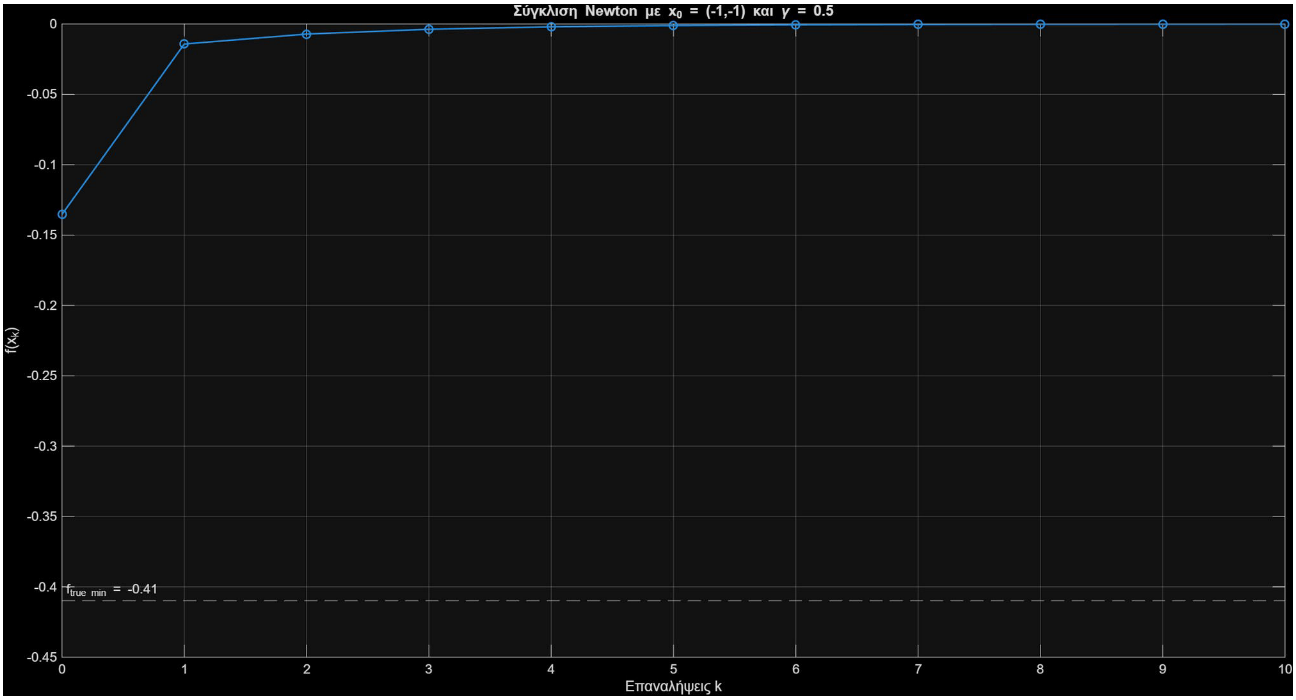
Το πρώτο βήμα Newton είναι

$$p_0=\begin{bmatrix}\frac{3}{8}\\-\frac{5}{8}\end{bmatrix} \text{ και } x_1=x_0+\gamma\, p_0\approx(-0.8125,-1.3125)$$

και οι τιμές της συνάρτησης $f(-1,-1) \approx -0.1353$ και $f(-0.8125,-1.3125) = -0.0143$

Δηλαδή το πρώτο βήμα με γ=0.5 μετακινεί το σημείο προς τα δεξιά το x μεγαλώνει από -1 σε -0,8125 και το y πάει πιο κάτω από -1 σε -1,3125. Κάνει δηλαδή την f λιγότερο αρνητική από -0.135 σε -0.014 και δεν πλησιάζει το τοπικό ελάχιστο -0.41.
Αυτό συμβαίνει προφανώς λόγω της Hessian καθώς η διεύθυνση p_0 δεν είναι διεύθυνση καθόδου και όπως φαίνεται από το πρώτο βήμα απομακρύνει την λύση μας από την λεκάνη του ελαχίστου.





<u>Σχολιασμός:</u>

Στο k=0: $f(x_0) \approx -0.135$

Στα επόμενα βήματα η τιμή ανεβαίνει: $-0.135 \rightarrow -0.02 \rightarrow -0.01 \rightarrow \ldots \rightarrow 0$

Η διακεκομμένη γραμμή είναι το πραγματικό τοπικό ελάχιστο της συνάρτησης $f_{min} \approx -0.41$ στο οποίο θα θέλαμε να συγκλίναμε αλλά η καμπύλη Newton απομακρύνεται και συγκλίνει στο 0 όπου σταθεροποιείται.

<u>Σύγκριση Μέγιστης Καθόδου – Newton με σταθερό γ από το (-1,-1):</u>

Η μέγιστη κάθοδος δουλεύει σωστά με κατάλληλο γ, είναι πάντα σε διεύθυνση καθόδου και συγκλίνει στο τοπικό ελάχιστο -0.41. Αντίθετα, η Newton λόγω αορίστου Hessian, δεν έχει διεύθυνση καθόδου και ο αλγόριθμος εγκλωβίζεται στο 0.

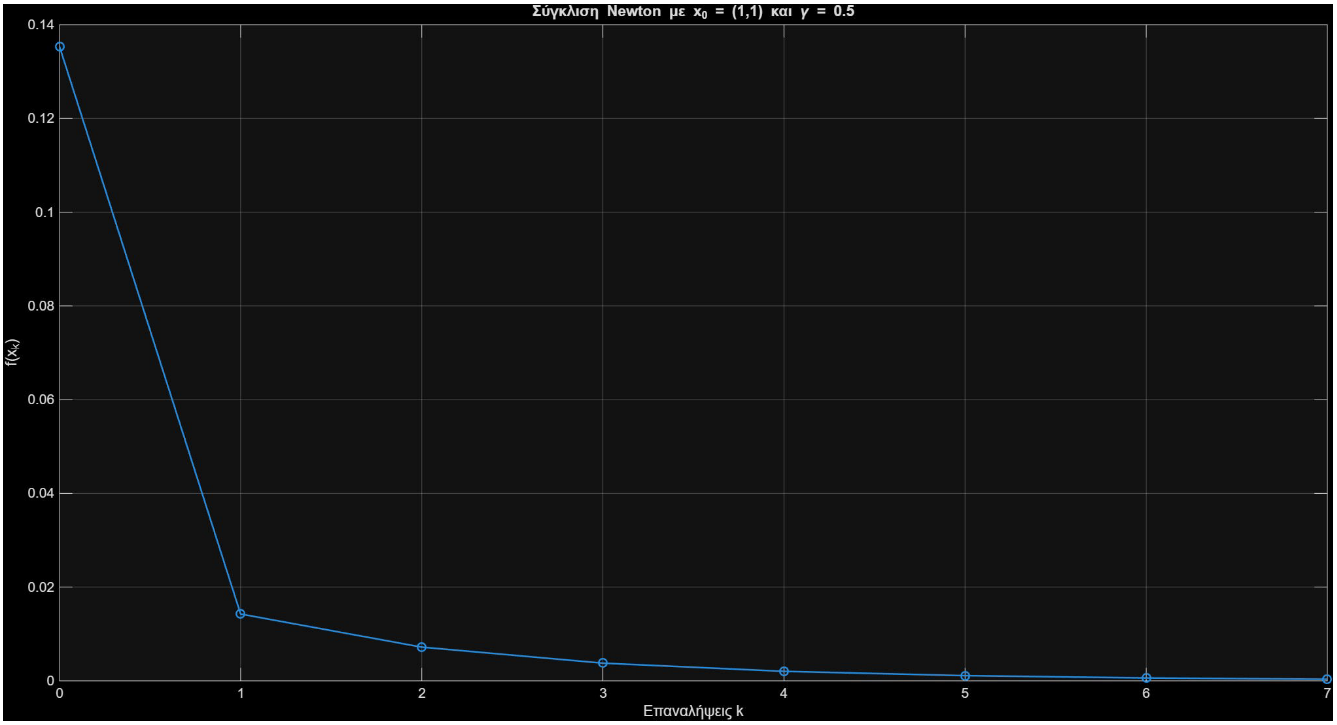
iii) Αρχικό Σημείο $x_0 = (1,1)$

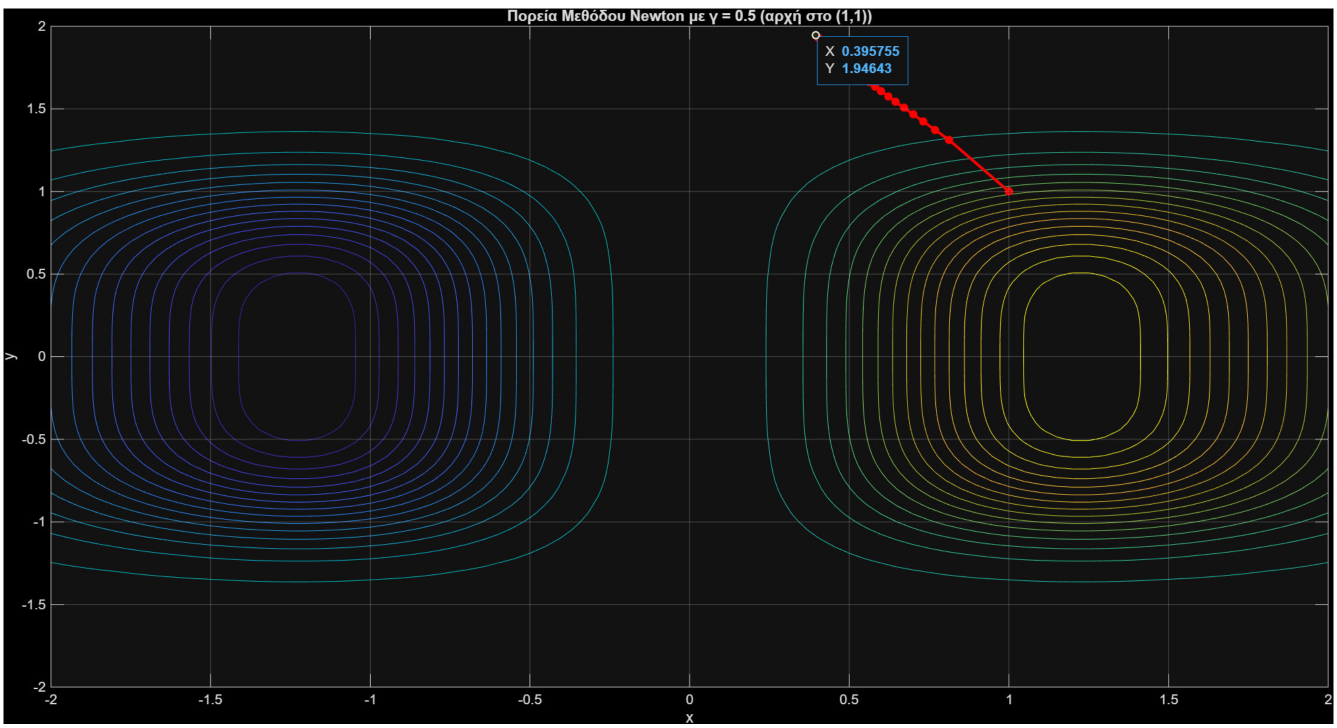
gradient $\nabla f(1,1) \approx \begin{bmatrix} 0.135 \\ -0.541 \end{bmatrix}$, η τιμή της f στο αρχικό σημείο $f(1,1) \approx 0.135$ και Hessian

$H(1,1) \approx \begin{bmatrix} -0.541 & -0.541 \\ -0.541 & 0.541 \end{bmatrix}$ που είναι αντιστρέψιμος αλλά μη θετικά ορισμένος.

Το πρώτο βήμα Newton είναι

$$p_0 = -H(1,1)^{-1} \nabla f(1,1) \text{ , } x_1 = x_0 + 0.5\, p_0$$





Σχολιασμός:

Το αρχικό σημείο (1,1) βρίσκεται στην περιοχή x>0 οπότε η τροχιά Newton μένει στην λεκάνη έλξης και τείνει την τιμή της f στο μηδέν και όχι προς το αρνητικό ελάχιστο στης αριστερή πλευρά.

Τελικά το αποτέλεσμα εξαρτάται από το σημείο εκκίνησης και από τα πολλαπλά στάσιμα σημεία και την περίπλοκη μορφή της συνάρτησης με λεκάνες και λόφους.

Σύγκριση Μέγιστης Καθόδου – Newton με σταθερό γ από το σημείο (1,1)

Και οι δύο μέθοδοι αποτυγχάνουν να συγκλίνουν στο τοπικό ελάχιστο. Ο βασικός λόγος που συμβαίνει αυτό είναι η γεωμετρία της f όπου με x>0 οι κατευθύνσεις καθόδου των δύο μεθόδων οδηγούν προς την επίπεδη περιοχή γύρω από το (0,0). Η Newton συγκλίνει πολύ πιο γρήγορα σε 5-6 βήματα ενώ η Μέθοδος Καθόδου χρειάζεται δεκάδες επαναλήψεις για να κατέβει στο ίδιο σημείο. Παρόλαυτα έχουμε αστοχία σε κάθε περίπτωση και αυτό καταδεικνύει την έντονη εξάρτηση και των δύο μεθόδων από το αρχικό σημείο εκκίνησης.

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

Exact Line Search για την επιλογή του γ_k

Σε κάθε βήμα βρίσκω το γ_k που ελαχιστοποιεί την συνάρτηση κατά μήκος της ευθείας που ορίζεται από το p_k:

$$\varphi_k(\gamma)=f(x_k+\gamma p_k)$$

Η τιμή του βέλτιστου βήματος ορίζεται ως :

$$\gamma_k=arg\min f(x_k+\gamma p_k), \gamma>0$$

i) Αρχικό σημείο $x_0=(0,0)$

gradient $\nabla f(0,0)=0$, Hessian $H(0,0)=\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ μη αντιστρέψιμος, μη θετικά ορισμένος.

Η Μέθοδος Newton δεν μπορεί να ξεκινήσει διότι δεν υπάρχει κατεύθυνση καθόδου. Το Hessian είναι μη αντιστρέψιμο και άρα δεν μπορεί να υπολογίσει κατεύθυνση Newton.

