Τεχνικες Βελτιστοποιησης - Εργασια 2

Ρούσος Σταμάτης

30 Νοεμβρίου 2025

Εισαγωγή

Η αναφορα βγηκε πολυ μεγαλη γιατι με επιασε το μπλα μπλα και πορωθηκα επειδη μου αρεσε η εργασια. Διαβαστε οσα θελετε. Πολλα τα εγραψα και προς δικη μου κατανοηση, γιατι οσο αναλυεις και γραφεις, τα καταλαβαινεις και καλυτερα. Τον περισσοτερο χωρο τον επιασαν τα plots, τα οποια δεν καταφερα να βαλω διπλα-διπλα για εξοικονομηση χωρου στο Latex, καθως γινονταν πολυ μικρα. Προτιμησα να τα αφησω ετσι για να βλεπουμε και τι γινεται. Η εργασια βασιστηκε στο βιβλιο του κ.Ροβιθακη:
Τεχνικες Βελτιστοποιησης εκδοσεις ΤΖΙΟΛΑ.

Περιγραφή Προβλήματος

Στην εργασια αυτη θα ασχοληθουμε με την ελαχιστοποιηση συναρτησης πολλων μεταβλητων, χωρις περιορισμους. Θα χρησιμοποιησουμε τρεις βασικες μεθοδους κλισης:

- **Τη μεθοδο της Μεγιστης Καθοδου (Steepest Descent)**
- **Τη μεθοδο Newton**
- **Τη μεθοδο Levenberg-Marquardt**

Στοχος μας ειναι η υλοποιηση αυτων των μεθοδων και η συγκριση τους, σχετικα με τη συγκλιση στο ελαχιστο. Η πορεια που θα ακολουθησω ειναι, αρχικα η μαθηματικη και διαισθητικη αναλυση της καθε μεθοδου (οπως τις καταλαβαινω εγω), και κατοπιν παρουσιαση των αποτελεσματων και σχολιασμος τους.

Ανάλυση των Μεθόδων Κλίσης

Ας φανταστουμε οτι βρισκομαστε πανω σε καποιο σημειο x_1. Και θελουμε απο αυτο το σημειο, να αρχισουμε να κανουμε βηματα προς το ολικο ελαχιστο της δοσμενης συναρτησης. Απο διανυσματικη αναλυση, ξερουμε οτι καθε σημειο αντιστοιχει σε ενα διανυσμα, και μπορουμε να γραψουμε ενα νεο διανυσμα

x_{k+1} ως:

$$x_2 = x_1 + \gamma * d$$

οπου:

- d το διανυσμα κατευθυνσης που μου δειχνει προς τα που θα παω,
- γ ενας συντελεστης που εκφραζει το μηκος του βηματος πανω στο διανυσμα κατευθυνσης, δηλαδη ποσο μακρια θα περπατησω πανω στο διανυσμα κατευθυνσης.

Επειδη σε καθε βημα του αλγοριθμου το διανυσμα κατευθυνσης και το μηκος βηματος αλλαζουν, γενικευουμε τη σχεση για το βημα k ως:

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

οπου:

- x_k: το σημειο εκκινησης
- x_{k+1}: το σημειο προσγειωσης

Το δευτερο μελος της (1), για $\gamma \geq 0$ εκφραζει μια ημιευθεια με αρχη το x_k. Για συγκεκριμενο γ_k, παιρνουμε το ευθυγραμμο τμημα $x_k x_{k+1}$. **Επειδη παντα θα προσγειωνομαστε σε ενα νεο σημειο x_{k+1}, θα φανταζομαστε οτι κινουμαστε πανω σε ενα ευθυγραμμο τμημα, με αρχη το x_k και τελος το x_{k+1}.**

Ολες οι παραπανω μεθοδοι, λεγονται μεθοδοι κλισης, διοτι χρησιμοποιουν την κλιση της συναρτησης δηλαδη το gradient της, για να κανουν το επομενο βημα. Βασιζονται στην ιδεα της επαναληπτικης καθοδου. Δηλαδη, σε καθε βημα τους, κινουνται προς την κατευθυνσης μειωσης (καθοδου) της συναρτησης, προσεγγιζοντας το ελαχιστο. Για να συμβαινει αυτο, πρεπει το διανυσμα κατευθυνσης που θα κινηθουμε, να σχηματιζει αμβλεια γωνια με το $\nabla f(x)$, ή αλλιως:

$$\nabla f^T(x_k) d_k < 0 \tag{2}$$

δηλαδη το εσωτερικο τους γινομενο να ειναι αρνητικο. Μονο τοτε, το βημα που θα κανω θα ειναι καθοδος, δηλαδη η συναρτηση θα μικραινει. Αυτο που αλλαζει απο μεθοδο σε μεθοδο, ειναι η επιλογη του διανυσματος κατευθυνσης d_k. Σε καθε μεθοδο, θα μελετησουμε τρεις διαφορετικους τροπους επιλογης του μηκους βηματος γ_k:
α) Σταθερο και ιδιο σε καθε βημα του αλγοριθμου
β) Τετοιο που να ελαχιστοποιει τη συναρτηση $f(x_k + \gamma_k d_k)$
γ) Βασει του κανονα Armijo

Παρουσίαση τρόπων επιλογής του βήματος γ_k

Θα εξηγησω λιγο τους τρεις παραπανω τροπους επιλογης του βηματος. Οπως ειπαμε, βρισκομαστε σε ενα σημειο x_k και θελουμε να κινηθουμε προς το ελαχιστο της συναρτησης. Αφου διαλεξουμε την κατευθυνση που θα κινηθουμε (d_k), πρεπει να αποφασισουμε ποσο θα κινηθουμε πανω στο διανυσμα κατευθυνσης. Αυτο το κανουμε με την επιλογη του μηκους βηματος γ_k.
Προσοχη!!!
Εαν επιλεξουμε πολυ μικρο γ_k τοτε τα βηματα που θα κανουμε θα ειναι πολυ μικρα και θα χρειαστουν πολλες επαναληψεις για να πλησιασουμε το ελαχιστο $\rightarrow$ αργη συγκλιση.
Εαν παλι επιλεξουμε πολυ μεγαλο γ_k, τοτε ναι μεν κανουμε μεγαλο βημα, αλλα μπορει να ξεφυγουμε μακρια απο το ελαχιστο ή να εχουμε ταλαντωσεις πηγαινοντας περα-δωθε, και ετσι δεν θα μπορεσουμε να φτασουμε στο ελαχιστο.

α) Στην πρωτη περιπτωση επιλεγουμε ενα σταθερο μηκος γ_k (fixed step size) το οποιο παραμενει ιδιο για ολα τα βηματα του αλγοριθμου. Σε αυτη την περιπτωση, δεν εχουμε περιορισμους ουτε ελεγχους για τα βηματα που κανουμε, απλα κινουμαστε κατα ενα συντελεστη γ_k πανω στο διανυσμα κατευθυνσης. Ομως, οπως ειπαμε, η επιλογη που θα κανουμε πρεπει να ειναι σωστη, γιατι πολυ μεγαλα ή πολυ μικρα μηκη προκαλουν προβληματα.

β) Στην δευτερη περιπτωση, ακολουθουμε ενα πιο θεωρητικο τροπο. Μετατρεπουμε το προβλημα αναζητησης ελαχιστου πολλων μεταβλητων, σε μονοδιαστατο προβλημα ελαχιστοποιησης ως προς το βημα γ_k, πανω στη γραμμη αναζητησης η οποια δινεται απο το δευτερο μελος της (1). Ουσιαστικα αναζητουμε το γ_k το οποιο θα μας προσγειωσει σε ενα σημειο x_{k+1} με την ελαχιστη τιμη της συναρτησης σε αυτη τη γραμμη (line search). Συγκεκριμενα, ελαχιστοποιουμε την συναρτηση:

$$\phi(\gamma) = f(x_k + \gamma d_k), \qquad \gamma \geq 0$$

ως προς γ. Λογω των παραπανω, η μεθοδος αυτη ειναι πολυ ακριβης και μαλιστα, σε κυρτες συναρτησεις ειναι σιγουρο πως θα συγκλινει. Ωστοσο, εχει τα εξης μειονεκτηματα:

- Σε καθε βημα απαιτει την ελαχιστοποιηση της συναρτησης ϕ το οποιο σημαινει πολλαπλους υπολογισμους της αντικειμενικης συναρτησης και κατα συνεπεια μεγαλο υπολογιστικο κοστος.
- Σε μη κυρτες συναρτησεις με περιεργη γεωμετρια, ειναι πιθανο να μας οδη-

γησει σε σημειο μακρια απο το ολικο ελαχιστο που θελουμε να προσεγγισουμε.

γ) Στον κανονα Armijo, προσπαθουμε να αποφυγουμε την επιλογη μεγαλου βηματος γ_k, οδηγωντας ετσι σε περισσοτερο ελεγχομενα και σωστα βηματα. Η μεθοδος ουσιαστικα λεει:
"Ξεκινα με ενα αρχικο μηκος γ_k. Θελω να επιλεξεις το βημα γ_k ως:

$$\gamma_k = s\beta^{m_k} \tag{3}$$

με το m_k τον μικροτερο θετικο φυσικο, ωστε να ικανοποιειται το Κριτηριο 4 του βιβλιου, δηλαδη η σχεση:

$$f(x_{k+1}) \leq f(x_k) + \alpha\beta^{m_k} s d_k^T \nabla f(x_k) \tag{4}$$

οπου α,β,s σταθεροι προεπιλεγμενοι παραμετροι με:

$$0 < \alpha < 1$$
$$0 < \beta < 1$$

και s το αρχικο βημα δηλαδη το γ_0."
Αυτο που μας λεει η σχεση (4) ειναι οτι: "θελω να εχω μια ικανοποιητικη μειωση". Θυμομαστε οτι, κινουμαστε προς μια κατευθυνση καθοδου, δηλαδη πρεπει να ισχυει η (2). Επομενως, ο παραγοντας που προστιθεται στο $f(x_k)$ στην (4) ειναι αρνητικος. Αρα, το δευτερο μελος της (4) ειναι μικροτερο του $f(x_k)$. Αυτο μας λεει, οτι "θελω, οχι μονο $f(x_{k+1}) < f(x_k)$, αλλα να μειωθει ακομα περισσοτερο, τοσο οσο ο παραγοντας $\alpha\beta^{m_k} s d_k^T \nabla f(x_k)$.

Στην ουσια, η μεθοδος ξεκιναει με ενα αρχικο $\gamma_k = s$ και ελεγχει αν ικανοποιειται η (4). Αν ικανοποιειται, τοτε κανει το βημα με το αρχικο γ_k. Αν οχι, τοτε αυξανει το m_k κατα 1 και ξανα-υπολογιζει το βημα απο την (3). Επειδη το β ειναι μικροτερο του 1, οσο αυξανεται ο εκθετης m_k, το β^{m_k} μικραινει, και κατα συνεπεια μικραινει και το γ_k. Επομενως, καθε φορα που δεν ικανοποιειται η (4), ριχνει το βημα β φορες.

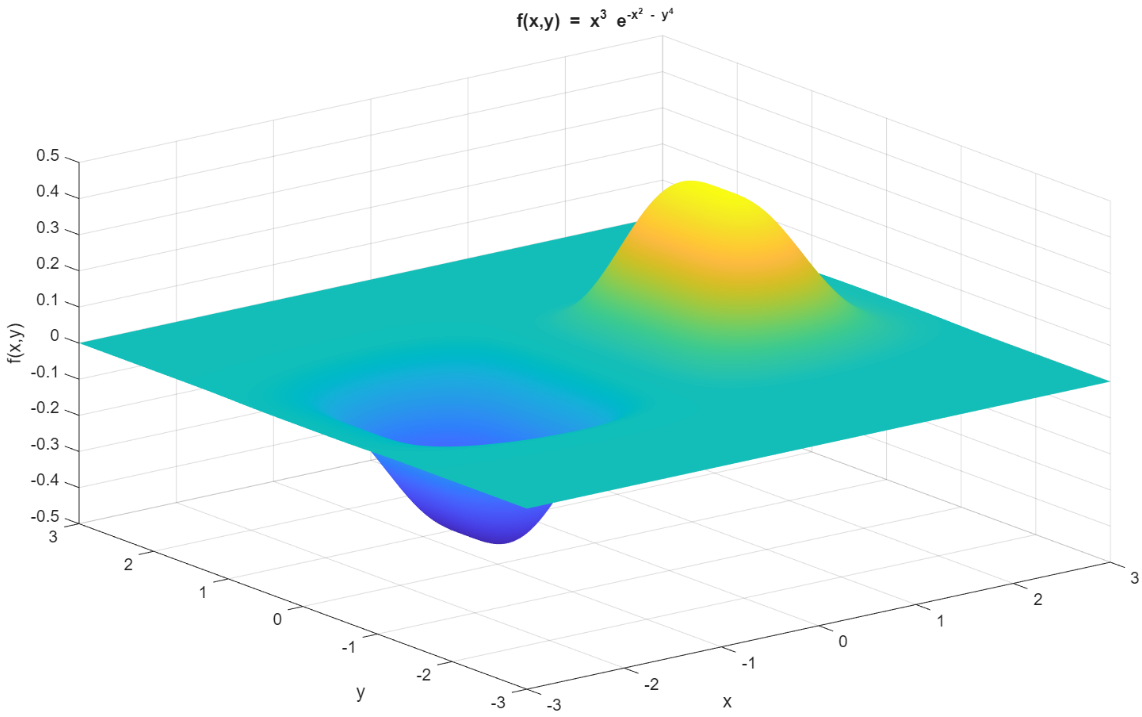
Με αυτο τον τροπο, ο κανονας αυτος ουσιαστικα ξεκιναει "προσγειωνοντας" σε ενα σημειο πανω στην γραμμη για $\gamma_k = s$, το σημειο $x_k + sd_k$, και "σκαναρει" την γραμμη προς τα πισω, ψαχνει δηλαδη κοντινοτερα σημεια στο x_k, μειωνοντας εκθετικα το βημα (για καθε τιμη τιμη γ_k παιρνουμε ενα σημειο πανω στη γραμμη κινησης), ωσπου να πετυχει την ικανοποιητικη μειωση που απαιτει η (4).