Άρα δεν μπορώ να εφαρμόσω την μέθοδο Newton από αυτό το αρχικό σημείο.

ii) Αρχικό σημείο $x_0=(-1.-1)$

gradient $\nabla f(-1,-1)=\begin{bmatrix} e^{-2} \\ -4e^{-2} \end{bmatrix}$, Hessian $H(-1,-1)=e^{-2}\begin{bmatrix} 4 & 4 \\ 4 & -4 \end{bmatrix}$ $det(H)=-32e^{-4}\neq 0$ άρα το

hessian είναι αντιστρέψιμο.

Κατεύθυνση Newton:

$$p_k=-H^{-1}\nabla f=\begin{bmatrix} \frac{3}{8} \\ -\frac{5}{8} \end{bmatrix}$$

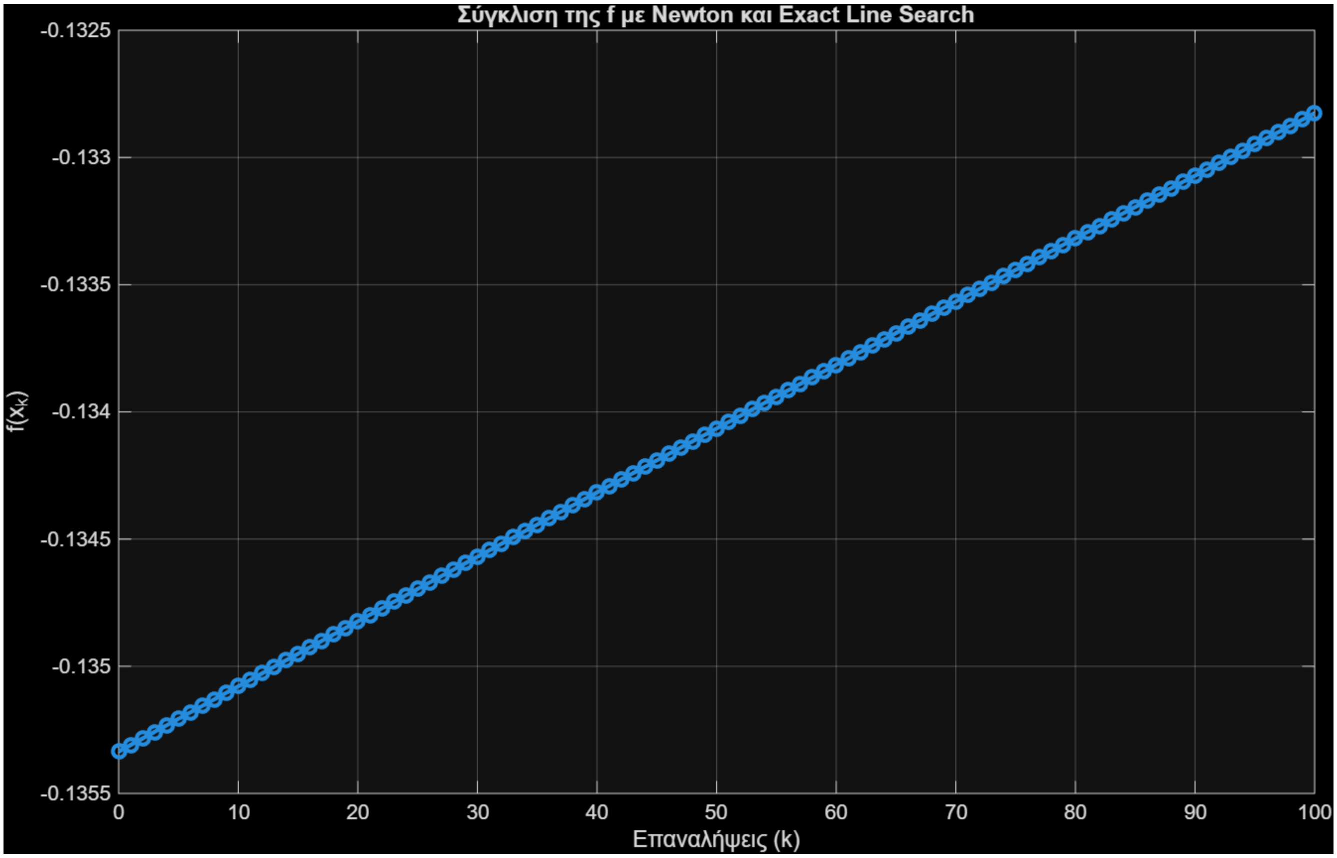
Ορίζω:

$$\varphi_k(\gamma)=f(-1+\frac{3}{8}\gamma, -1-\frac{5}{8}\gamma)$$

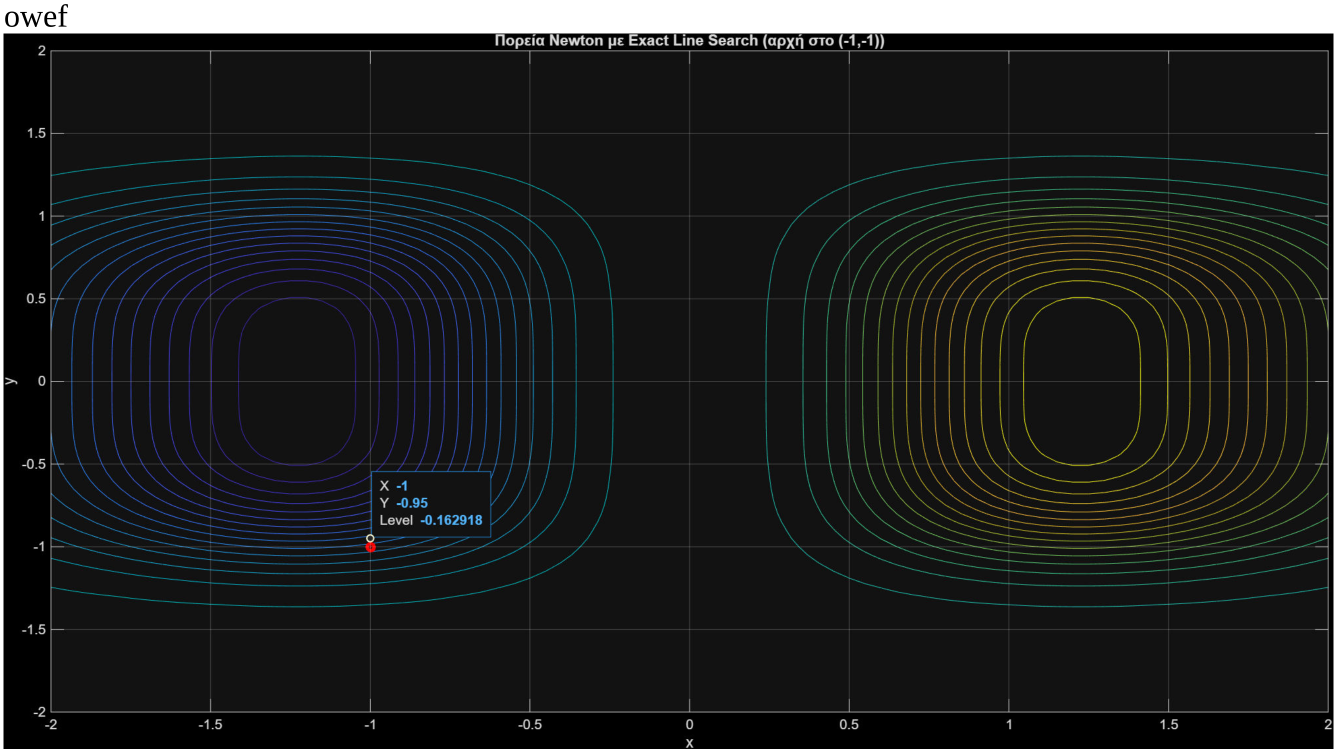
Βήμα γ_k:

$$\gamma_k=arg\min \varphi_k(\gamma)$$

χρησιμοποιώ στον κώδικα την fminbnd στο διάστημα $[0{,}0.5] = [0, \gamma_{max}]$



owef



Σχολιασμός:

Η τιμή της f αυξάνεται γραμμικά με τον αριθμό των επαναλήψεων.

Η Newton με Exact Line Search οδηγείται σε στάσιμο σημείο γρήγορα επειδή το gradient είναι αρκετά μικρό έχουμε πολύ μικρή σχεδόν μηδενική μετακίνηση.

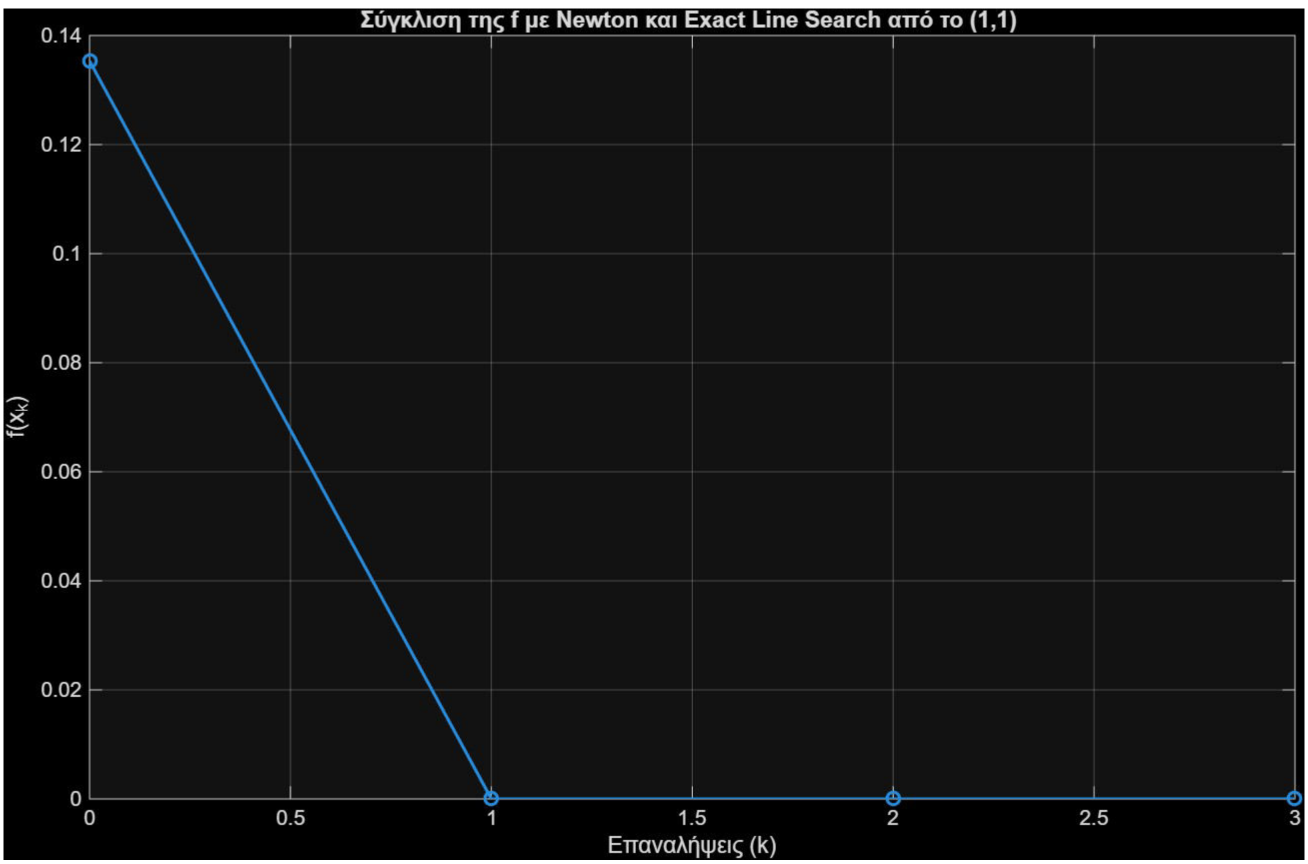
Αποτυχία σύγκλισης στο τοπικό ελάχιστο.

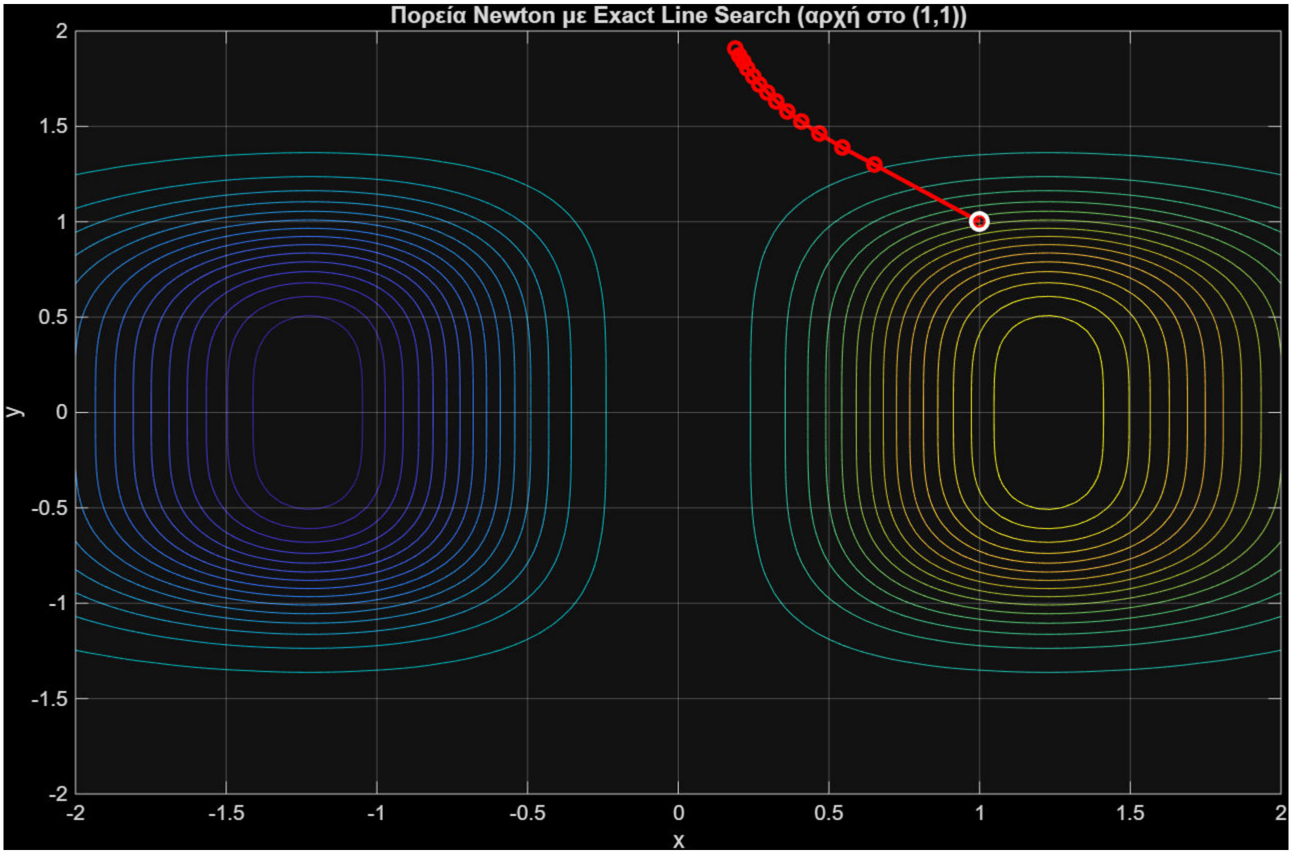
<u>Σύγκριση Μεθόδου Newton – Μεθόδου Καθόδου σε Exact Line Search για (-1,-1)</u>

Με την Μέθοδο Newton έχω πολύ αργή σύγκλιση. Η f αυξάνεται ελαφρώς και έχει εγκλωβιστεί λόγω μη αρνητικής κατεύθυνσης. Από την άλλη με την Μέθοδος Καθόδου η πορεία της είναι ομαλή και πλησιάζει το τοπικό ελάχιστο. Αυτή η μέθοδος για αυτό το αρχικό σημείο είναι πιο σταθερή και ασφαλής. Η μέθοδος Newton επηρεάζεται αρκετά από αυτή την αρχική θέση λόγω του Hessian.

iii) Αρχικό σημείο $x_0=(1,1)$

gradient$\nabla f(1,1)$ και Hessian $H(1,1)$, διεύθυνση $d_0=-H^{-1}\nabla f$





Σχολιασμός:
Βλέπουμε ότι η μέθοδος Newton ξεκινά από το (1,1), ακολουθεί λάθος περιοχή και καταλήγει σε ένα κρίσιμο σημείο της επίπεδης περιοχής της συνάρτησης όπου η f είναι μηδέν και το gradient μηδέν.
Το Exact Line Search ελαχιστοποιεί σωστά την f πάνω στην διεύθυνση που θεωρεί ο αλγόριθμος ως σωστή αλλά δεν φτάνει ποτέ στην αριστερή κοιλάδα στο τοπικό ελάχιστο -0.41.
Έτσι ο αλγόριθμος παγιδεύεται σε ένα κρίσιμο σημείο και αστοχεί.

Σύγκριση Μεθόδου Newton – Καθόδου με Exact Line Search με αρχή το (1,1)

Η Newton χρησιμοποιεί την πληροφορία 2ης τάξης μέσω του Hessian και πετάγεται στην κοιλάδα 0 με μεγάλο βήμα. Από την άλλη η Μέθοδος Καθόδου χρησιμοποιεί μόνο το gradient ως διεύθυνση και με μεγάλο βήμα πάει στην μηδενική κοιλάδα. Και οι δύο μέθοδοι αποτυγχάνουν την σύγκλιση στο ελάχιστο -0.41 από το (1,1) αρχικό σημείο. Μοιάζουν αρκετά στην συμπεριφορά και το αποτέλεσμά τους, διαφέρουν λίγο στην τροχιά προσέγγισης του λανθασμένου σημείου.

γ) Βήμα γ_k με βάση τον κανόνα Armijo

Κανόνας Armijo:

Θεωρώ το $\varphi_k(\gamma)=f(x_k+\gamma d_k),\gamma>0$ και ο κανόνας ικανοποιεί την σχέση:

$$f(x_k+\gamma_k d_k)\leq f(x_k)+\alpha\gamma_k \nabla f(x_k)^T d_k, 0<\alpha<1$$

Διαλέγω σταθερές: 0<β<1 (πχ β=0.5), 0<α<0.5, ένα αρχικό βήμα s>0 θέτω $\gamma_k = s\,\beta^{m_k}$ όπου m_k είναι ο ελάχιστος μη αρνητικός ώστε να ισχύει η ανισότητα Armijo: (5.2.39)

$$f(x_k + s\,\beta^{m_k} d_k) \leq f(x_k) + \alpha\, s\,\beta^{m_k} \nabla f(x_k)^T d_k$$

Στον Newton πρακτικά για k=1,2,3,…

Θα υπολογίσω gradient:

$$g_k = \nabla f(x_k)$$

Αν $\|g_k\| \leq \varepsilon$ τερματίζει

Θα υπολογίσω Hessian:

$$H_k = \nabla^2 f(x_k)$$

Αν είναι αντιστρέψιμος και θετικά ορισμένος για να οδηγηθούμε σωστά στο ελάχιστο, ορίζω διεύθυνση:

$$d_k = -H_k^{-1} g_k$$

και επιλέγω βήμα Armoji:

Θέτω αρχικά $m_k = 0$. Όσο δεν ισχύει η ανισότητα :

$$f(x_k + s\,\beta^{m_k} d_k) \leq f(x_k) + \alpha\, s\,\beta^{m_k} g(x_k)^T d_k$$

Θα αυξάνει $m_k = m_k + 1$ και όταν βρεθεί τέτοιο m_k θέτω $\gamma_k = s\,\beta^{m_k}$ και ενημερώνω $x_{k+1} = x_k + \gamma_k d_k$.

i) Αρχικό σημείο $x_0 = (0,0)$

gradient $\nabla f(0,0) = 0$ Hessian $H(0,0)\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Λόγω μηδενικού gradient ο αλγόριθμος σταματά στο πρώτο βήμα. Ο Armijo δεν ξεκινά καν να λειτουργεί.

Σχόλιο: Το σημείο (0,0) έχει μηδενική κλίση άρα η μέθοδος το θεωρεί στάσιμο σημείο και τερματίζει αμέσως. Η τιμή της f(0,0)=0 προφανώς δεν είναι το ελάχιστο της συνάρτησης και εκεί μένει κολλημένη η μέθοδος.

ii) Αρχικό σημείο $x_0 = (-1,-1)$

gradient $\nabla f(-1,-1) \approx \begin{bmatrix} 0.1353 \\ -0.5413 \end{bmatrix}$ και Hessian $H(-1,-1) \approx \begin{bmatrix} 0.5413 & 0.5413 \\ 0.5413 & -0.5413 \end{bmatrix}$ όπου $detH_0 = -32\,3^{-4} < 0$ μη θετικά ορισμένος.

Διεύθυνση $d_0 = -H_0^{-1} g_0 = \begin{bmatrix} \frac{3}{8} \\ -\frac{5}{8} \end{bmatrix}$ και ελέγχω αν είναι διεύθυνση καθόδου $g_0^T d_0 \approx 0.3891 > 0$ είναι διεύθυνση ανόδου. Δεν ικανοποιείται το κριτήριο 2 του βιβλίου.

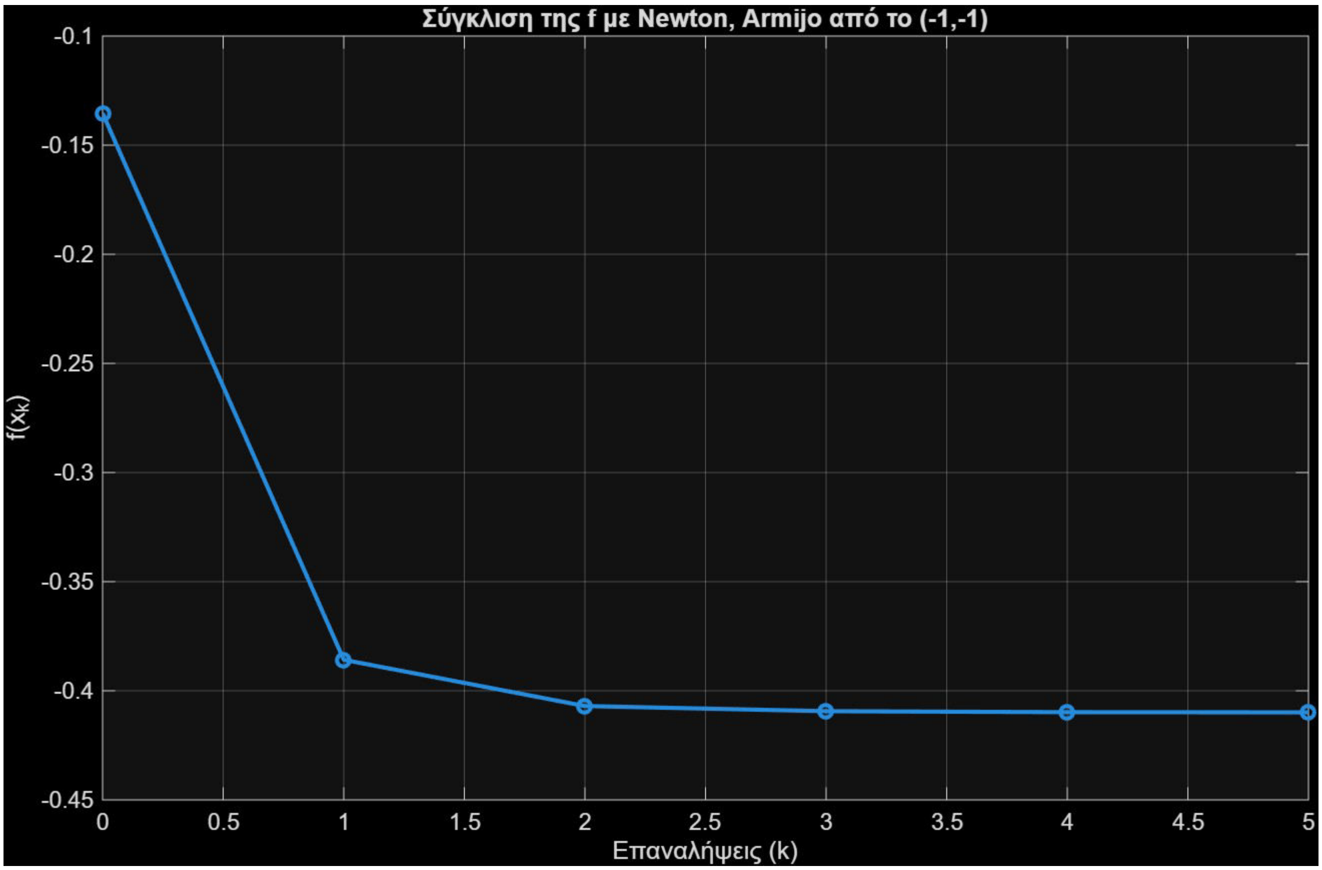
Για το Armijo θεωρώ τη συνάρτηση:

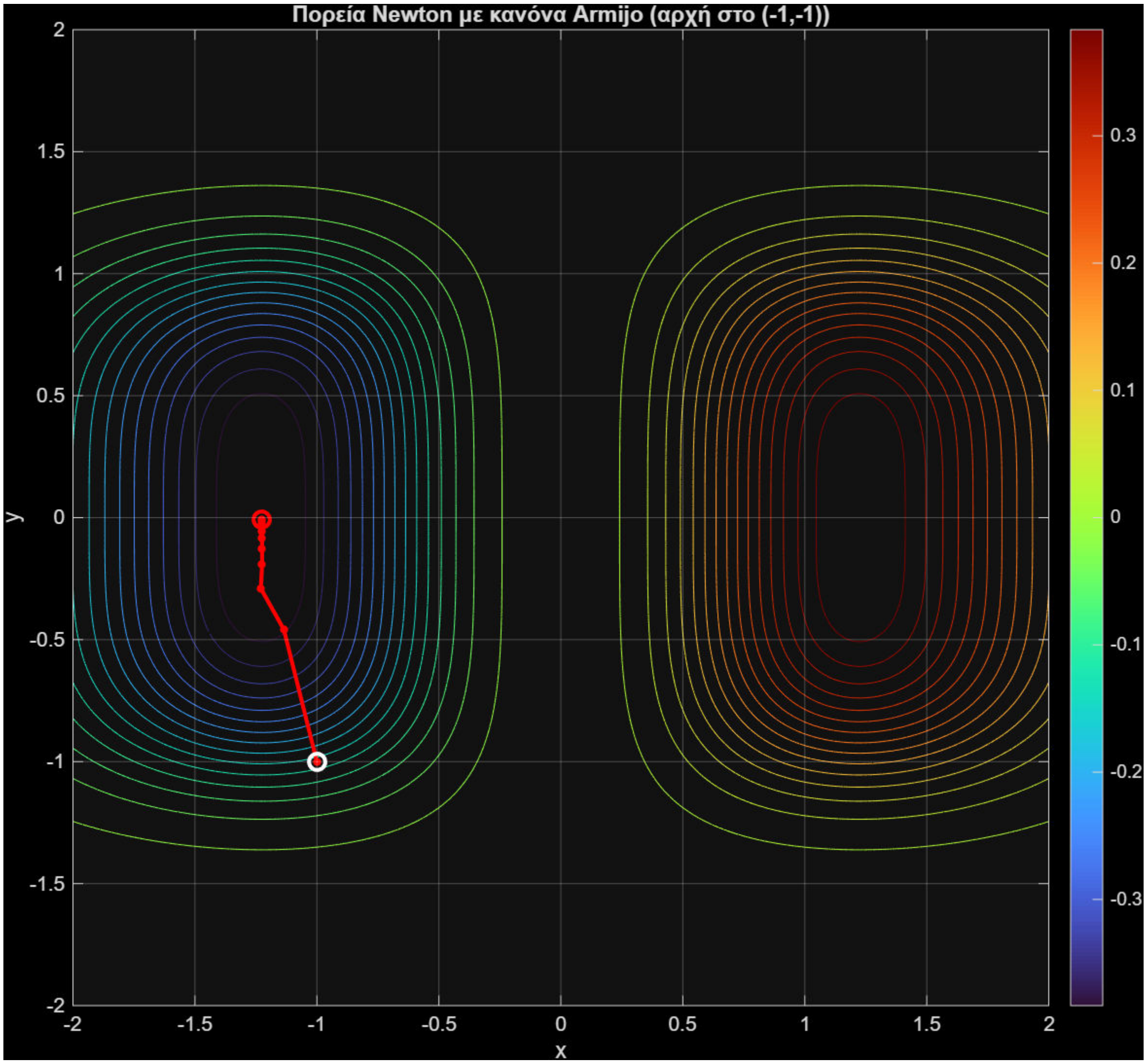
$$\varphi_0(\gamma)=f(x_0+\gamma d_0),\ x_0=(-1,-1)$$

Η παράγωγός της στο γ=0 είναι:

$$\varphi_0{}'(0)=\nabla f(x_0)^T d_0>0$$

Για μικρά θετικά γ η f αυξάνεται κατά μήκους του d_0. Από το Θεώρημα 5.2.6 του βιβλίου που εγγυάται ύπαρξη καταλλήλου γ_k , προϋποθέτει και $d_k^T \nabla f(x_k)<0$ που εδώ δεν ισχύει. Άρα θεωρητικά δεν έχουμε εγγύηση για την Armijo ότι θα βρει αποδεκτό βήμα.