Μελέτη Συνάρτησης - Θέμα 1

Παμε τωρα να αναλυσουμε την καθε μεθοδο. Η συναρτηση που θα προσπαθησουμε να ελαχιστοποιησουμε ειναι η:

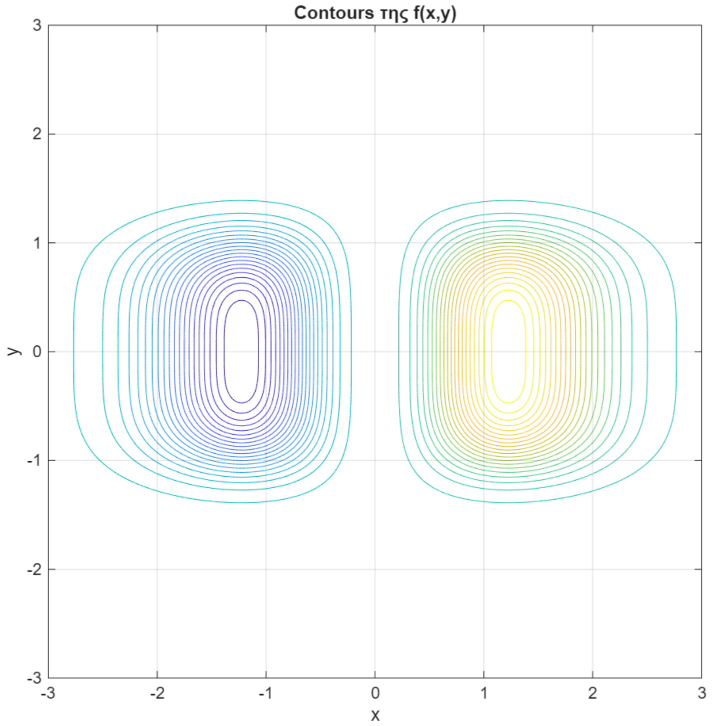
$$f(x,y) = x^3 e^{-x^2 - y^4}$$

με γραφικη παρασταση:



Σχήμα 1: 3D graph of $f(x,y)$

και ισουψεις καμπυλες (contour plot) :



Σχήμα 2: Contour plot of $f(x, y)$

Gradient

Το gradient της παραπανω συναρτησης προκυπτει, με απλη μερικη παραγωγιση της f ως προς x και y, οτι ειναι **το διανυσμα**:

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

Hessian

Ο χεσσιανος της πινακας ειναι:

$$H(x,y) = \nabla^2 f(x,y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial f}{\partial x \partial y} \\ \frac{\partial f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} =$$

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

οπου παρατηρουμε οτι τα στοιχεια στην μη κυρια διαγωνιο ειναι ισα, το οποιο συμβαινει γιατι η f εχει συνεχεις μερικες παραγωγους οποτε δεν παιζει ρολο η σειρα παραγωγισης. Οπως θα δουμε και παρακατω, ο εσσιανος δεν ειναι θετικα ορισμενος σε ολα τα σημεια που οριζεται η συναρτηση, **και ακριβως γι'αυτο το λογο, η f δεν ειναι κυρτη**. Θα δουμε οτι αυτο θα μας προκαλεσει διαφορα προβληματα, **ειδικα στη μεθοδο Newton**.

Κρίσιμα σημεία

Παμε να δουμε τι γινεται με τα ακροτατα τωρα. Αυτα τα βρισκουμε, λυνοντας:

$$\nabla f(x,y) = 0 \Leftrightarrow f_x(x,y) = f_y(x,y) = 0$$

Θελουμε να μηδενιζονται και οι 2 μερικες παραγωγοι ταυτοχρονα. Εχουμε:

$$f_x = 0 \Leftrightarrow -4x^3y^3e^{-x^2-y^4} = 0 \Leftrightarrow x = 0 \text{ ή } y = 0$$

$$f_y = 0 \Leftrightarrow (3-2x^2)x^2e^{-x^2-y^4} = 0 \Leftrightarrow x = 0 \text{ ή } x = \pm\sqrt{\frac{3}{2}}$$

Παρατηρουμε, οτι για $x = 0$ επαληθευονται και οι 2 εξισωσεις ταυτοχρονα. Επισης, επαληθευονται ταυτοχρονα στα σημεια $(-\sqrt{\frac{3}{2}}, 0)$, $(\sqrt{\frac{3}{2}}, 0)$. Επομενως, τα κρισιμα σημεια της συναρτησης ειναι τα: $(-\sqrt{\frac{3}{2}}, 0)$, $(\sqrt{\frac{3}{2}}, 0)$, και ολη η ευθεια $x = 0$, ή αλλιως τα σημεια της μορφης $(0, y)$.

Βαζοντας καθε σημειο στον εσσιανο πινακα, και βλεποντας αν ειναι θετικα ή αρνητικα ορισμενος, συμπεραινουμε οτι, η συναρτηση εχει:

- τοπικο μεγιστο στο $(\sqrt{\frac{3}{2}}, 0)$ το $f(\sqrt{\frac{3}{2}}) \approx 0.41$,
- ολικο ελαχιστο στο $(-\sqrt{\frac{3}{2}}, 0)$ το $f(-\sqrt{\frac{3}{2}}) \approx -0.41$
- και τα σημεια $(0, y)$ ειναι σαγματικα σημεια (saddle points), με $f(0, y) = 0$. Τα σαγματικα σημεια ειναι κρισιμα σημεια (δηλαδη μηδενιζουν το gradient) στα οποια η συναρτηση αλλαζει συμπεριφορα: αυξανεται προς μια κατευθυνση και μειωνεται προς καποια αλλη. Οποτε δεν αποτελουν ακροτατο.

Παμε να δουμε τις μεθοδους.

Κριτήριο Τερματισμού Αλγορίθμων

Καθε αλγοριθμος τερματιζει μολις ικανοποιηθει η συνθηκη:

$$\nabla f(x_k) < \varepsilon$$

οπου ϵ μια πολυ μικρη σταθερα. Εγω την εθεσα ιση με $\varepsilon = 10^{-6}$. Εαν ποτε δεν ικανοποιηθει αυτο το κριτηριο, ο αλγοριθμος σταματαει στις **200 επαναληψεις**.

Θέμα 2 - Μέθοδος Μέγιστης Καθόδου (Steepest Descent)

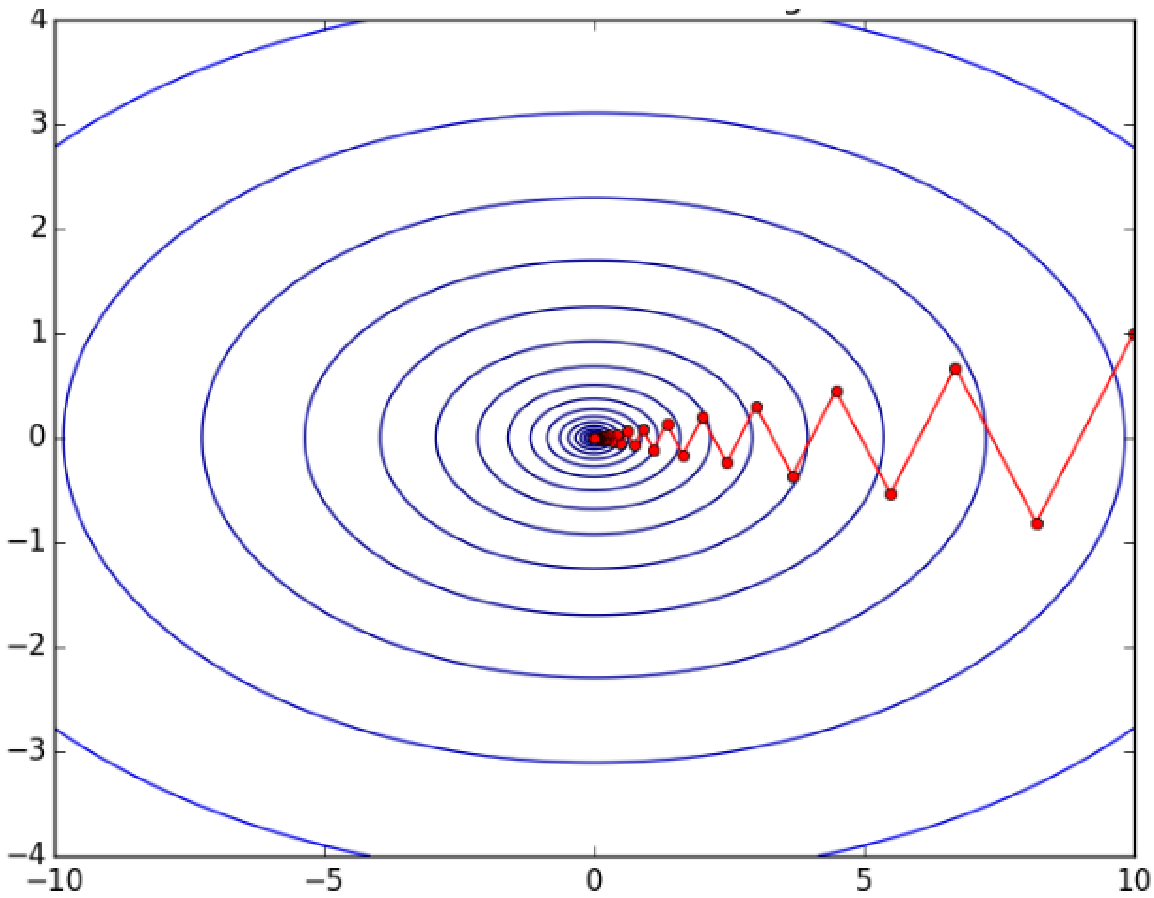
Παρουσίαση Μεθόδου

Η μεθοδος αυτη χρησιμοποιει:

$$\mathbf{d_k} = -\nabla \mathbf{f}(\mathbf{x_k})$$

Γνωριζουμε, οτι το gradient μιας συναρτησης, ειναι ενα διανυσμα που δειχνει σε καθε σημειο προς την κατευθυνση της μεγιστης αυξησης της. Δηλαδη, αν κινηθουμε (στον χωρο των μεταβλητων) πανω στο gradient, η συναρτηση θα αυξανεται με τον μεγιστο δυνατο τροπο. **Προσοχη!** Το gradient "ζει" στο χωροδιασταση των μεταβλητων. Οχι στην γραφικη παρασταση της συναρτησης. Επομενως, στη δικια μας περιπτωση, η γραφικη παρασταση της συναρτησης ειναι 3D, ομως το gradient "ζει" στο επιπεδο (2D) και το βλεπουμε στο contour plot.

Ετσι, η μεθοδος αυτη λεει "Θα παω προς την αντιθετη ακριβως κατευθυνση του $\nabla f(x_k)$, δηλαδη προς την κατευθυνση μεγιστης μειωσης." Η μεθοδος αυτη, ναι μεν ειναι θεωρητικα ακριβης και ειναι σιγουρο πως θα συγκλινει σε κυρτες συναρτησεις, εχει μειονεκτηματα. Συγκεκριμενα, αν η συναρτηση ειναι τετοια που η γεωμετρια της να ειναι για παραδειγμα πλατιες ελλειψουδεις ισουψεις καμπυλες, οπως για παραδειγμα η $f(x, y) = 100x^2 + y^2$, και ξεκιναω απο ενα σημειο στο πλαι, τοτε θα οδηγηθω σε μια κινηση "ζιγκ-ζαγκ" η οποια θα αργησει παρα πολυ να συγκλινει. Οπως για παραδειγμα εδω:



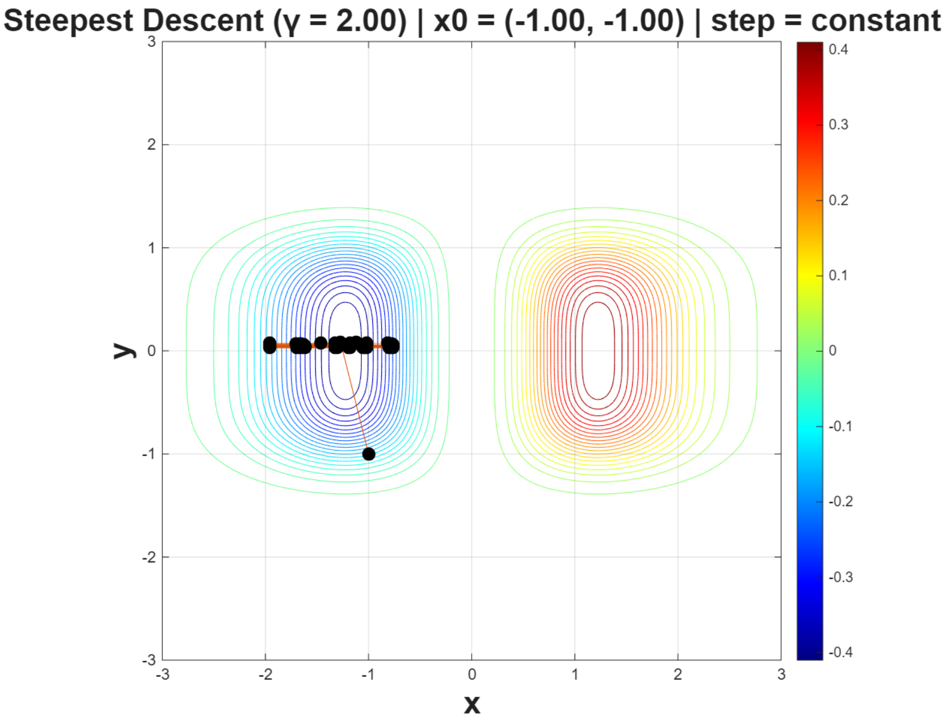
Επομενως, η μεθοδος λογω της γραμμικοτητας που χρησιμοποιει, δεν λαμβανει υποψην τη γεωμετρια/καμπυλοτητα της συναρτησης.