Σχολιασμός:
Ο συνδυασμός Newton και Armoji με αρχικό σημείο (-1,-1) οδηγούν την f στο σωστό τοπικό ελάχιστο.

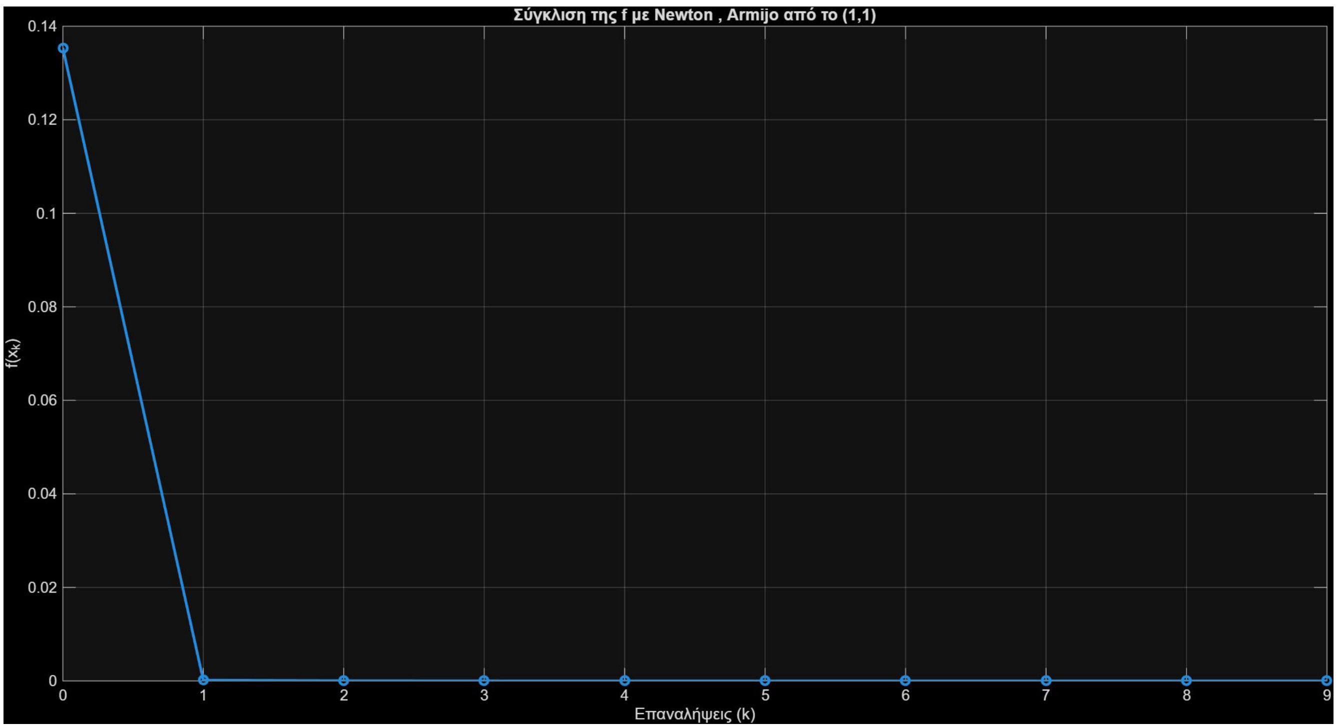
Σύγκριση Μεθόδου Newton- Μέγιστης Καθόδου με κανόνα Armijo και αρχικό σημείο (-1.-1)
Οι δύο μέθοδοι Newton -Μέγιστη Κάθοδος για αρχικό σημείο (-1,-1) συγκλίνουν στο ίδιο τοπικό ελάχιστο το οποίο είναι και τοπικό ελάχιστο της συνάρτησης. Έχουν λίγες επαναλήψεις μέχρι την σύγκλιση, μεγάλες διορθώσεις στην αρχή και μικρές προσαρμογές στη συνέχεια γύρω από το σημείο. Η Newton είναι πιο αποδοτική με λιγότερες επαναλήψεις και πιο άμεση πορεία ενώ η Μέγιστη Κάθοδος χρειάζεται περισσότερα βήματα για να προσεγγίσει το ίδιο ελάχιστο.

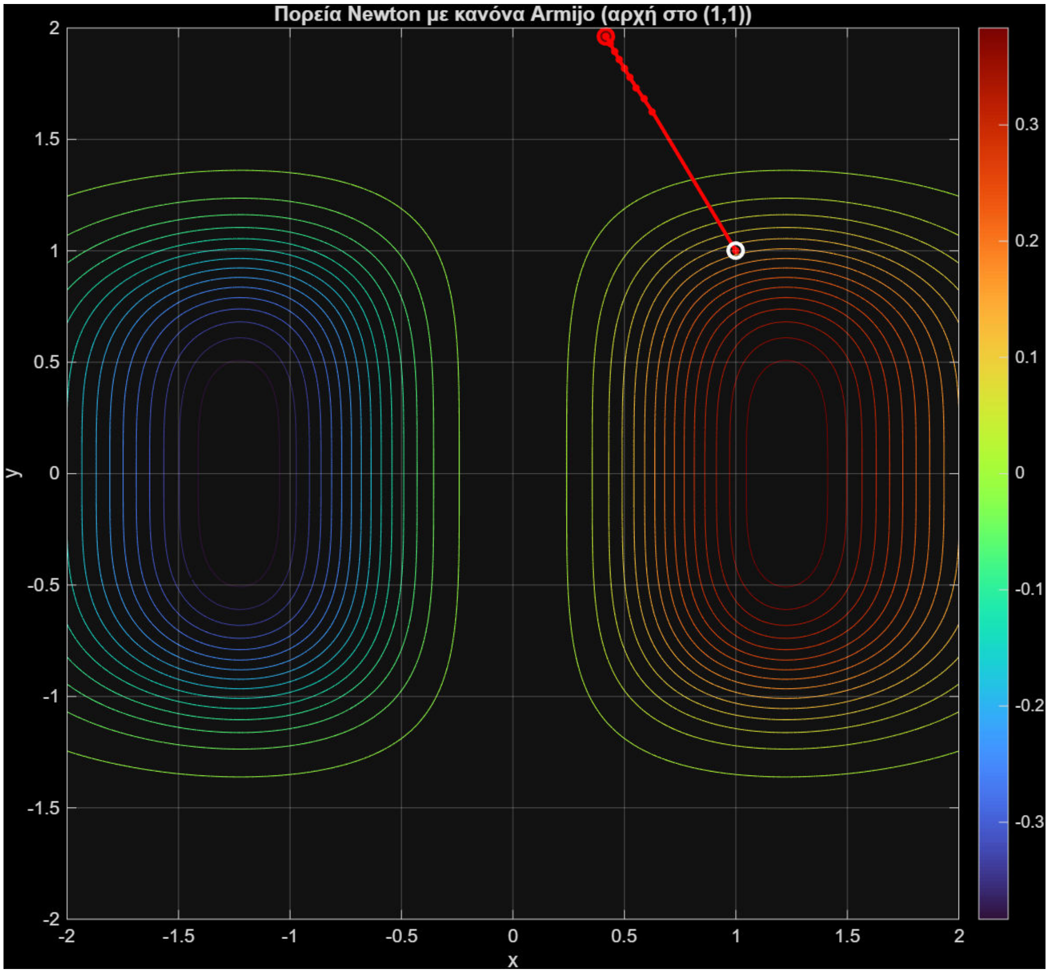
iii) Αρχικό σημείο $x_0=(1,1)$

gradient $\nabla f(1,1)=\begin{bmatrix} e^{-2} \\ -4e^{-2} \end{bmatrix}$ Hessian $H(1,1)e^{-2}\begin{bmatrix} -4 & -4 \\ -4 & 4 \end{bmatrix}$ ιδιοτιμές $\pm 4e^{-2}\sqrt{2}$ άρα είναι αόριστος

διεύθυνση $d_0=-H(1,1)^{-1}\nabla f(1,1)=\begin{bmatrix} -0.375 \\ 0.625 \end{bmatrix}$ κατεύθυνση καθόδου $\nabla f(1,1)^T d_0=-e^{-2}\frac{23}{8}<0$

Κατά μήκος της διεύθυνσης θέτω $x(\gamma)=1-\frac{3}{8}\gamma$ και $y(\gamma)=1+\frac{5}{8}\gamma$ και $\varphi(\gamma)=f(x(\gamma),y(\gamma))$





Σχολιασμός:
Η Newton - Armijo από το (1,1) συγκλίνει γρήγορα αλλά σε λάθος σημείο στο στάσιμο σημείο (0,0) όπου f=0. Επειδή ο Hessian είναι αόριστος και η διεύθυνση Newton οδηγεί στην μηδενική πεδιάδα και όχι στην αρνητική κοιλάδα που υπάρχει το τοπικό ελάχιστο.

<u>Σύγκριση Newton – Μέγιστης Καθόδου με κανόνα Armijo με αρχικό σημείο το (1,1)</u>

Και οι δύο μέθοδοι Newton και Μέγιστης Καθόδου με αρχικό σημείο (1,1) οδηγούν γρήγορα στο ίδιο λανθασμένο σημείο (0,0) όπου f=0 αντί για το πραγματικό τοπικό ελάχιστο της f=-0.41. Παρόλαυτα επειδή και οι δύο χρησιμοποιούν κανόνα Armijo από το συγκεκριμένο σημείο καταλήγουν και οι δύο στην ίδια μηδενική επιφάνεια έλξης και καταλήγουν στο (0,0).

<u>ΘΕΜΑ Δ</u>

Η Μέθοδος Levenberg – Marquardt είναι μια τροποποιημένη μέθοδος Newton που μπορούμε να την χρησιμοποιήσουμε όταν ο Hessian $\nabla^2 f(x_k)$ δεν είναι θετικά ορισμένος.

<u>Σε κάθε επανάληψη για ένα δεδομένο σημείο x_k:</u>

Υπολογίζουμε το διάνυσμα κλίσης $\nabla f(x_k)$ και τον πίνακα Hessian $\nabla^2 f(x_k)$.

Επιλέγουμε μια παράμετρο $\mu_k > 0$ ώστε ο πίνακας

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

να είναι θετικά ορισμένος.

Λύνουμε το σύστημα:

$$B_k d_k = -\nabla f(x_k)$$

για την διεύθυνση d_k

Βρίσκουμε το μήκος βήματος γ_k με κανόνα Armijo.

Και τέλος ενημερώνουμε το

$$x_{k+1} = x_k + \gamma_k d_k$$

Όταν το μ_k είναι μεγάλο η μέθοδος συμπεριφέρεται σαν μέθοδος Μέγιστης Καθόδου, ενώ για μικρό μ_k συμπεριφέρεται σαν Μέθοδος Newton.

α) Σταθερό βήμα $\gamma_k = \gamma = 0.5$

Άρα σύμφωνα με τα παραπάνω $x_{k+1} = x_k + 0.5 d_k$

i) Αρχικό σημείο $x_0 = (0,0)$

gradient $\nabla f(0,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, Hessian $H(0,0) = 0$. Άρα $B_0 = \mu_0 I$ και το βήμα $(\mu_0 I) d_0 = -\nabla f(0,0) \Rightarrow d_0 = 0$

Οπότε $x_1 = x_0 + 0.5 d_0 = (0,0)$

Άρα με αρχικό σημείο το (0,0) η μέθοδος Levenberg – Marquardt με οποιοδήποτε μ_0 και σταθερό $\gamma_k = 0$ μένει κολλημένη στο (0,0) γιατί η κλίση είναι μηδέν. Το σημείο αυτό είναι στάσιμο σημείο της f αλλά όχι το τοπικό ελάχιστο που ψάχνουμε.

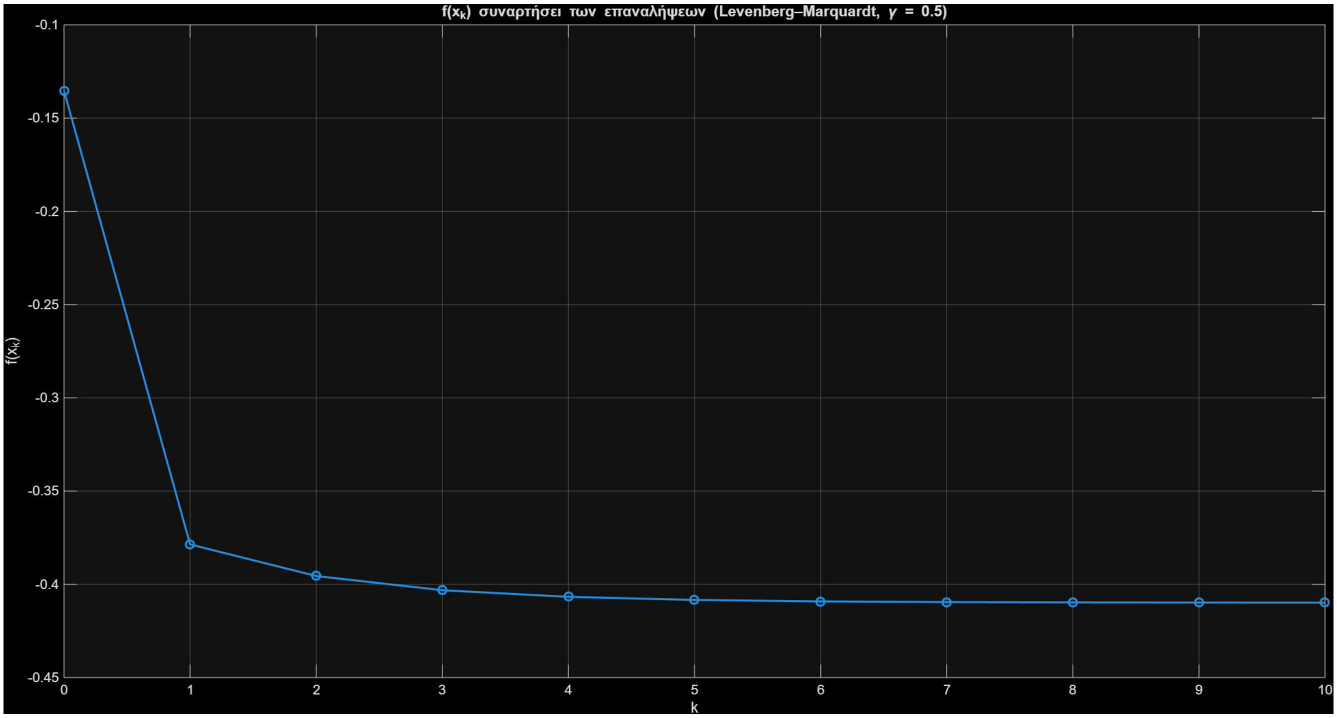
ii) Αρχικό σημείο $x_0 = (-1, -1)$

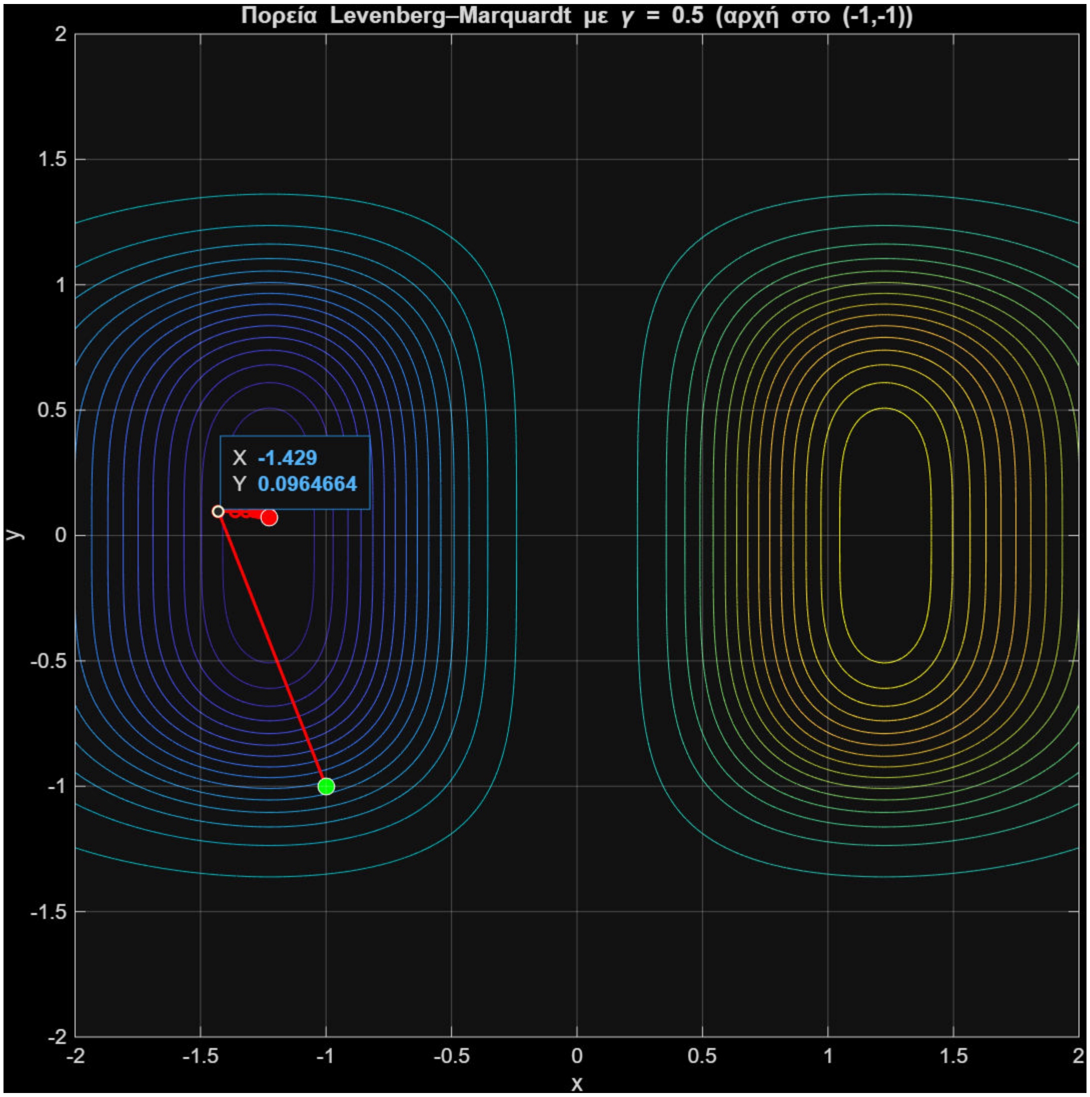
gradient $\nabla f(-1.-1) \approx \begin{bmatrix} 0.1353 \\ -0.5413 \end{bmatrix}$, Hessian $H(-1,-1) \approx \begin{bmatrix} 0.5413 & 0.5413 \\ 0.5413 & -0.5413 \end{bmatrix}$ $f(-1-1) \approx 0.1353$

Με μ=1: $B_0 = H(-1,-1) + I \approx \begin{bmatrix} 1.5413 & 0.5413 \\ 0.5413 & 0.4587 \end{bmatrix}$, λύνω $B_0 d_0 = -\nabla f(x_0) \Rightarrow d_0 \approx \begin{bmatrix} -0.858 \\ 2.1929 \end{bmatrix}$

Ενημερώνω : $x_1 = x_0 + 0.5\, d_0 \approx \begin{bmatrix} -1.429 \\ 0.0965 \end{bmatrix}$ και παίρνω $f(x_1) \approx -0.3786 < f(x_0)$ άρα η μέθοδος έχει

καθοδική κίνηση συνεχίζω με k επαναλήψεις τον αλγόριθμο και βλέπουμε την σύγκλιση παρακάτω:





<u>Σχολιασμός:</u>
Ξεκινώντας από (-1,-1), το πρώτο βήμα της Levenberg – Marquardt είναι μεγάλο διότι έχω $H+\mu I$ το οποίο συμπεριφέρεται σύμφωνα με την παραπάνω θεωρητική με την οποία το πρώτο βήμα συμφωνεί απόλυτα. Στη συνέχει η Μέθοδος συμπεριφέρεται σχεδόν σαν Newton και τραβάει το σημείο στο τοπικό ελάχιστο πού γρήγορα και αποτελεσματικά. Τα επόμενα βήματα είναι μικρά και διορθωτικά όπως ακριβώς είδαμε στην Newton.

Σύγκριση Μεθόδων Levenberg – Marquardt , Newton, Μέγιστης Καθόδου με σταθερό γ =0.5 και αρχικό σημείο το $x_0=(-1,-1)$.

Με αρχικό σημείο (-1,-1) και σταθερό βήμα, η μέθοδος μέγιστης καθόδου κινείται σωστά αλλά με μικρά βήματα, η τροχιά της ακολουθεί ομαλά την κοιλάδα προς το τοπικό ελάχιστο στην αριστερή αρνητική κοιλάδα και το γράφημα της f δείχνει αργή, σταδιακή μείωση που απαιτεί πολλές επαναλήψεις για να πλησιάσει το -0.41.
Αντίθετα η μέθοδος Newton με σταθερό γ είναι πολύ πιο επιθετική , η τροχιά της φεύγει προς τα κάτω μακριά από την περιοχή ελαχίστου και το f αυξάνει προς το 0, οπότε αποτυγχάνει να ελαχιστοποιήσει την συνάρτηση.

Η Levenberg – Marquardt με το ίδιο βήμα συμπεριφέρεται στο πρώτο βήμα σαν την Newton (έντονο και μεγάλο άλμα) και στην συνέχεια λόγω του όρου μI ο τροποποιημένος Hessian είναι πιο ήπιος, κάνει την τροχιά να μείνει στην περιοχή της αρνητικής κοιλάδας και γύρω από το ελάχιστο. Στο γράφημα της f δείχνει γρήγορη και ομαλή σύγκλιση στο -0.41 σε λίγες επαναλήψεις.

Άρα στο συγκεκριμένο σημείο η Levenberg – Marquardt είναι η πιο αποτελεσματική μέθοδος σύγκλισης. Η μέθοδος μεγίστης καθόδου είναι αποτελεσματική αλλά πολύ αργή (δεκάδες βήματα μέχρι την σύγκλιση). Η χειρότερη- αποτυχημένη μέθοδος είναι η Newton.

iii) Αρχικό σημείο $x_0=(1,1)$

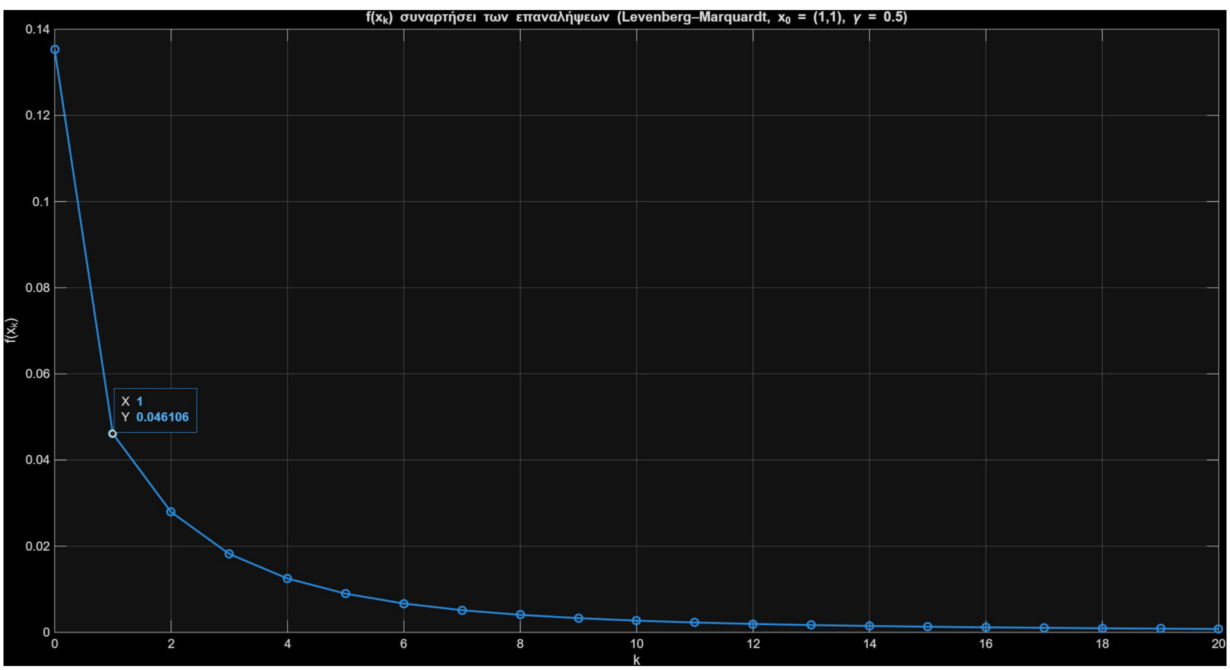
gradient : $\nabla f(1,1)=\begin{bmatrix} 0.1353 \\ -0.5413 \end{bmatrix}$ Hessian $H=\begin{bmatrix} -0.5413 & -0.5413 \\ -0.5413 & 0.5413 \end{bmatrix}$ $f(1,1)\approx 0.1353>0$