Πειραματικά Αποτελέσματα:

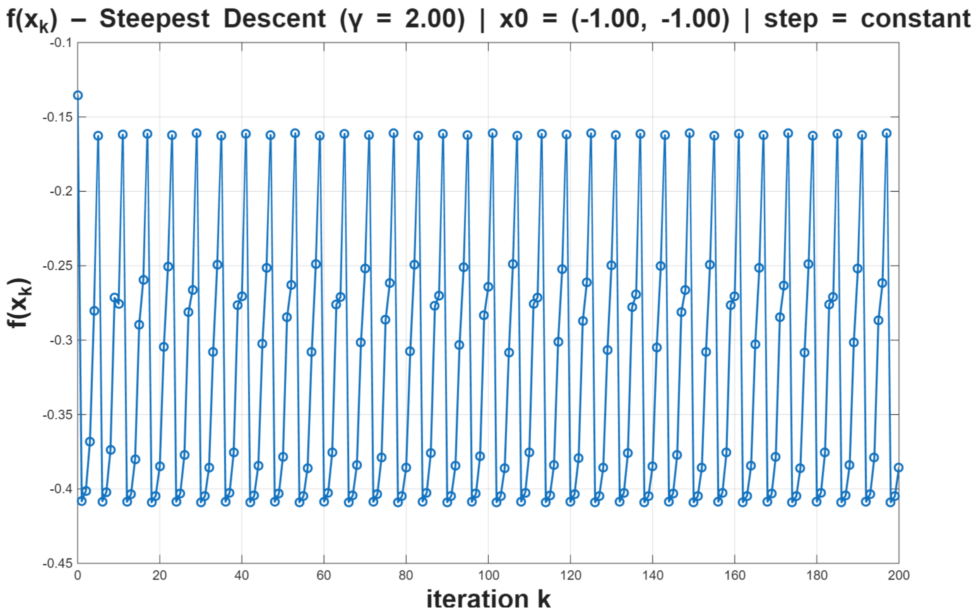
Α - Σταθερο βημα $\gamma_{\mathbf{k}}$

Διαλεξα 3 διαφορετικες τιμες σταθερου βηματος (μια πολυ μεγαλη, μια κανονικη, και μια πολυ μικρη), για να δω τι συμβαινει. Συγκεκριμενα, επελεξα τις τιμες $\gamma_k = (2, 0.5, 0.01)$. Στο πανω σχημα θα βλεπουμε τη συγκλιση στο contour plot, ενω στο κατω τις τιμες $f(x_k)$ σε καθε βημα k. Κανονικα, η κατω γραφικη παρασταση ειναι διακριτα σημεια και οχι ενιαια γραμμη, αφου το k παιρνει διακριτες τιμες. Για αισθητικους λογους παρουσιαζεται ετσι, ομως τα κυκλακια αναπαριστουν το πραγματικο διαγραμμα.

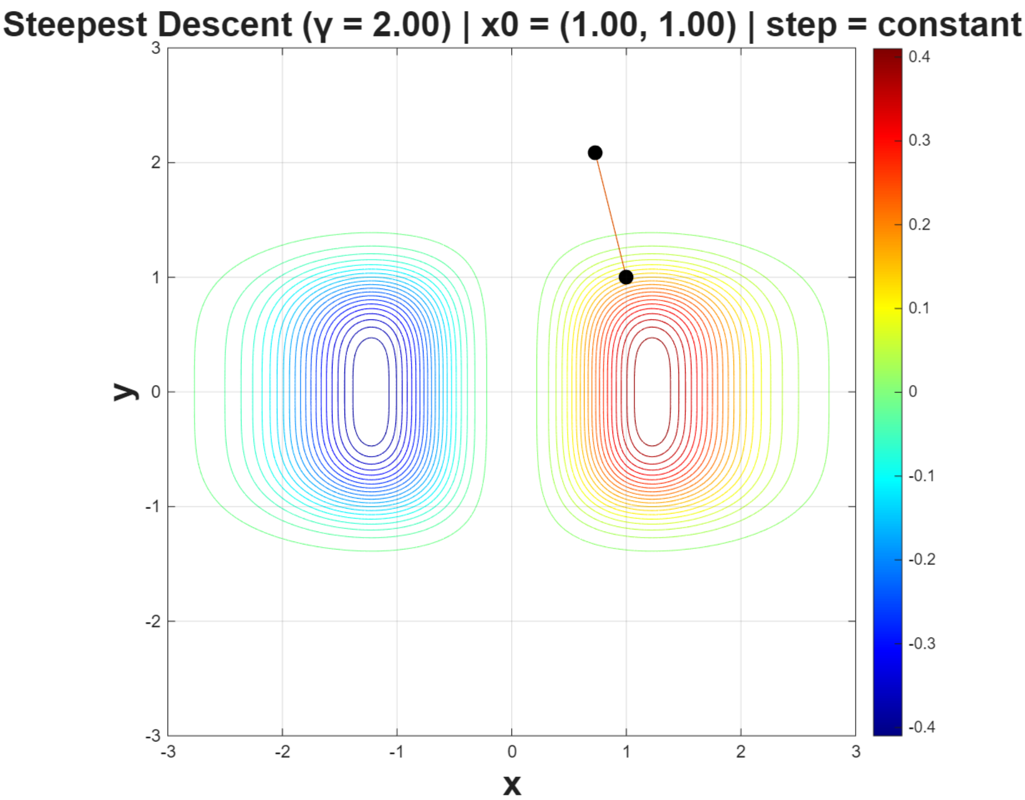
-Για $\gamma_{\mathbf{k}} = \mathbf{2}$:



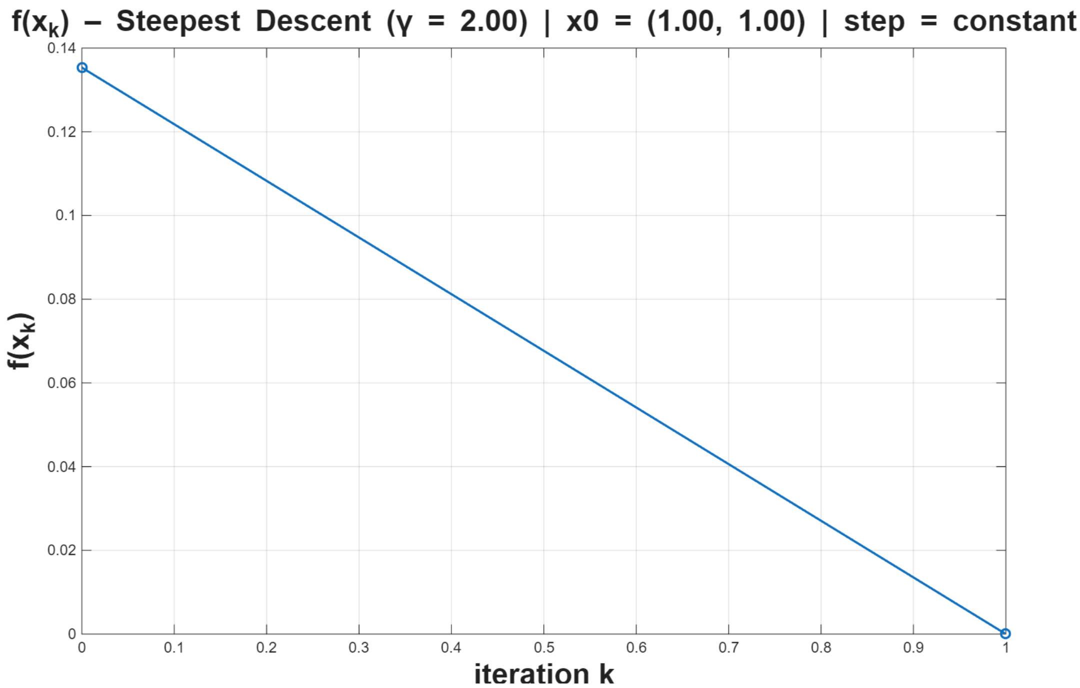
Steepest Descent (γ = 2.00) | x0 = (-1.00, -1.00) | step = constant
x
y



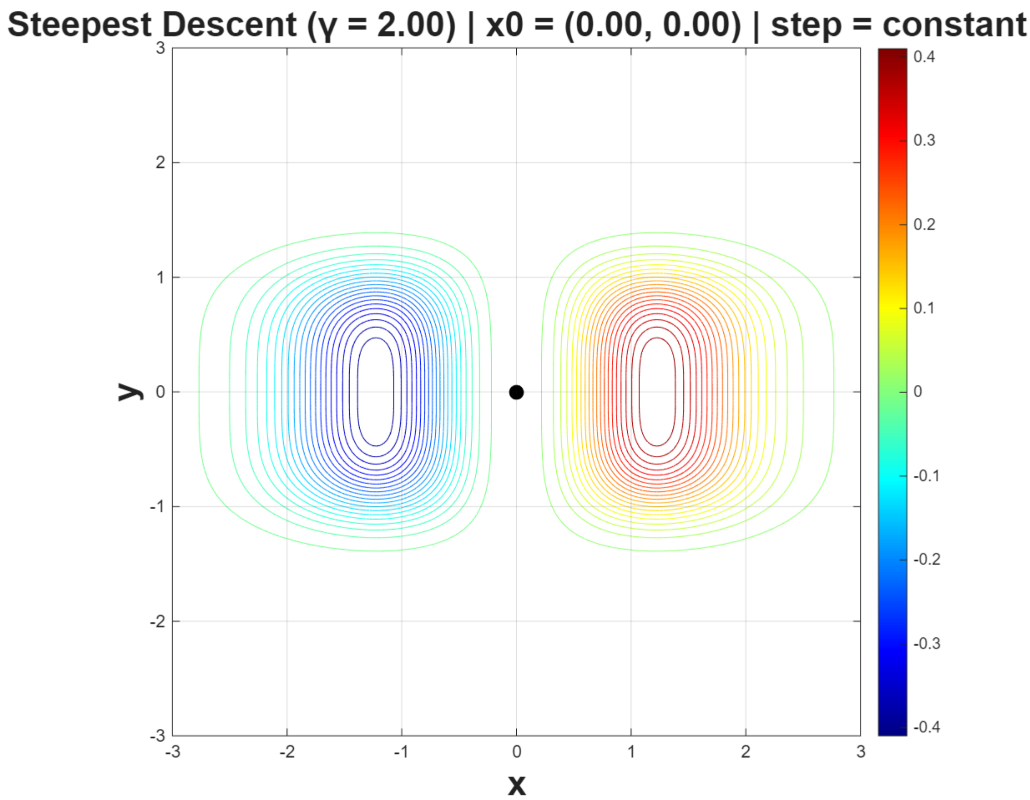
f(x_k) – Steepest Descent (γ = 2.00) | x0 = (-1.00, -1.00) | step = constant
iteration k
f(x_k)

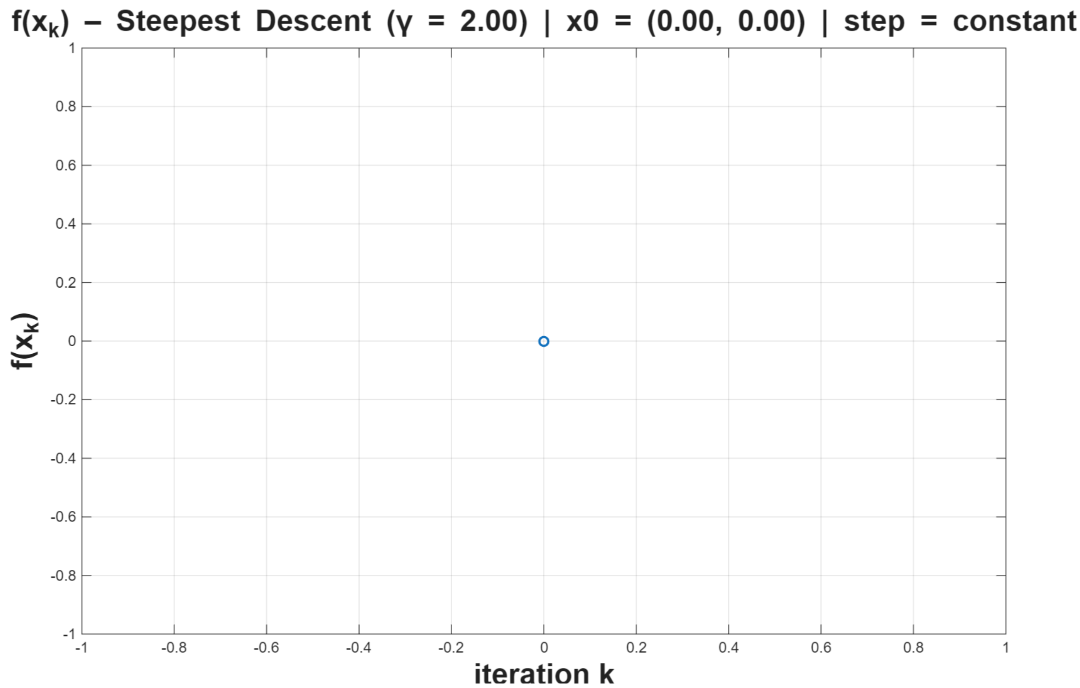


Steepest Descent (γ = 2.00) | x0 = (1.00, 1.00) | step = constant
x
y



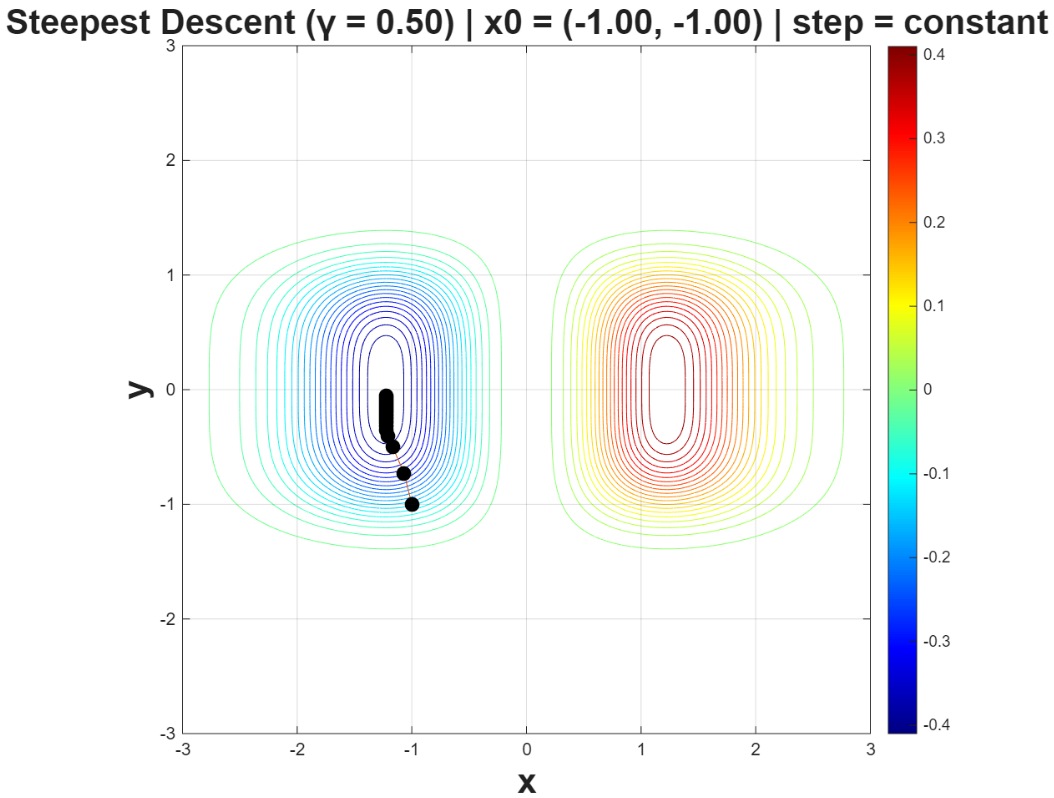
f(x_k) – Steepest Descent (γ = 2.00) | x0 = (1.00, 1.00) | step = constant
iteration k
f(x_k)