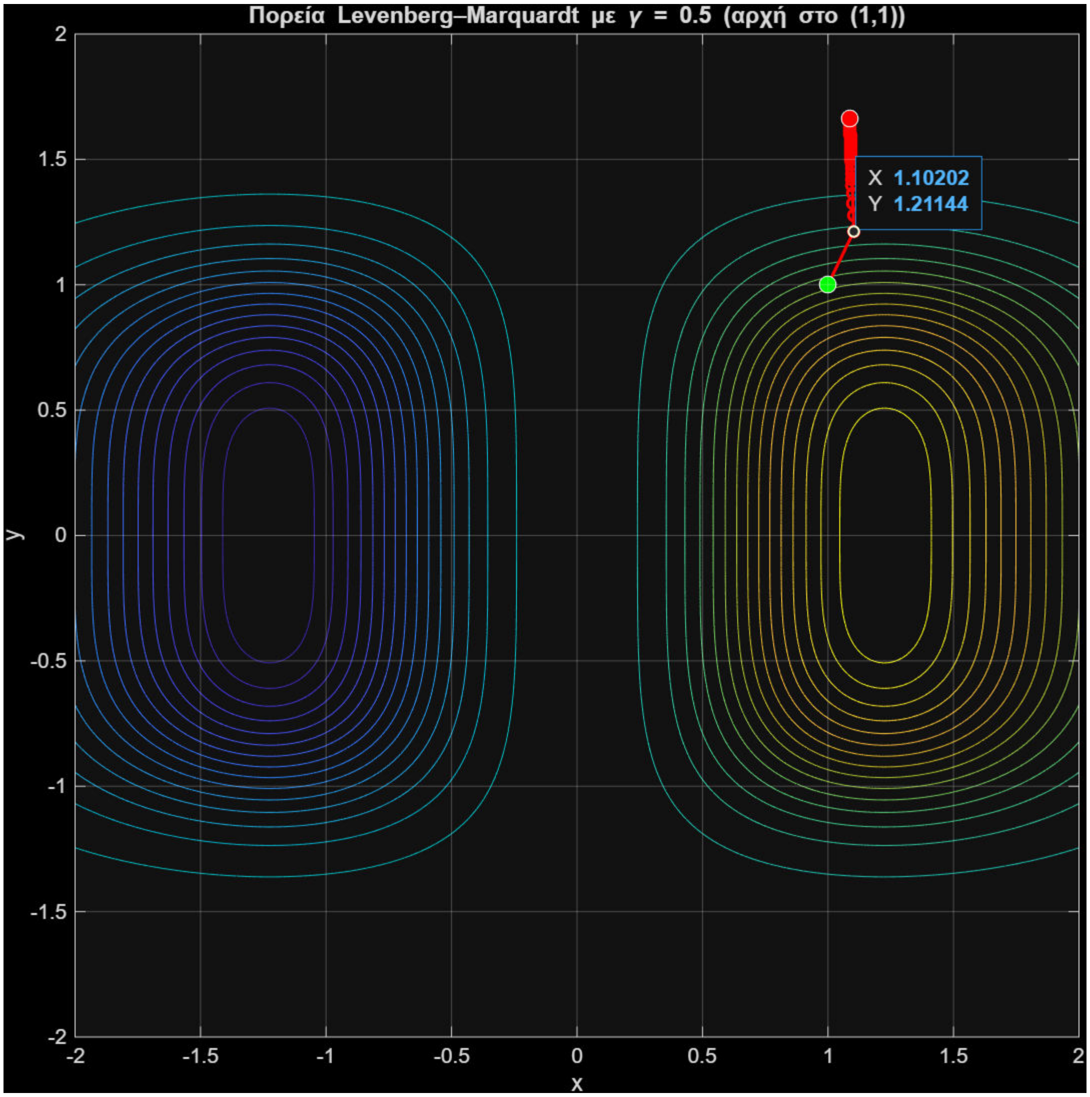
παίρνω μ=1 και έχω $B_0=\nabla^2 f(x_0)+\mu I\approx\begin{bmatrix} 0.4587 & -0.5413 \\ -0.5413 & 1.5413 \end{bmatrix}$ το οποίο έχει θετικές ιδιοτιμές

$$B_0 d_0=-\nabla f(x_0)\Rightarrow d_0\approx\begin{bmatrix} 0.204 \\ 0.423 \end{bmatrix}$$

Ενημερώνω $x_1=x_0+\gamma d_0\approx\begin{bmatrix} 1.102 \\ 1.211 \end{bmatrix}$, $f(x_1)\approx 0.0461<f(x_0)$ άρα το πρώτο βήμα είναι μεγάλο και ρίχνει την f και την μετακινεί προς τα δεξιά και προς τα πάνω.

Μετά από κάποιες επαναλήψεις παίρνω την παρακάτω σύγκλιση:





<u>Σχολιασμός:</u>

Ξεκινώντας από το (1,1) η Levenberg – Marquardt και γ=0,5 κάνει το πρώτο βήμα μεγάλο και μεταφέρει το επόμενο σημείο δεξιά και προς τα πάνω και στην συνέχει οδεύει προς την μηδενική επιφάνεια της f. Αυτό ήταν αναμενόμενο γιατί βρισκόμαστε στον δεξί θετικό λόφο όπου η f για μεγάλα y θετικά σβήνει μέσα από τον εκθετικό της όρο προς το 0. Το πραγματικό ελάχιστο της συνάρτησης βρίσκεται στην αριστερή αρνητική κοιλάδα. Σε κάθε βήμα ο αλγόριθμος μειώνει την f. Βαλτώνει όμως στην πεδιάδα όπου τα x=0, ξεκινώντας από θετικά f (θετικό λόφο) και δεν περνά στην αρνητική κοιλάδα όπου έχω το πραγματικό ελάχιστο της f. Καταλήγει σε περιοχή όπου f=0.

Σύγκριση Μεθόδων Levenberg – Marquardt, Newton και Μέγιστης Καθόδου με σταθερό γ και αρχικό σημείο το $x_0=(1,1)$

Για αρχικό σημείο (1,1) και σταθερό γ και οι τρεις μέθοδοι κινούνται μέσα στον δεξί θετικό λόφο και δεν φτάνουν ποτέ στο πραγματικό ελάχιστο f=-0.41 στην αριστερή κοιλάδα. Όλες οδηγούν σε περιοχές με μεγάλο y όπου ο όρος e^{-y^4} σβήνει την συνάρτηση προς το μηδέν. Η μέθοδος Newton κάνει μεγάλα βήματα και το f πέφτει γρήγορα προς το μηδέν σε λίγες επαναλήψεις. Η Levenberg – Barquardt έχει παρόμοια κατεύθυνση αλλά λίγο πιο ήπια και ομαλή μετά το δεύτερο βήμα λόγω του μI. Η μέθοδος

Μέγιστης Καθόδου κινείται πιο συντηρητικά κάνει πολλά μικρά βήματα και η f μειώνεται πολύ αργά απαιτώντας πολλές επαναλήψεις για να φτάσει στο μηδέν.

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

Exact Line Search για γ_k, ορίζω:

$$\varphi_k(\gamma)=f(x_k+\gamma d_k), \gamma>0$$

Διαλέγω :

$$\gamma_k=arg\,min\,\varphi_k(\gamma)$$

Ψάχνω ρίζα της παραγώγου:

$$\frac{d\varphi_k(\gamma)}{d\gamma}=\nabla f(x_k+\gamma d_k)^T d_k=0$$

και επίσης $\ddot{\varphi}_k(\gamma_k)>0$ για να είναι ελάχιστο.

Ενημέρωση σημείου:

$$x_{k+1}=x_k+\gamma_k d_k$$

Επειδή d_k κατευθύνετε προς ελάχιστα και γ_k είναι ο βέλτιστος μέσα από το exact line search έχω:

$$f(x_{k+1})=min\,f(x_k+\gamma d_k)\leq f(x_k)$$

i) Αρχικό σημείο $x_0=(0,0)$

gradient $\nabla f=\begin{bmatrix}0\\0\end{bmatrix}$ Hessian $H=\begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$ άρα βήμα $B_0=H(0,0)+\mu_0 I$, $\mu_0>0$ και διεύθυνση

$B_0 d_0=-\nabla f(x_0)\Rightarrow d_0=0$

Για το Exact Line Search $\varphi_0(\gamma)=f(0+\gamma 0)=\varphi(0,0)=0$ οποιοδήποτε γ ελαχιστοποιεί τη φ_0.
Το επόμενο σημείο $x_1=x_0+\gamma_0 d_0=(0,0)$.

Από εδώ και πέρα όλες οι επαναλήψεις είναι ίδιες δηλαδή ο αλγόριθμος μένει κολλημένος στο (0,0).

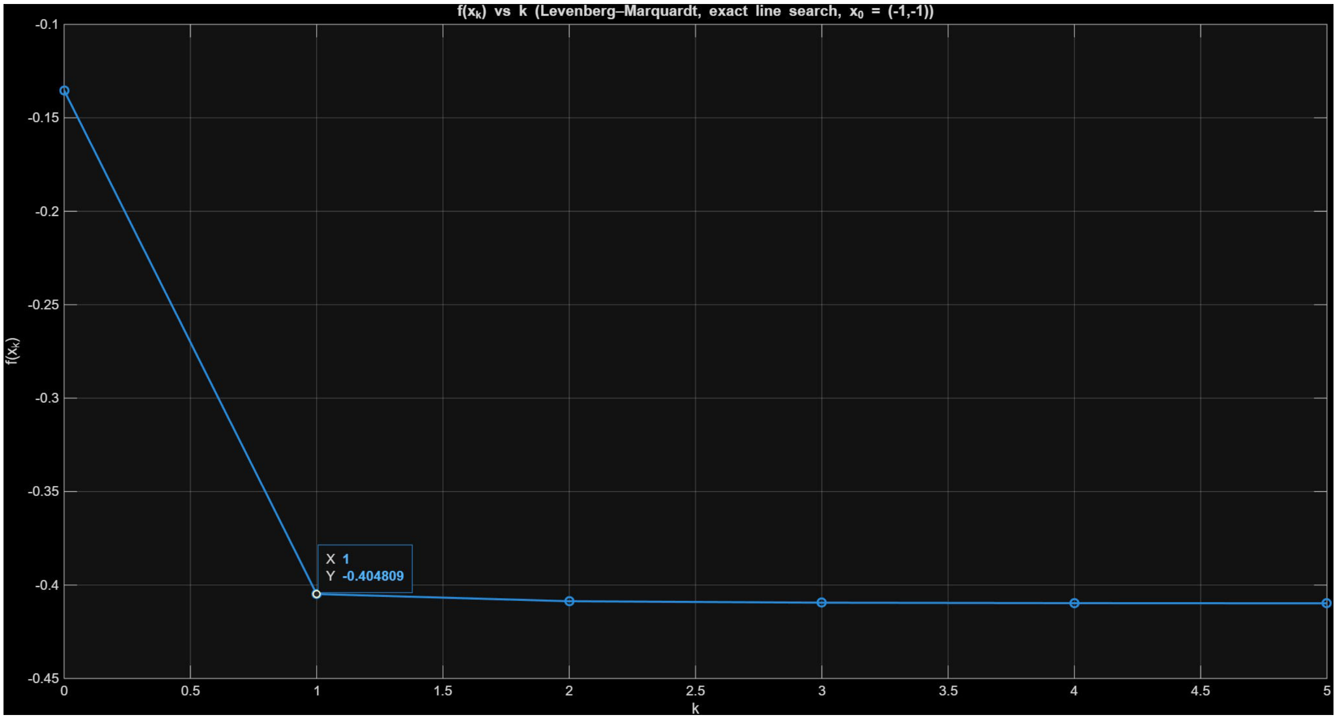
Σχολιασμός: Η μέθοδος συγκλίνει σε λάθος κρίσιμο σημείο λόγω κακής επιλογής αρχικού σημείου. Δεν καταφέρνουμε να μετακινηθούμε με τον αλγόριθμο στο τοπικό ελάχιστο, άρα έχουμε αποτυχία.

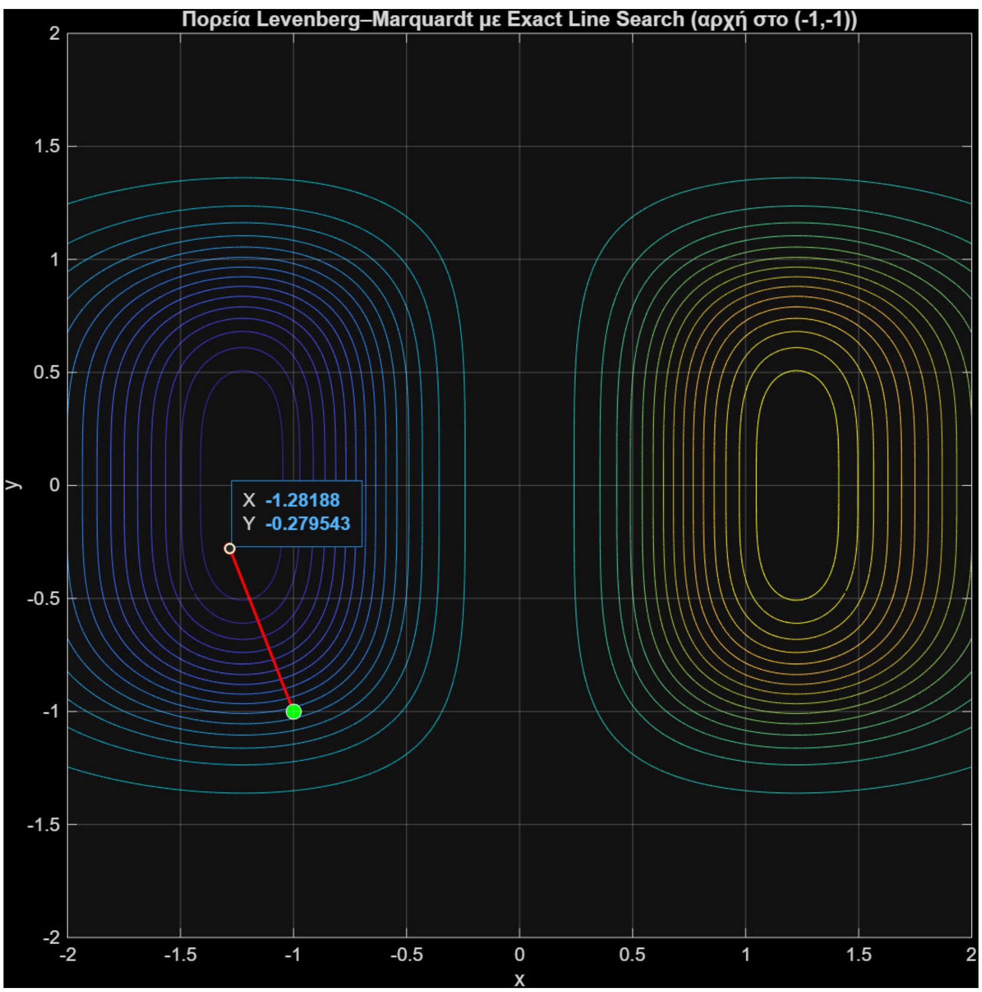
ii) Αρχικό σημείο $x_0=(-1,-1)$

gradient $\nabla f(-1,-1)\approx\begin{bmatrix}0.1353\\-0.5413\end{bmatrix}$ Hessian $H(-1.-1)\approx\begin{bmatrix}0.5413 & 0.5413\\0.5413 & -0.5413\end{bmatrix}$, $d_0\approx\begin{bmatrix}-0.858\\2.193\end{bmatrix}$ και κατιούσα.

Exact Line Search $\gamma_0\approx 0.33$ οπότε $x_1\approx(-1.28,-0.28)$ και $f(x_1)\approx-0.405<f(x_0)\approx-0.135$

Κάνω k επαναλήψεις και προκύπτει το διάγραμμα:


f(x_k) vs k (Levenberg–Marquardt, exact line search, x_0 = (-1,-1))
f(x_k)
k
X 1
Y -0.404809
-0.1
-0.15
-0.2
-0.25
-0.3
-0.35
-0.4
-0.45
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5


Πορεία Levenberg–Marquardt με Exact Line Search (αρχή στο (-1,-1))
X -1.28188
Y -0.279543
y
x
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2

Σχολιασμός:
Με αρχικό σημείο το (-1,-1) η Levenberg -Marquardt με exact line search κάνει ένα ιδανικό πρώτο βήμα μέσα στην αρνητική κοιλάδα του τοπικού ελαχίστου σε σημείο πολύ κοντά στο πραγματικό τοπικό ελάχιστο της συνάρτησης. Και στα δυο διαγράμματα βλέπουμε να συμφωνούν οι τιμές του πρώτου βήματος που υπολογίσαμε. Επίσης στο διάγραμμα $f(x_k)$ βλέπουμε τεράστια πτώση απο -0.135 σε -0.405 ήδη από το k=1. Η μέθοδος φαίνεται να πετυχαίνει την πιο βέλτιστη προσέγγιση του τοπικού ελαχίστου της συνάρτησης -0.41 με ελάχιστες επαναλήψεις και μεγάλο σωστό πρώτο βήμα.

Σύγκριση Μεθόδων Levenberg – Marquardt, Newton, Μέγιστης Καθόδου με Exact Line Search και αρχικό σημείο (-1,-1)

Με αρχικό σημείο (-1,-1) και Exact Line Search η Μέθοδος Levenberg – Marquardt είναι ξεκάθαρα η πιο αποδοτική, δίνει μεγάλο πρώτο βήμα προς τη σωστή κατεύθυνση της αριστερής αρνητικής κοιλάδας του τοπικού ελαχίστου και πέφτει αμέσως κοντά στην τελική τιμή -0.41. Η Μέθοδος Μέγιστης Καθόδου με Exact Line Search δίνει και αυτή τεράστια μείωση της f στο πρώτο βήμα και φτάνει σε τιμή πολύ κοντά στο ελάχιστο αλλά η τροχιά της δεν είναι τόσο ομαλή και θέλει περισσότερες επαναλήψεις για να σταθεροποιηθεί. Αντίθετα, η Newton με Exact Line Search εδώ αποτυγχάνει διότι η Hessian δεν είναι θετικά ορισμένη και οι κατευθύνσεις δεν είναι καθοδικές, οπότε κάνει μικρά βήματα και μένει σχεδόν στάσιμη.

γ) Βήμα γ_k βάσει του κανόνα Armijo

Κανόνας Armijo, ορίζω : 0 < α < 1, 0<β<1 , s>0 και επιλέγω $\gamma_k = s\beta^{m_k}$, $m_k = 0,1,2,\ldots$
Ο κανόνας ζητά το μικρότερο m_k τέτοιο ώστε:

$$f(x_k + \gamma_k d_k) \le f(x_k) + \alpha \gamma_k g_k^T d_k$$

Δηλαδή ξεκινάμε με βήμα s και αν η μείωση της f δεν είναι αρκετή σε σχέση με την πρόβλεψη μειώνουμε το βήμα κατά β και επαναλαμβάνουμε μέχρι να ικανοποιηθεί η ανισότητα.

i) Αρχικό σημείο $x_0 = (0,0)$

gradient $\nabla f(0,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, Hessian $H(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $B_0 = H(0,0) + \mu_0 I$ και διεύθυνση $B_0 d_0 = -\nabla f(0,0)$ είναι $d_0 = 0$.

Ο κανόνας Armijo ψάχνει γ_0 ώστε να ικανοποιεί την ανισότητα:

$$f(x_0 + \gamma_0 d_0) \le f(x_0) + \alpha \gamma_0 g_0^T d_0$$

Όμως $d_0 = 0 \Rightarrow x_0 + \gamma d_0 = x_0$ για κάθε γ_0. Άρα $f(x_0 + \gamma_0 d_0) = f(x_0) = 0$.

Το νέο σημείο είναι $x_1 = x_0 + \gamma_0 d_0 = (0,0)$.

Στις επόμενες επαναλήψεις συμβαίνει ακριβώς το ίδιο για κάθε νέο σημείο. Άρα η Μέθοδος κολλάει στο (0,0) και τερματίζει εκεί άρα αποτυγχάνει να προσεγγίσει το τοπικό ελάχιστο -0.41.

ii) Αρχικό σημείο $x_0=(-1,-1)$

gradient $g_0=\nabla f(x_0)\approx\begin{bmatrix}0.135\\-0.541\end{bmatrix}$ Hessian $H_0\approx\begin{bmatrix}0.54 & 0.54\\0.54 & -0.54\end{bmatrix}$ Παίρνω $\mu_0=1$ και έχω

$B_0 d_0=-g_0\Rightarrow d_0\approx\begin{bmatrix}-0.858\\2.193\end{bmatrix}$ και $g_0{}^T d_0<0$ άρα καθοδική πορεία.

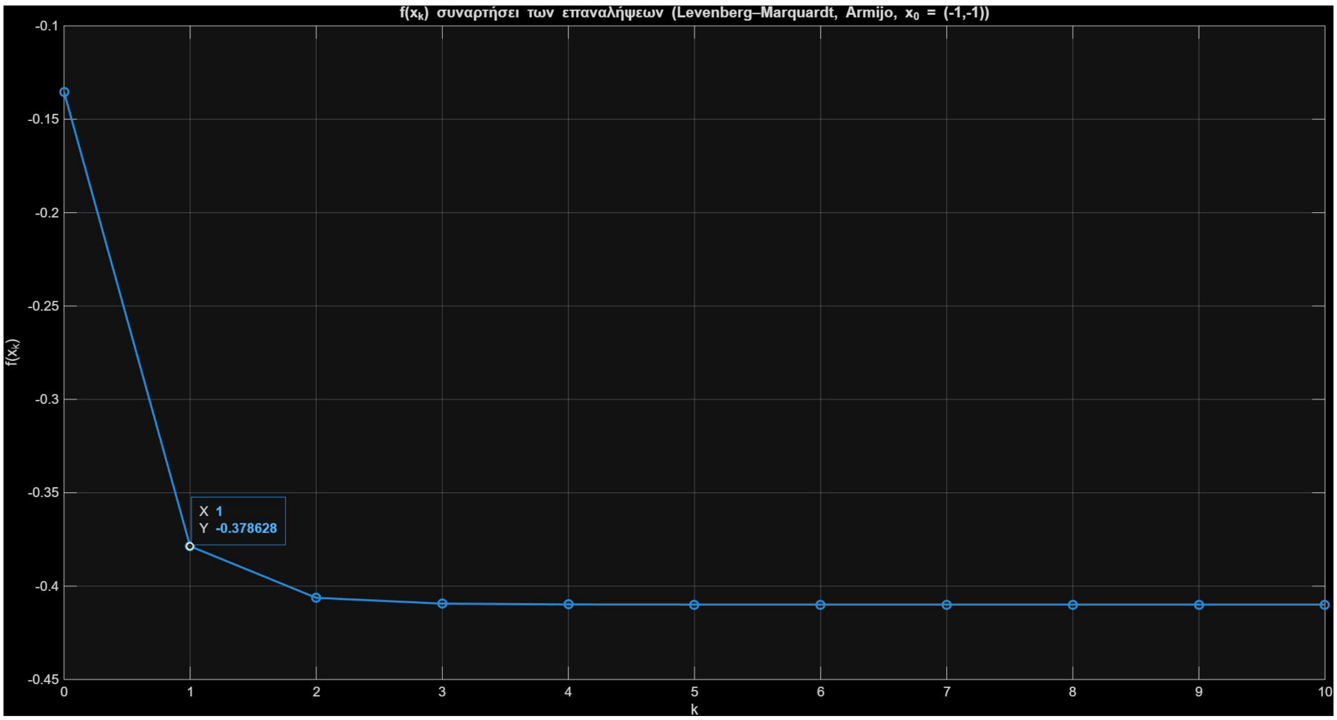
Κανόνας Armoji ορίζω: α=0.1 και β=0.5 και βήμα s=1 και θέτω $\gamma_0=s\beta^{m_0}$, $m_0=0,1,2,\ldots$
Μέχρι να ισχύει η ανισότητα

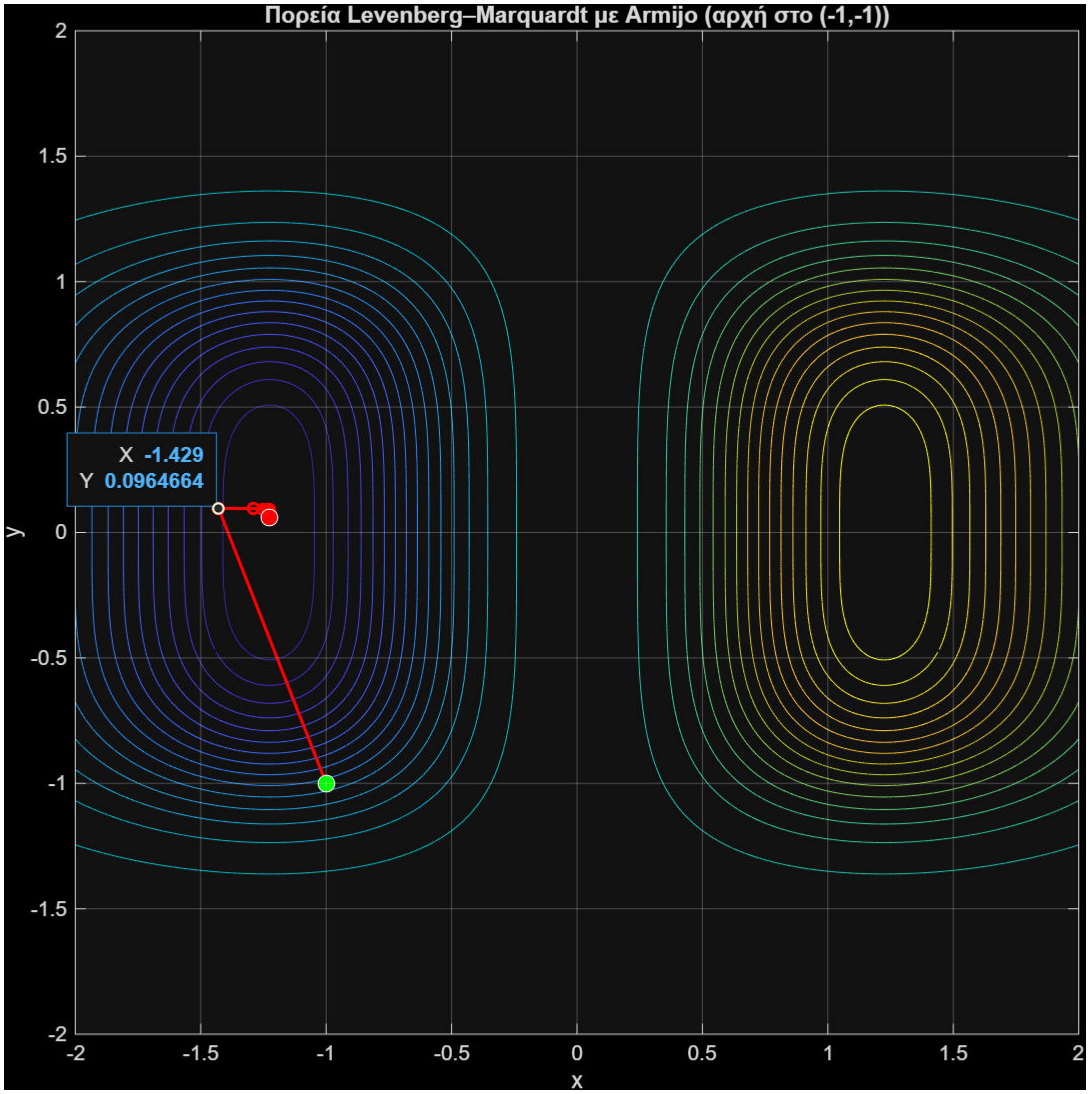
$$f(x_0+\gamma_0 d_0)\leq f(x_0)+\alpha\gamma_0 g_0^T d_0$$

Για γ=0,5 η τιμή της $f(x_0+\gamma d_0)$ ικανοποιεί την ανισότητα.

Το νέο σημείο είναι $x_1=x_0+0.5d_0\approx(-1.429,0.096)$ με $f(x_1)\approx-0.379<f(x_0\approx-0.135)$

Επαναλαμβάνω για κάθε νέο σημείο την ίδια διαδικασία και τα αποτελέσματα σύγκλισης φαίνονται στο παρακάτω διάγραμμα:





<u>Σχολιασμός:</u>
Ξεκινώντας από το (-1,-1) ο Levenberg – Marquardt με κανόνα Armijo κάνει ένα μεγάλο πρώτο βήμα προς τα πάνω και αριστερά φέρνοντας το σημείο μέσα στην αρνητική κοιλάδα κοντά στο τοπικό ελάχιστο και με τιμές ίδιες με αυτές που υπολογίσαμε στην θεωρητική ανάλυση. Αυτό φαίνεται στο διάγραμμα $f(x_k)$ όπου η τιμή της συνάρτησης πέφτει από το -0.135 σε περίπου 0.379 στην πρώτη επανάληψη και στη συνέχεια μειώνεται και σταθεροποιείται γύρω από το -0.41 που είναι η τιμή του ελαχίστου. Πετυχημένη λοιπόν μέθοδος ελαχιστοποίησης.