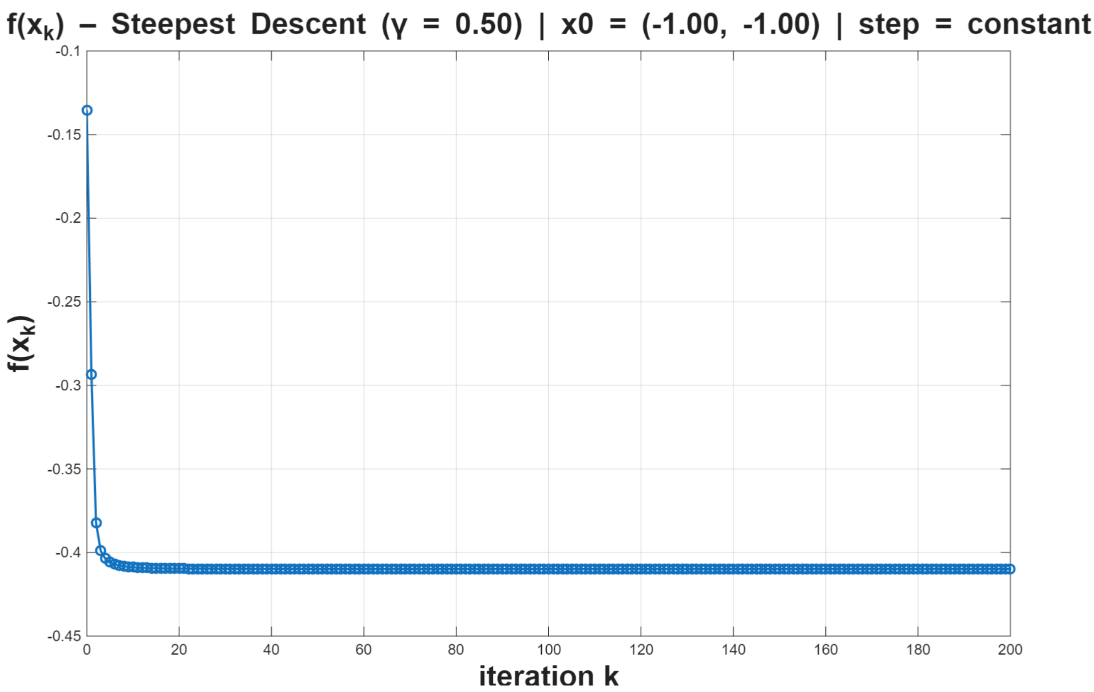


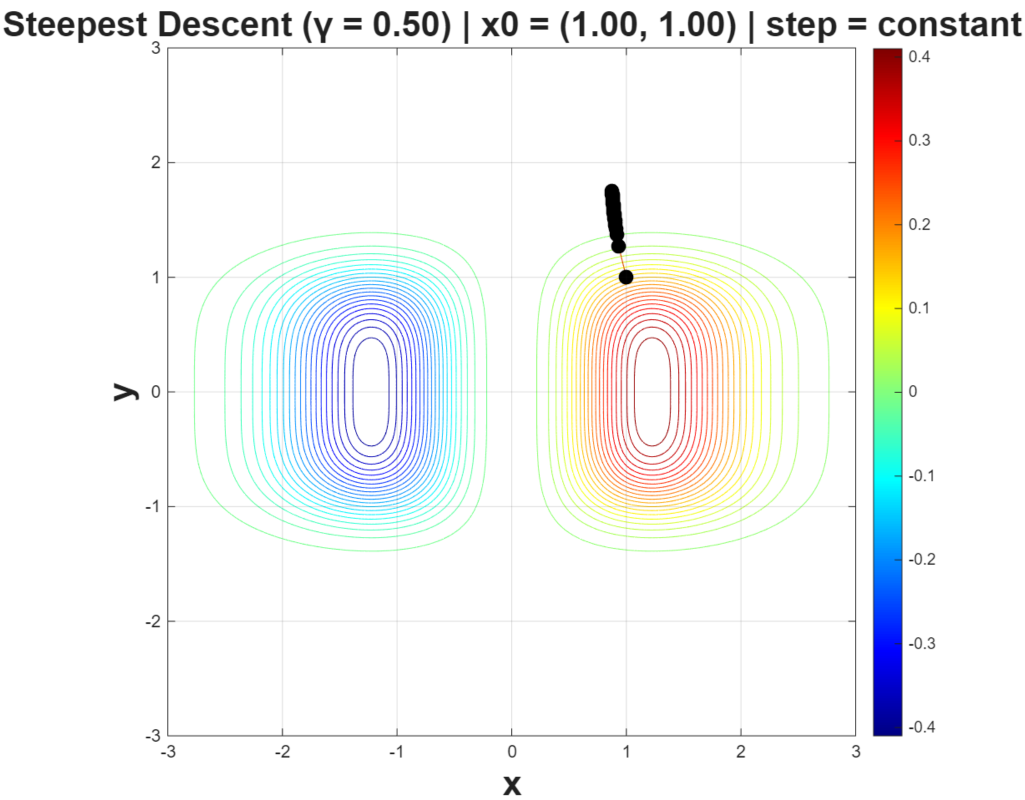


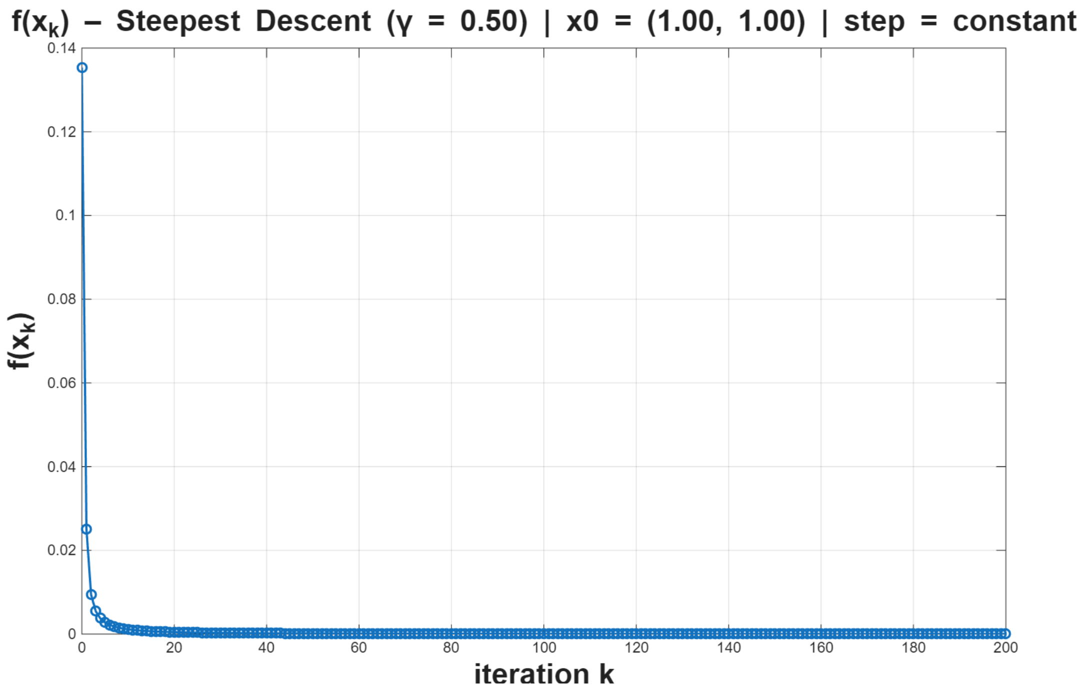
Σημειωση! Το σημειο $x_0 = (0,0)$, οπως ειπαμε παραπανω ειναι σαγματικο. Δηλαδη εκει ισχυει $\nabla f(0,0) = 0$. Επομενως, λογω του κριτηριου τερματισμου που εχει ο καθε αλγοριθμος, καμια μεθοδος με οποιαδηποτε επιλογη του γ δεν κανει κανενα βημα. Επομενως, δεν θα ξανα συμπεριλαβω κανενα πειραματικο αποτελεσμα αυτου του σημειου αφου ειναι ολα τα ιδια, για εξοικονομηση χωρου.

-Για $\gamma_k = 0.5$:

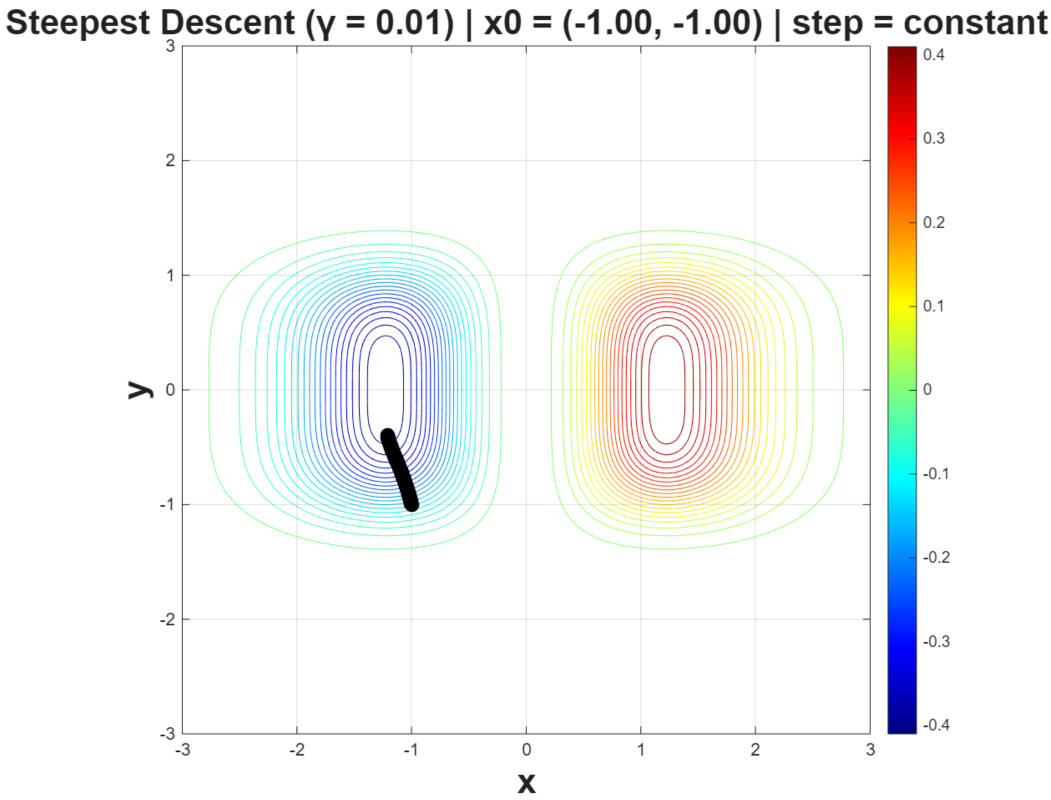


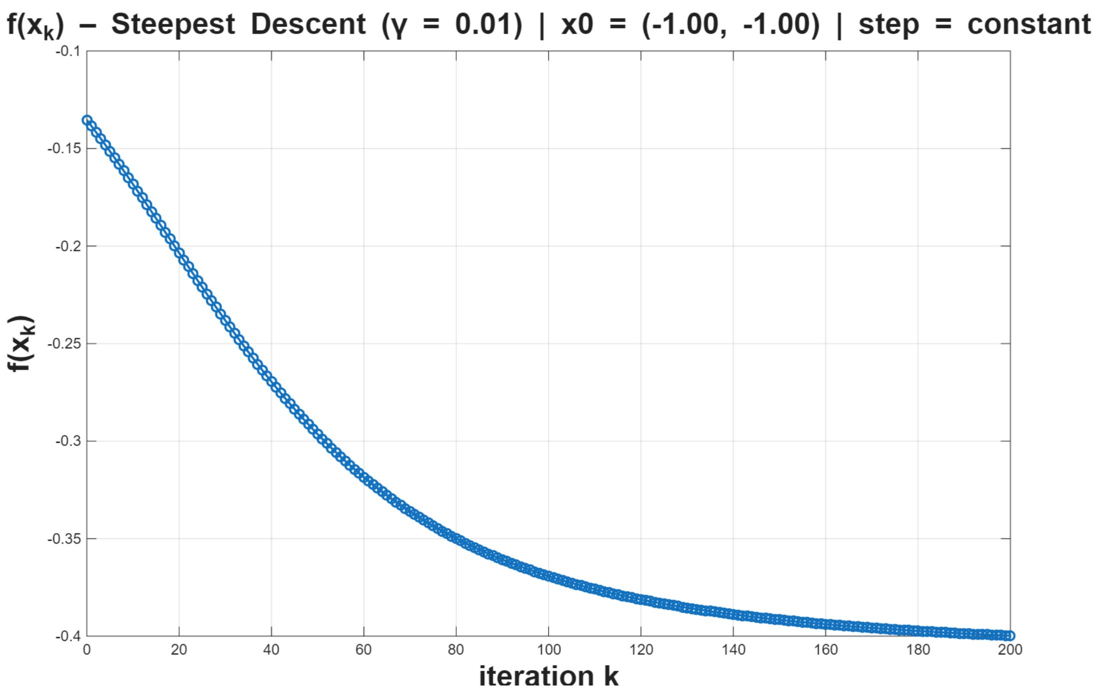


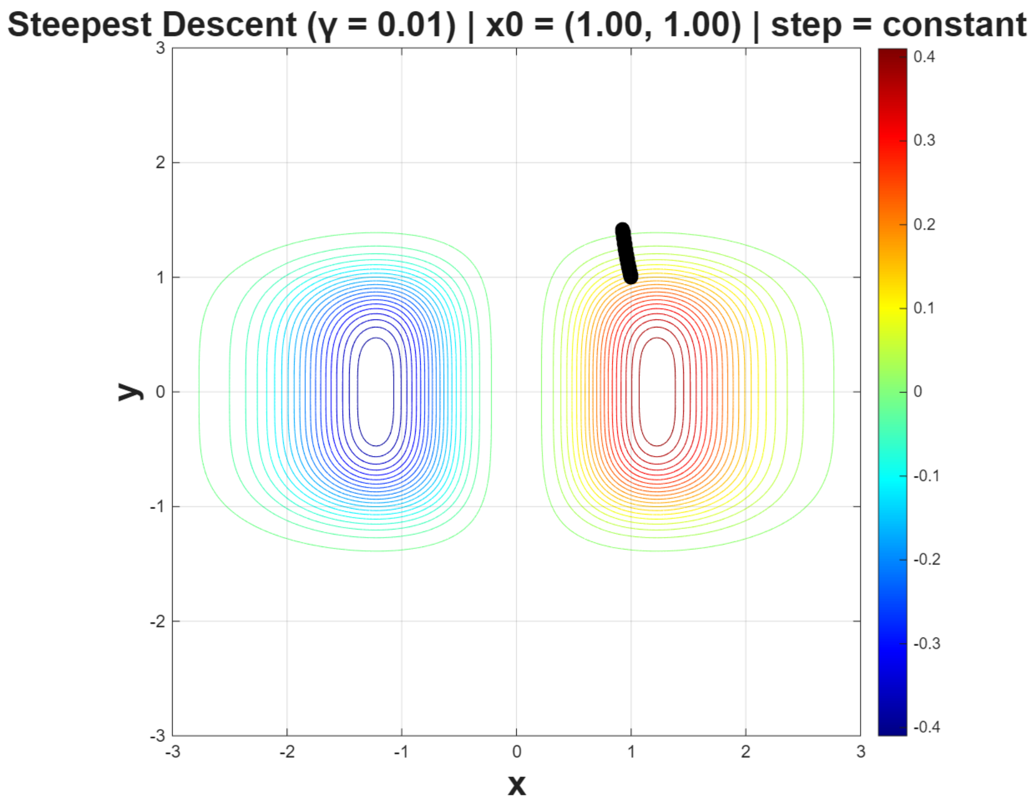

Steepest Descent (γ = 0.50) | x0 = (1.00, 1.00) | step = constant
x
y

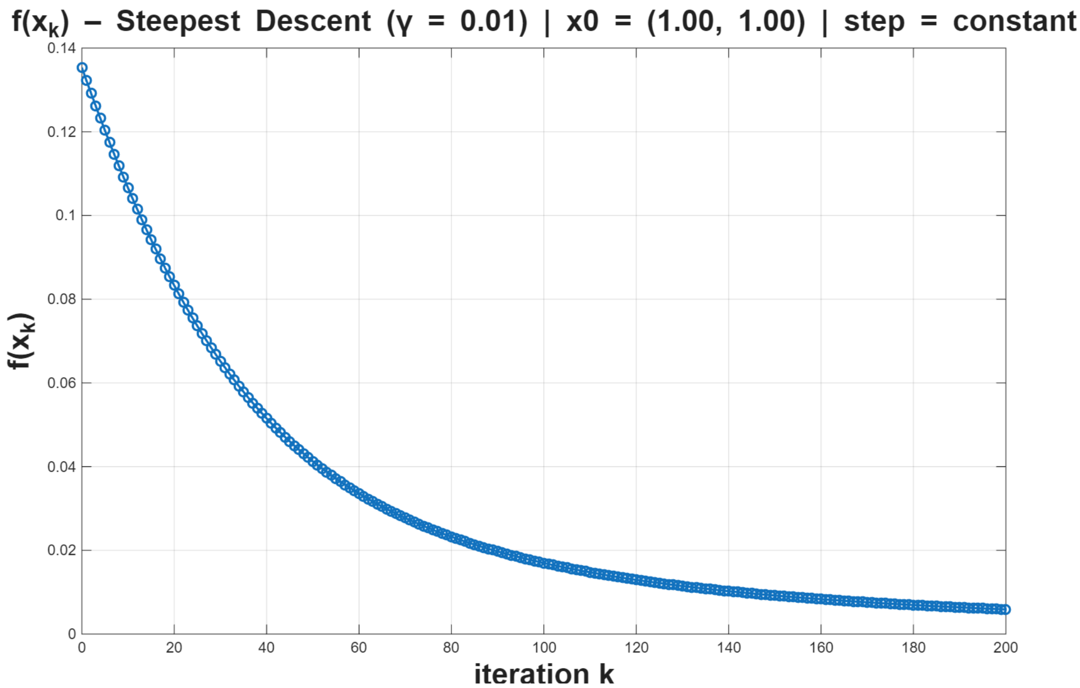

f(x_k) – Steepest Descent (γ = 0.50) | x0 = (1.00, 1.00) | step = constant
iteration k
f(x_k)

-Για $\gamma_k = 0.01$:



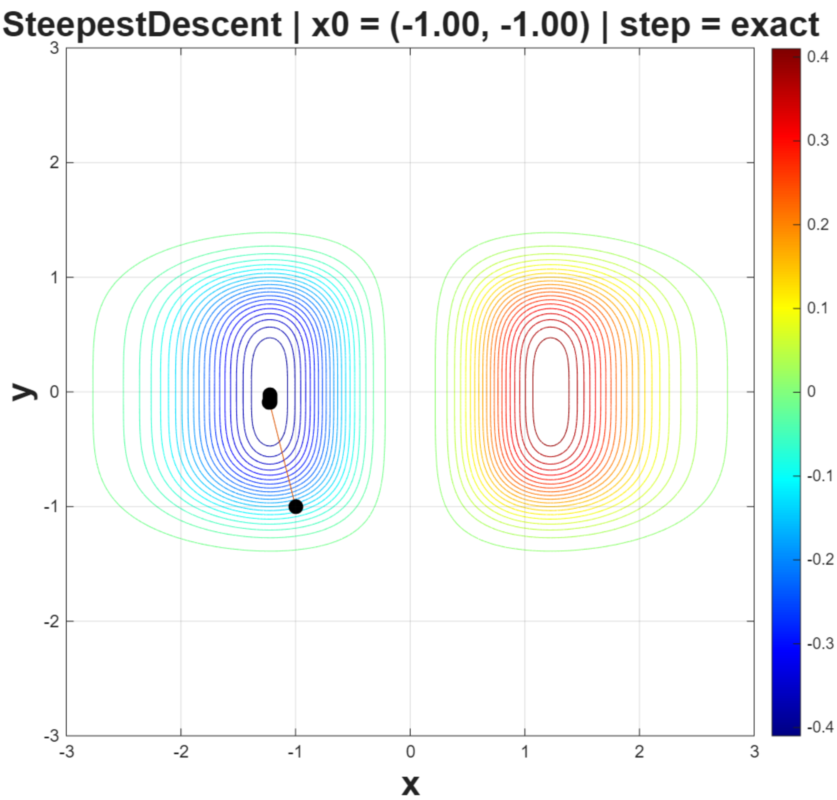


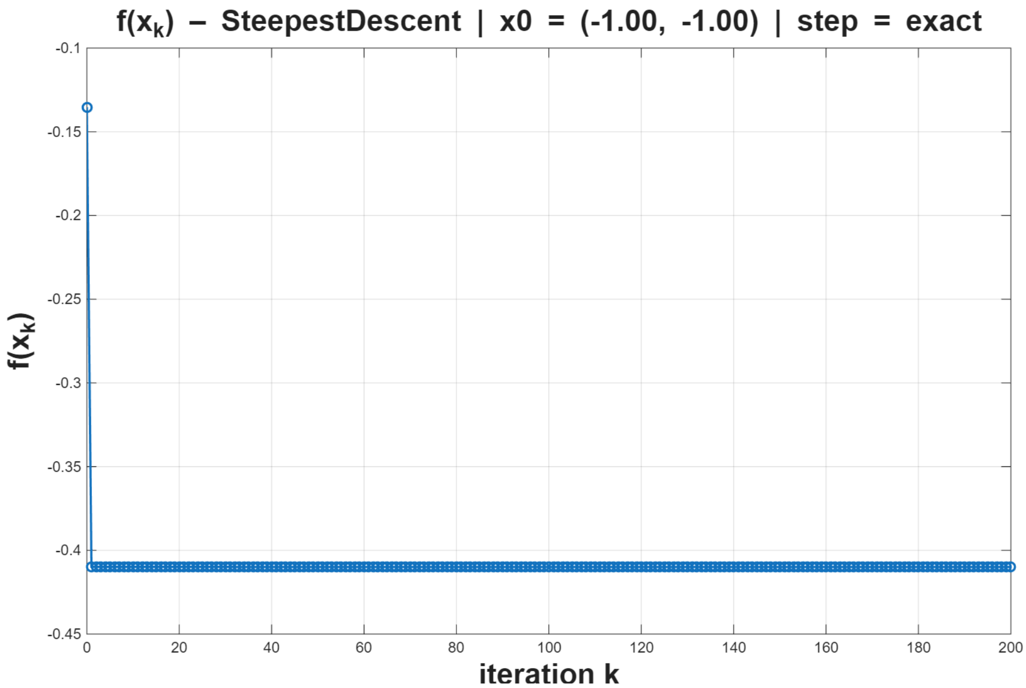


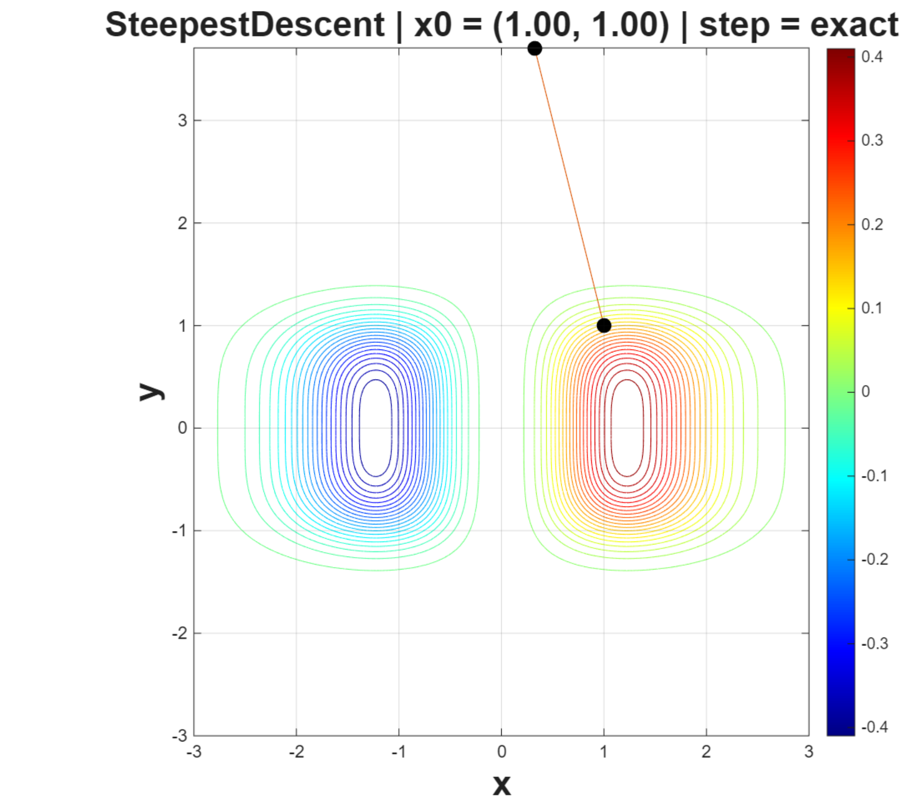


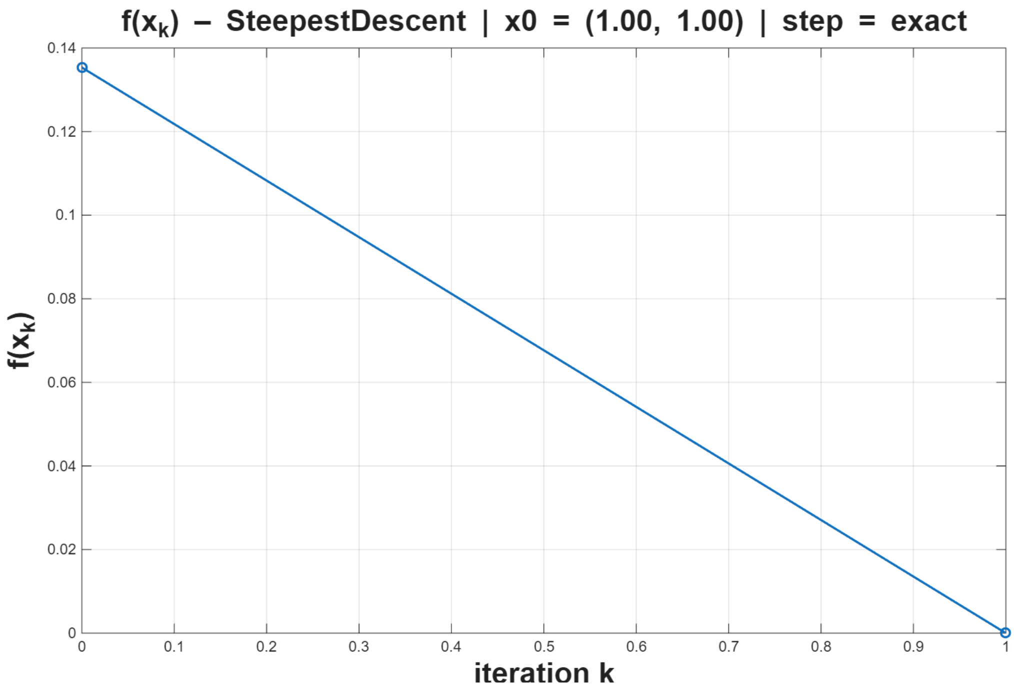
B - Βημα γ_k που ελαχιστοποιει την $\phi(\gamma_k) = f(x_k + \gamma_k d_k)$

Εχοντας αναλυσει την μεθοδο αυτη παραπανω, το μονο που θα πω εδω ειναι οτι για να βρω το γ_k που ελαχιστοποιει την σε ϕ καθε βημα, χρησιμοποιησα την μεθοδο του Χρυσου Τομεα που υλοποιησαμε στην προηγουμενη εργασια, με αρχικο διαστημα ευρους 5 και ακριβεια $l = 10^{-4}$. Επομενως, εαν η μεθοδος δεν κανει κανενα βημα, ειναι γιατι δεν βρηκε κανενα γ_k που να δινει μικροτερη τιμη στην f πανω στη γραμμη αναζητησης σε αυτο το διαστημα.