<u>Σύγκριση μεθόδων Levenberg- Marquardt, Newton, Μεγίστης Καθόδου με κανόνα Armijo και αρχικό σημείο το (-1,-1)</u>

Με αρχικό σημείο το (-1,-1) και κανόνα Armijo και οι τρεις μέθοδοι φτάνουν στο ίδιο σωστό τοπικό ελάχιστο f=-0.41, αλλά με διαφορετική τροχιά. Η Newton με Armijo έχει την καλύτερη σύγκλιση στο διάγραμμα της $f(x_k)$ βλέπουμε μεγάλη πτώση στο πρώτο βήμα και μετά από 2-3 διορθώσεις σταθεροποιείται στο ελάχιστο και η τροχιά της στο επίπεδο x,y είναι σχετικά ευθύγραμμη. Η Levenberg-Marquardt συμπεριφέρεται παρόμοια με λίγο περισσότερες επαναλήψεις μέχρι την σύγκλιση. Η Μέθοδος Μέγιστης Καθόδου με Armijo μειώνει και αυτή απότομα την f στο πρώτο βήμα

αλλά η πορεία της έχει πολλά ζικ-ζακ και χρειάζεται περισσότερες επαναλήψεις για να σταθεροποιηθεί στο ελάχιστο.
Συνολικά από τις τρεις μεθόδους η πιο αποδοτική είναι η Newton η οποία είναι πολύ κοντά με την Levenberg -Marquardt ενώ η Μέθοδος Μέγιστης Καθόδου είναι η πιο αργή.

iii) Αρχικό σημείο $x_0=(1,1)$
gradient $g_0=\nabla f(x_0)\approx\begin{bmatrix}0.1353\\-0.5413\end{bmatrix}$, Hessian $H_0\approx\begin{bmatrix}-0.5413 & -0.5413\\-0.5413 & 0.5413\end{bmatrix}$, $f(x_0)\approx 0.1353$

$B_0=H_0+\mu_0 I\approx\begin{bmatrix}0.4587 & -0.5413\\-0.5413 & 1.5413\end{bmatrix}$ και $B_0 d_0=-g_0\Rightarrow d_0\approx\begin{bmatrix}0.204\\0.423\end{bmatrix}$ ελέγχω αν είναι κατιούσα

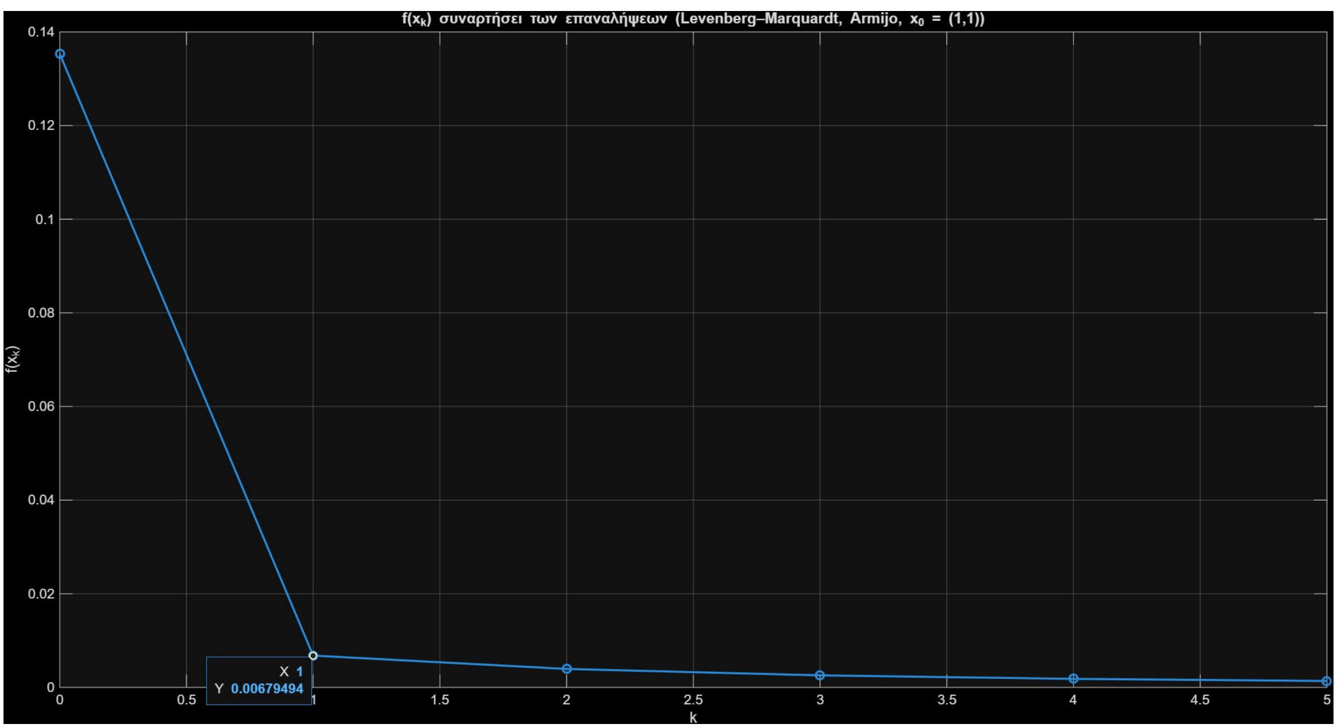
$g_0^T d_0\approx -0.2013<0$ άρα η f μειώνεται.

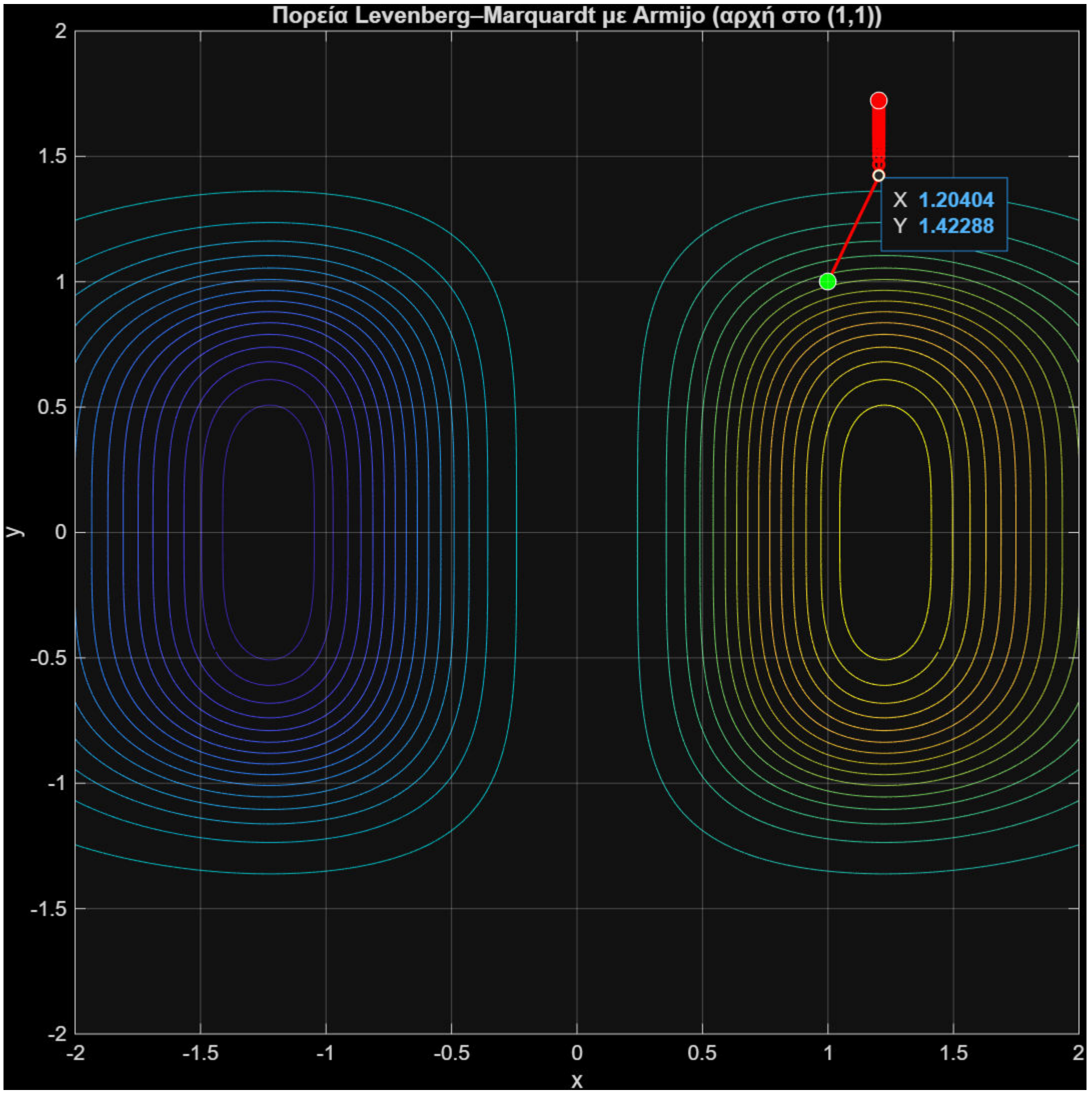
Κανόνας Armijo για το πρώτο βήμα: α=0,1 , β=0,5 και αρχικό βήμα s=1 θέτω $\gamma_0=s\beta^{m_0}$, $m_0=0,1,\ldots$ μέχρι να ισχύει :

$$f(x_0+\gamma_0 d_0)\leq f(x_0)+\alpha\gamma_0 g_0^T d_0$$

Για $\gamma_0=1$ βρίσκω : $x_1=x_0+d_0\approx(1.204,1.423)$ και $f(x_1)\approx 0.0068$ ενώ το δεξί μέλος δίνει $f(x_0)+\alpha\gamma_0 g_0^T d_0\approx 0.1353+0.1(-0.2013)\approx 0.1152$ άρα θα έχω μεγάλη μείωση της συνάρτησης από το πρώτο βήμα.

Επαναλαμβάνω την παραπάνω διαδικασία για κάθε νέο σημείο και παρακάτω στο διάγραμμα φαίνεται η σύγκλιση της μεθόδου:





<u>Σχολιασμός:</u>
Ξεκινώντας από το (1,1), η μέθοδος Levenberg -Marquardt με Armijo κάνει ένα πρώτο μεγάλο βήμα προς τα πάνω και δεξιά βγαίνοντας από τον θετικό λόφο και πηγαίνοντας σε περιοχή όπου ο εκθετικός όρος της f μηδενίζει την συνάρτηση. Αυτό φαίνεται και στο διάγραμμα $f\left(x_k\right)$ όπου από το 0.135 πάμε σε 0.0068 και στις επόμενες επαναλήψεις είμαστε στη γειτονιά του μηδέν. Δηλαδή με τον αλγόριθμο αυτόν οδηγούμαστε στην επίπεδη μηδενική επιφάνεια όπου f=0, και μένουμε εγκλωβισμένοι ενώ το τοπικό ελάχιστο βρίσκεται στην αρνητική κοιλάδα αριστερά. Η μέθοδος αποτυγχάνει.

<u>**Σύγκριση Μεθόδων Levenberg – Marquardt, Newton, Μεγίστης Καθόδου με κανόνα Armijo και αρχικό σημείο (1,1)**</u>

Με αρχικό σημείο (1,1) και κανόνα Armijo, και οι τρεις μέθοδοι φεύγουν από τον δεξί θετικό λόφο και πάνε στην επίπεδη περιοχή όπου η συνάρτηση μηδενίζεται. Αποτυγχάνουν και οι τρεις να βρουν το τοπικό ελάχιστο της συνάρτησης. Από τα διαγράμματα $f\left(x_k\right)$ βλέπουμε ότι τόσο η Newton όσο και η Μέγιστη Κάθοδος ρίχνουν την τιμή της f πρακτικά στο 0 μέσα σε 1-2 επαναλήψειςενώ η Levenberg-Marquardt κάνει λίγες ακόμα επαναλήψεις μέχρι να σταθεροποιηθεί στο μηδέν. Στις τροχιές η μέθοδος Newton ακολουθεί μια πιο καμπύλη πορεία προς το ελάχιστο και η άλλες με παρόμοιες ομαλές τροχιές πάνε προς την ισορροπία. Συνολικά και οι τρεις μέθοδοι είναι πολύ γρήγορες με ελαφρό προβάδισμα της Newton παρόλο που συγκλίνουν αποτυχημένα στο 0 και όχι στο -0.41.

Τελικά Συμπεράσματα

Συνοψίζοντας, το πρόβλημα ελαχιστοποίησης της $f(x,y)=x^3 e^{-x^2-y^4}$ μας ανέδειξε την μεγάλη εξάρτηση των αλγορίθμων τόσο από το αρχικό σημείο όσο και από την επιλογή βήματος γ_k.
Για αρχικό σημείο (0,0) όλες οι μέθοδοι παγώνουν σε στάσιμο σημείο με f=0 ενώ για (1,1) οι περισσότερες τροχιές οδηγούνται στην επίπεδη μηδενική επιφάνεια γύρω από το x=0 και σε καμία περίπτωση δεν είχαμε σύγκλιση στο τοπικό ελάχιστο της συνάρτησης στην αριστερή αρνητική κοιλάδα.
Από το (-1,-1) όταν έχω καθοδικές κατευθύνσεις και βήμα τύπου Exact Line Search ή Armijo, όλες οι μέθοδοι κατάφεραν να συγκλίνουν στο σωστό ελάχιστο, με την Newton και την Levenberg – Marquardt να χρειάζονταν πολύ λίγες επαναλήψεις σε σχέση με την Μέθοδο Καθόδου .
Η Levenberg-Marquardt αποδεικνύεται πιο αποτελεσματική από την Newton σε αόριστο Hessian κάνοντας την έτσι μια πιο ασφαλή επιλογή ανάμεσα στις δύο. Ταυτόχρονα η ταχύτητά της την κάνει πιο αποδοτική απέναντι στην Μέθοδο Μέγιστης Καθόδου. Συνδυάζει γρήγορη και αξιόπιστη σύγκλιση.
Η αιτία που όλες οι μέθοδοι αποτυγχάνουν είναι κυρίως η περίπλοκη γεωμετρία της f που αποπροσανατολίζει τον αλγόριθμο με πολλαπλά στάσιμα σημεία, λεκάνες και λόφους έλξης που κάνουν την τοποθέτηση του αρχικού σημείου να κρίνει το τελικό αποτέλεσμα.