SteepestDescent | x0 = (-1.00, -1.00) | step = exact
x
y

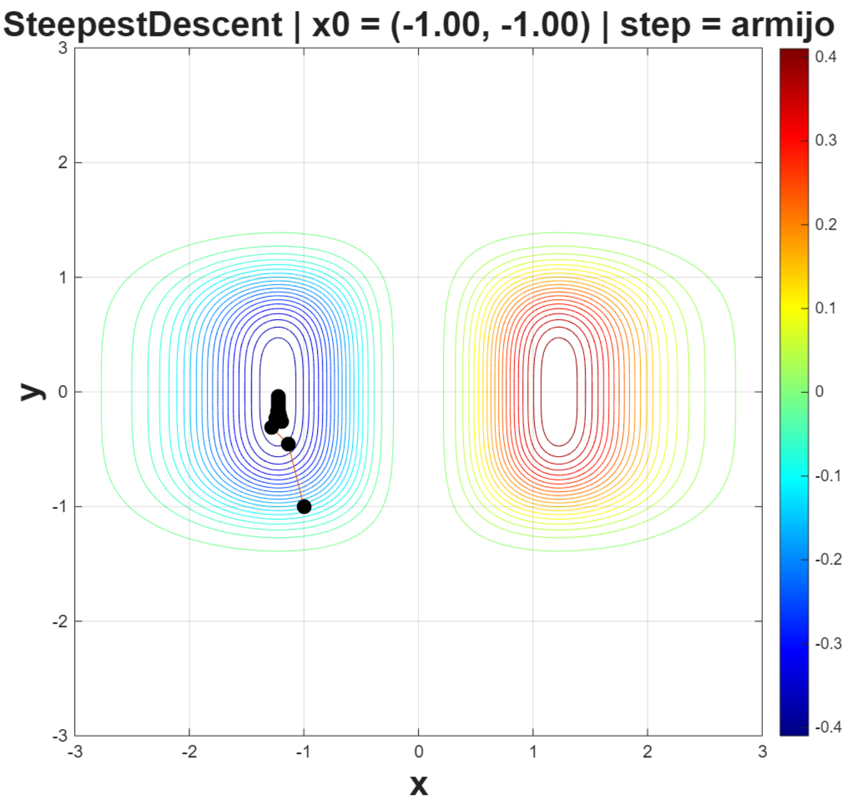

f(x_k) – SteepestDescent | x0 = (-1.00, -1.00) | step = exact
iteration k
f(x_k)

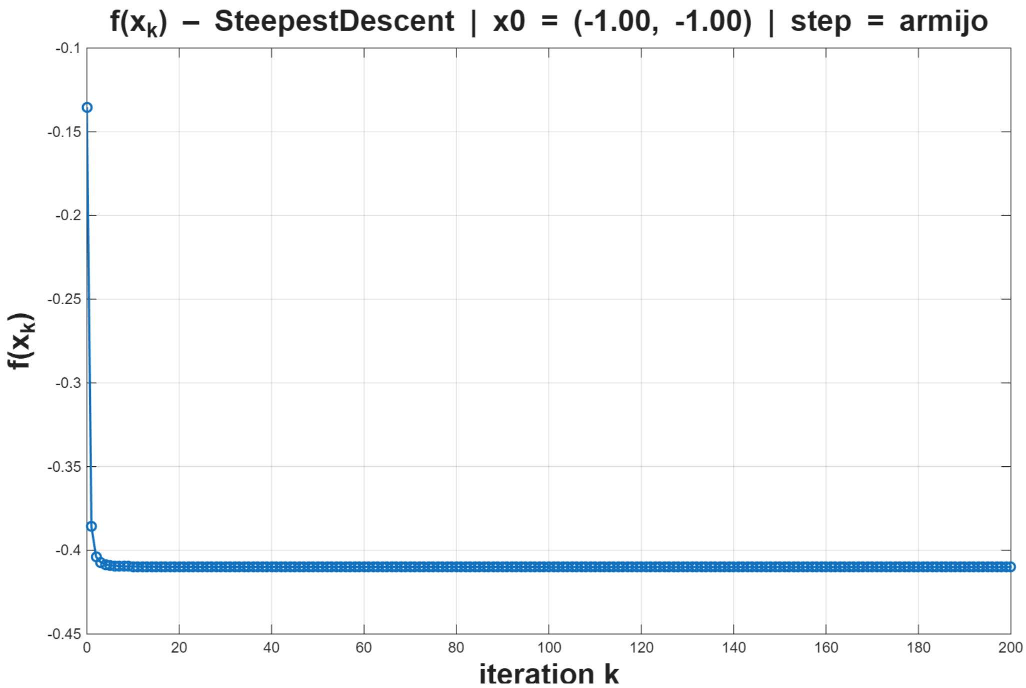


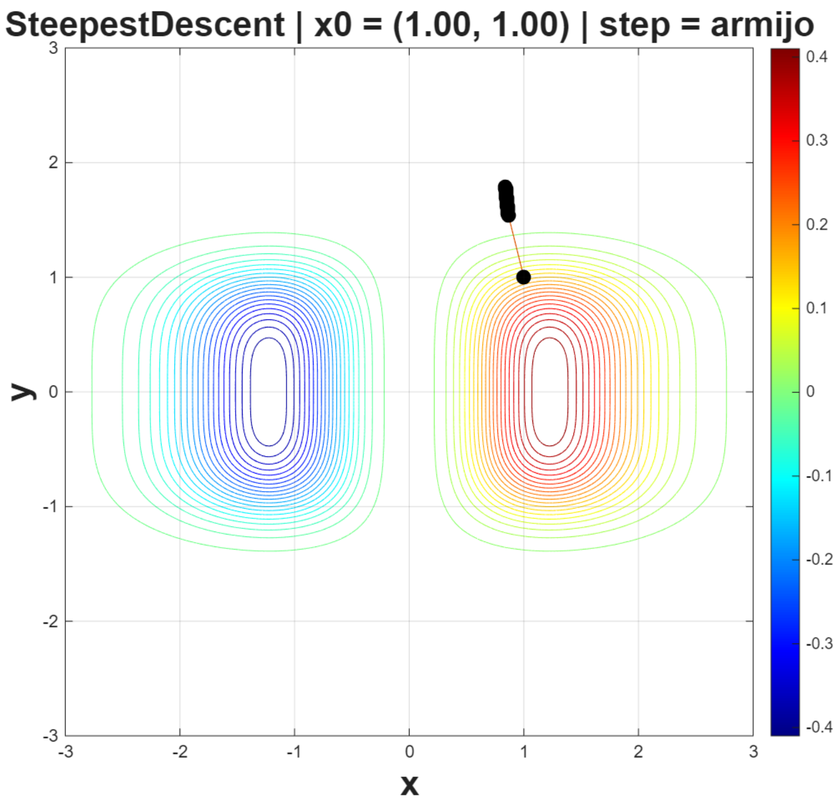


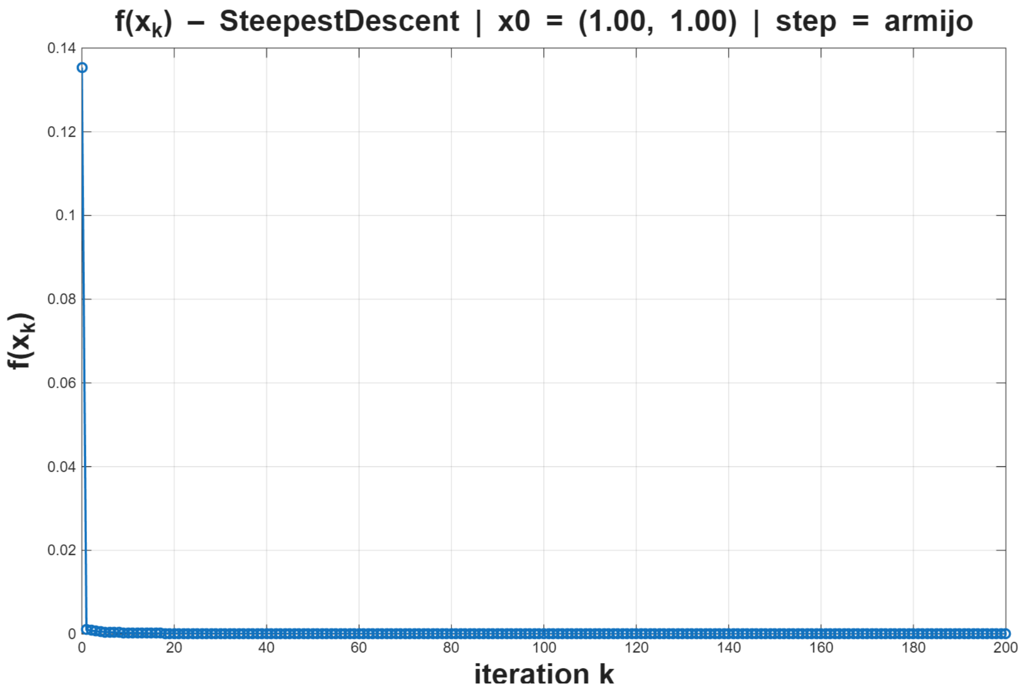
Γ - Βημα βασει του κανονα Armijo

Και αυτη η μεθοδος εχει αναλυθει παραπανω. Το μονο που θα πω ειναι οτι, το αρχικο βημα του αλγοριθμου το εθεσα μοναδα, δηλαδη $\gamma_0 = s = 1$, και τις παραμετρους α,β ισες με: $\alpha = 10^{-4}$, $\beta = 0.5$. Οι επιλογες αυτες βρισκονται στα διαστηματα που γραφει το βιβλιο.


SteepestDescent | x0 = (-1.00, -1.00) | step = armijo
y
x
3
2
1
0
-1
-2
-3
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4


f(x_k) – SteepestDescent | x0 = (-1.00, -1.00) | step = armijo
f(x_k)
iteration k
-0.1
-0.15
-0.2
-0.25
-0.3
-0.35
-0.4
-0.45
0
20
40
60
80
100
120
140
160
180
200





Σχόλια - Παρατηρήσεις

- Παρατηρουμε οτι οταν ξεκιναμε απο το σημειο $x_0 = (1, 1)$, ο αλγοριθμος σε ολες τις περιπτωσεις επιλογης του βηματος γ_k δεν συγκλινει στο ολικο ελαχιστο, αλλα φευγει προς τα πανω αριστερα. Αυτο οφειλεται στη γεωμετρια της συναρτησης και στην κατευθυνση του ∇f και κατα συνεπεια του

d_k. Μετα απο τα πρωτα βηματα, το gradient γινεται πολυ μικρο, και γι'αυτο βλεπουμε πολυ μικρα βηματακια.

- Στην περιπτωση του σταθερου βηματος, βλεπουμε οτι για $\gamma_k = 2$ ο αλγοριθμος σταματαει μολις με 1 βημα για σημειο εκκινησης το $(1, 1,$ ομως στο $(-1, -1)$ που βρισκομαστε κοντα στο ελαχιστο, εχουμε τεραστιες ταλαντωσεις και ο αλγοριθμος δεν συγκλινει ποτε. Με βημα $\gamma_k = 0.5$ φτανουμε πολυ κοντα στο ελαχιστο οταν $x_0 = (-1, -1)$ και μαλιστα πολυ νωρις, απο το 10 κι ολας βημα, οπως επισης και για $x_0 = (1, 1)$. Η συμπεριφορα του αλγοριθμου λοιπον ειναι μερικα αποτομα βηματα στην αρχη, αλλα μετα κολλαει γιατι το gradient εχει μικρυνει παρα πολυ, οποτε τα βηματα μου ειναι ασημαντα και η συγκλιση αργει. Για $\gamma_k = 0.01$ εχουμε την ιδια συμπεριφορα αλλα πολυ πιο αργα, οπως βλεπουμε και απο τα διαγραμματα των επαναληψεων, θελουμε πολυ περισσοτερα βηματα για να πλησιασουμε εκει που φτανει η μεθοδος με $\gamma_k = 0.5$ στις 200 επαναληψεις. Σε ολες σχεδον τις περιπτωσεις, το κριτηριο τερματισμου ειναι ο αριθμος των επαναληψεων. Ειναι προφανες, πως, οσο πιο μικρο σταθερο βημα βαλουμε τοσο πιο δυσκολα θα "περασουμε" το ελαχιστο, αλλα προφανως η συγκλιση θα ειναι ολο και πιο αργη.

- Στην exact μεθοδο, δηλαδη στην ευρεση του βηματος που ελαχιστοποιει την $\phi(\gamma)$ στην κατευθυνση του d_k, βλεπουμε οτι στο $(-1, -1)$ πλησιαζει πολυ στο ελαχιστο απο το πρωτο μολις βημα, ενω στο $(1, 1)$ κανει 1 μονο βημα και σταματαει λογω του κριτηριου τερματισμου $\|\nabla f\| < \varepsilon$.

- Στην περιπτωση του Armijo, βλεπουμε πολυ πιο σταθερα και ελεγχομενα βηματα καθως οπως ειπαμε ο κανονας αυτος εξασφαλιζει οτι θα εχουμε μια ικανοποιητικη μειωση της f σε καθε βημα.

Θέμα 3 - Μέθοδος Newton

Παρουσίαση Μεθόδου

Η μεθοδος αυτη χρησιμοποιει το αναπτυγμα Taylor 2ου βαθμου της αντικειμενικης συναρτησης. Η ιδεα της ειναι να ελαχιστοποιησει την προσεγγιση της f που δινεται απο το αναπτυγμα αυτο. Με τον τροπο αυτο, λαμβανει υποψην και την καμπυλοτητα της συναρτησης, μεσω του ορου $\nabla^2 f(x_k)$. Στη σελιδα 126 του βιβλιου, φαινεται πως παιρνοντας την τετραγωνικη προσεγγιση της f και θετοντας την παραγωγο της ιση με μηδεν προκυπτει οτι αυτη ελαχιστοποιειται οταν

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

. Αυτο λοιπον ειναι το διανυσμα κατευθυνσης που χρησιμοποιει αυτη η μεθοδος. Ο παραγοντας $\nabla^2 f(x_k)$ ειναι ενας πινακας που λεγεται Hessian τον οποιο και εχουμε υπολογισει παραπανω για την f. Προφανως βλεπουμε οτι πρεπει να ειναι αντιστρεψιμος ωστε να οριζεται η μεθοδος. Επισης, προκειμενου η κατευθυνση μας να ειναι καθοδικη, πρεπει να ισχυει:

$$\nabla^T f(x_k) d_k < 0 \Leftrightarrow -\nabla^T f(x_k)[\nabla^2 f(x_k)]^{-1} \nabla f(x_k) < 0$$

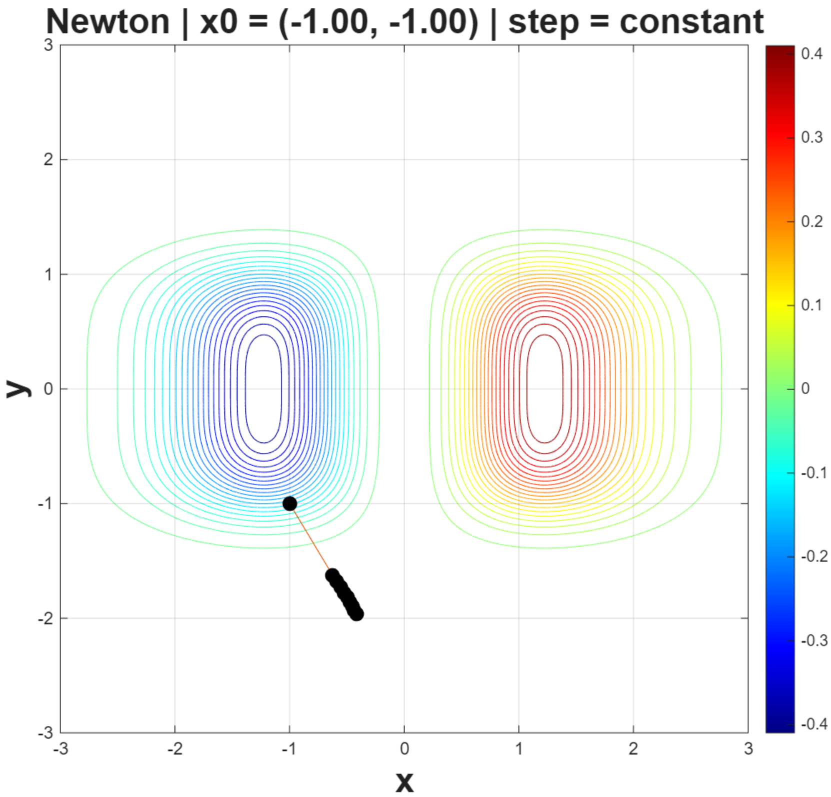
$$\Leftrightarrow \nabla^T f(x_k)[\nabla^2 f(x_k)]^{-1} \nabla f(x_k) > 0$$

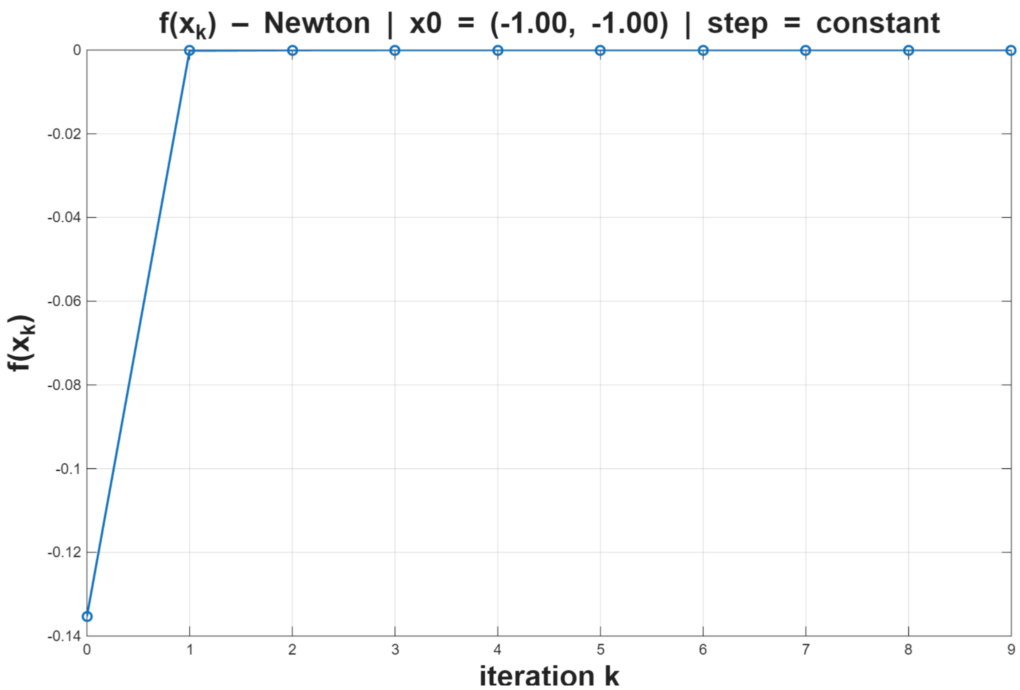
δηλαδη, πρεπει ο Hessian να ειναι θετικα ορισμενος. Δηλαδη να εχει μονο θετικες ιδιοτιμες. Αλλιως, ειναι πιθανο η κατευθυνση να γινει ανοδικη! Η μεθοδος Newton ειναι ιδανικη για τετραγωνικες συναρτησεις, καθως αποδεικνυεται (σελ 127) οτι συγκλινει με μολις ενα βημα!

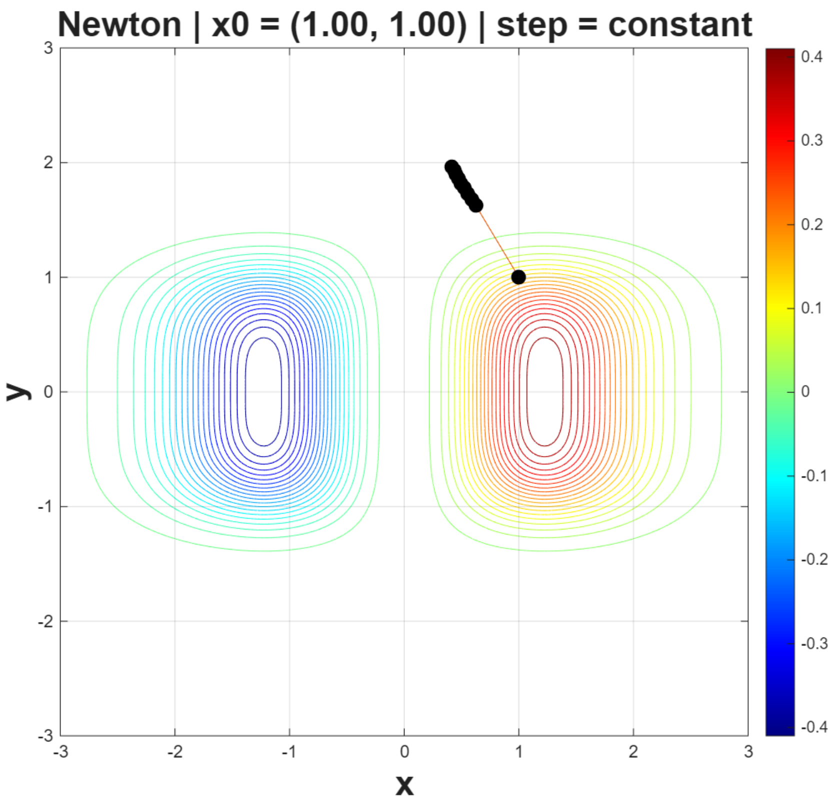
Πειραματικα αποτελεσματα:

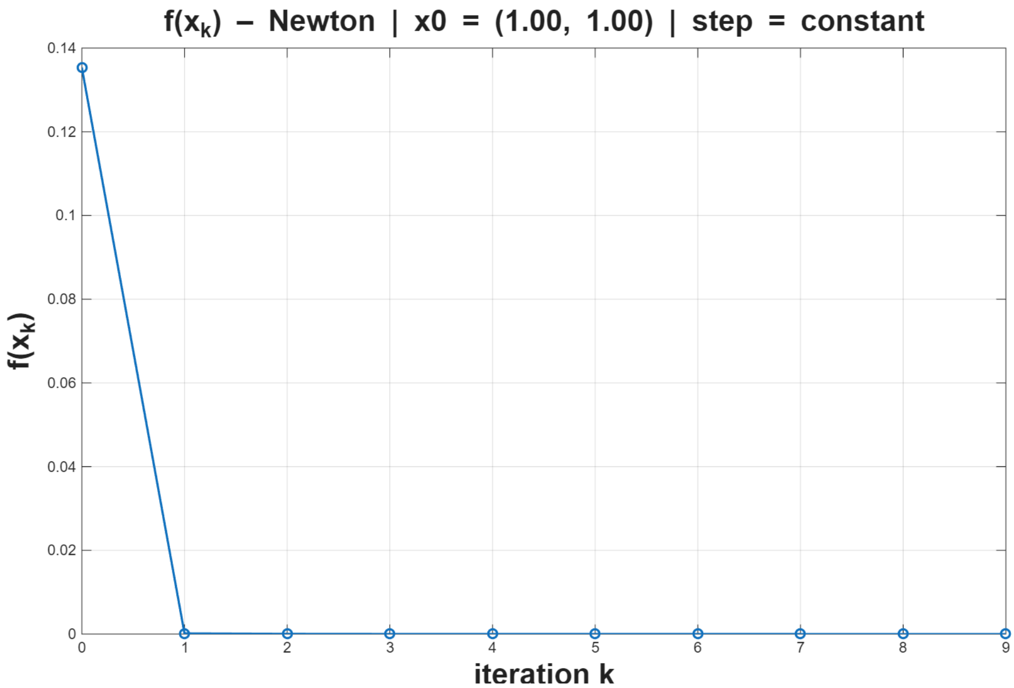
A - Σταθερο βημα γ_k :

Εδω διαλεξα μονο την τιμη $\gamma_k = 1$ ως σταθερο βημα.

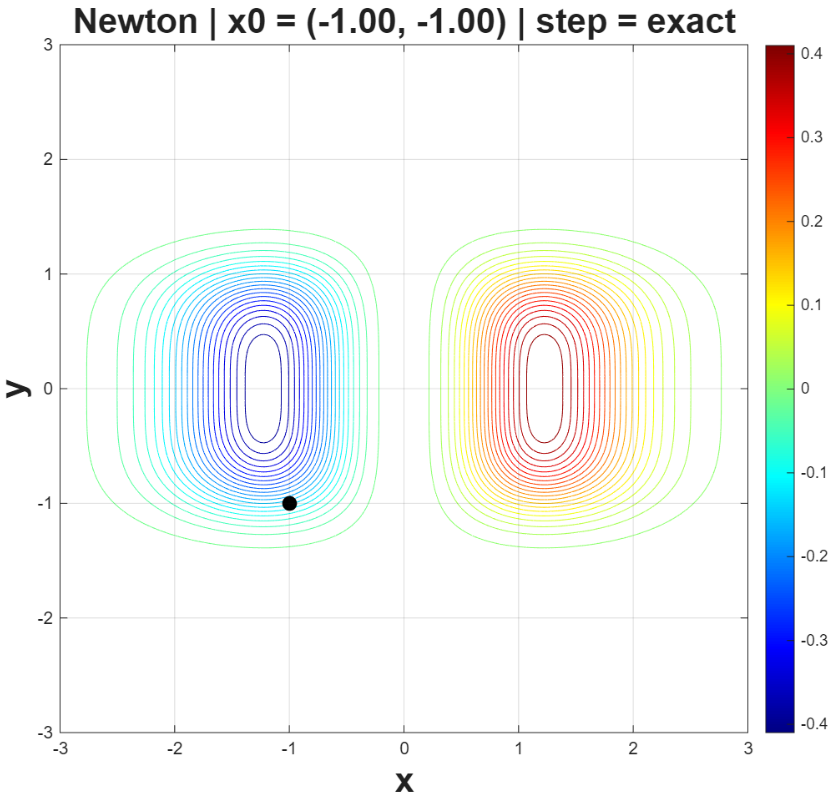


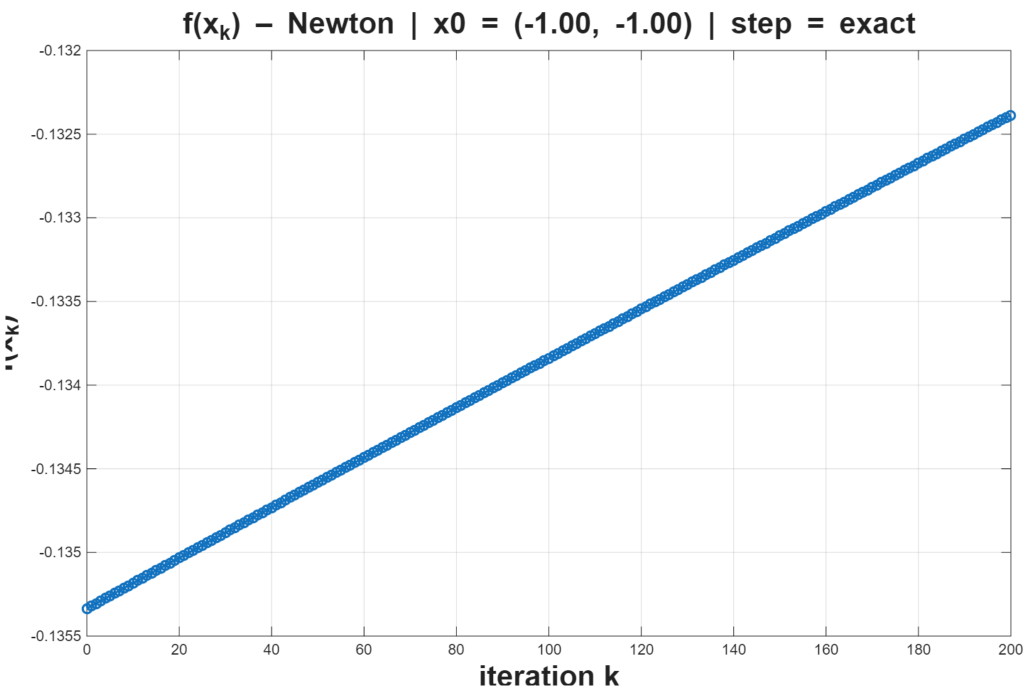


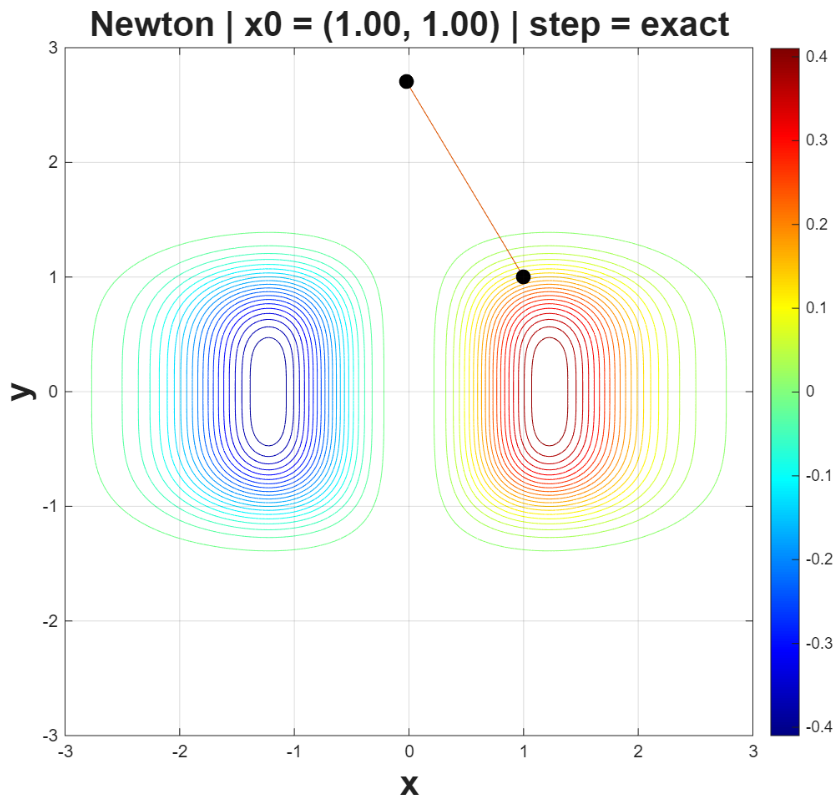

Newton | x0 = (1.00, 1.00) | step = constant
x
y
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4

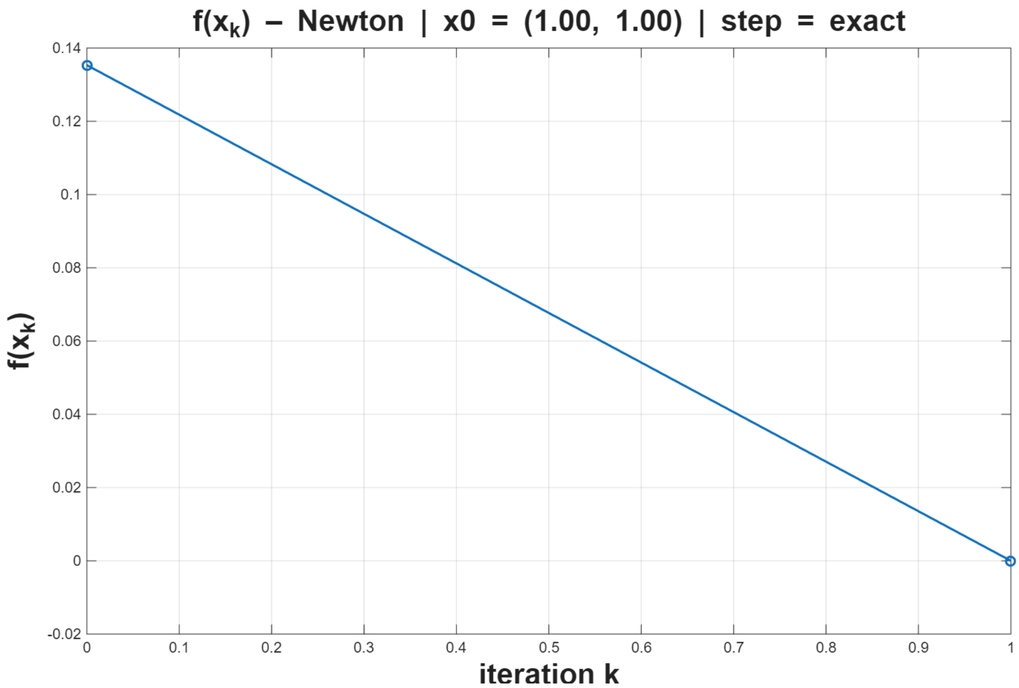

f(x_k) – Newton | x0 = (1.00, 1.00) | step = constant
f(x_k)
iteration k
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
1
2
3
4
5
6
7
8
9

B - Βημα γ_k που ελαχιστοποιει την $\phi(\gamma_k) = f(x_k + \gamma_k d_k)$



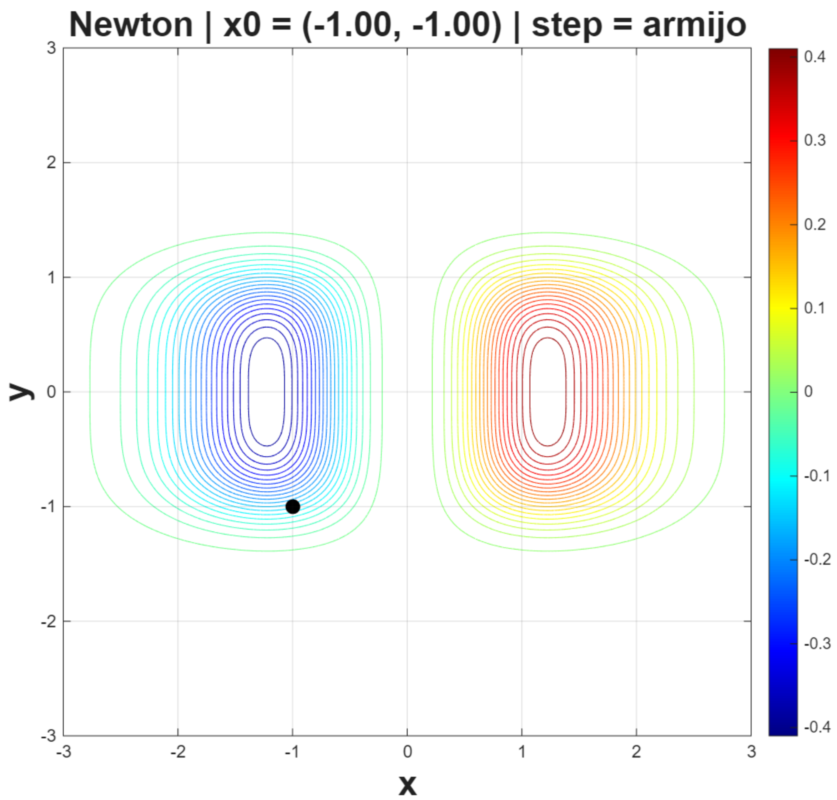


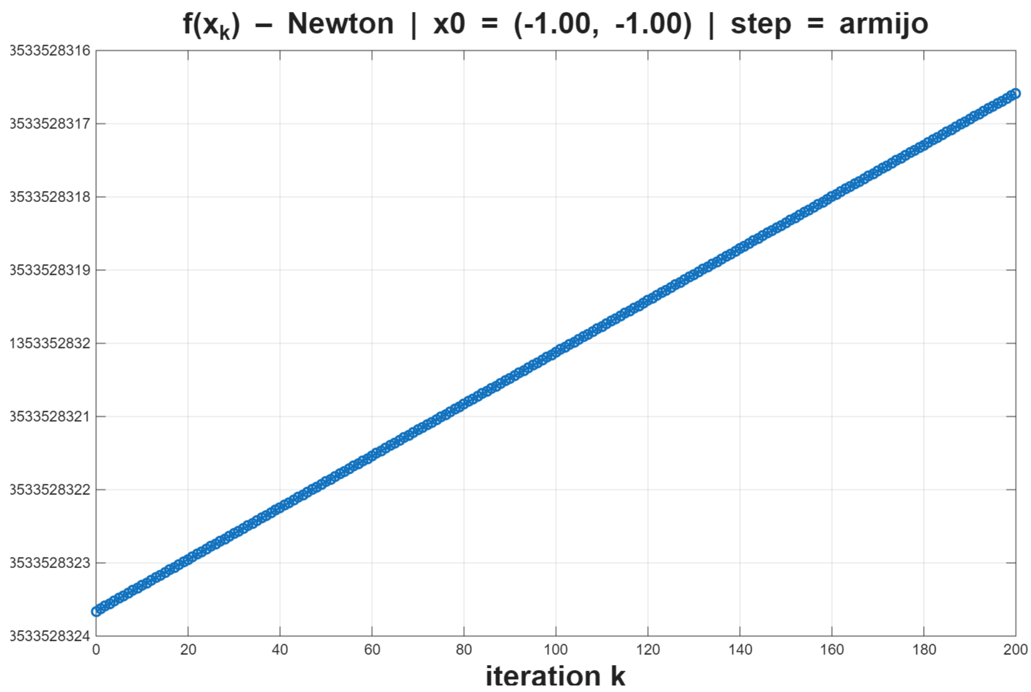

Newton | x0 = (1.00, 1.00) | step = exact
y
x
3
2
1
0
-1
-2
-3
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4

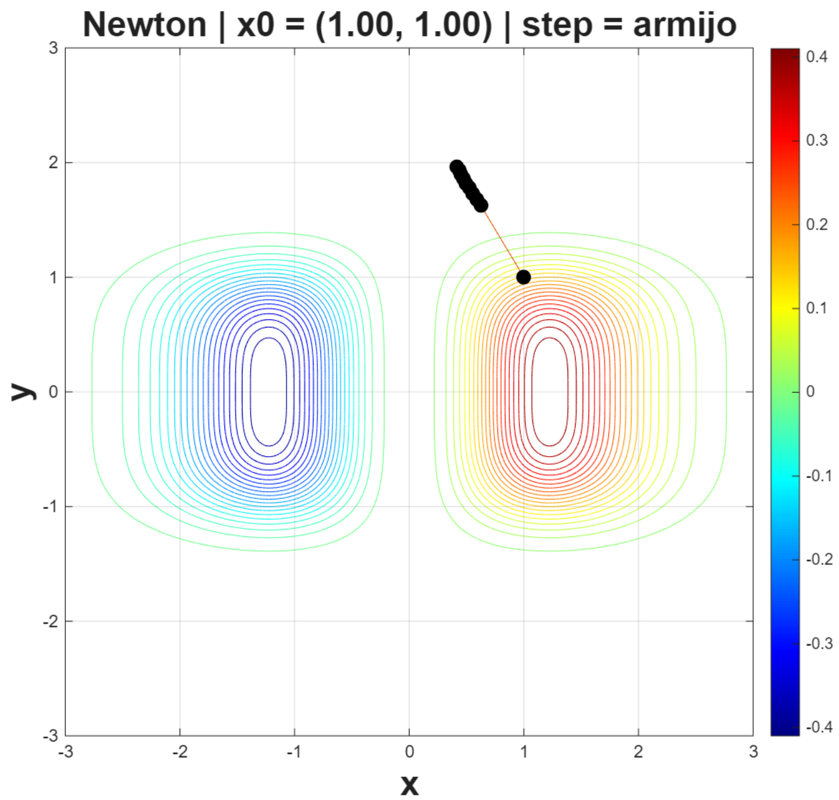

f(x_k) – Newton | x0 = (1.00, 1.00) | step = exact
f(x_k)
iteration k
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

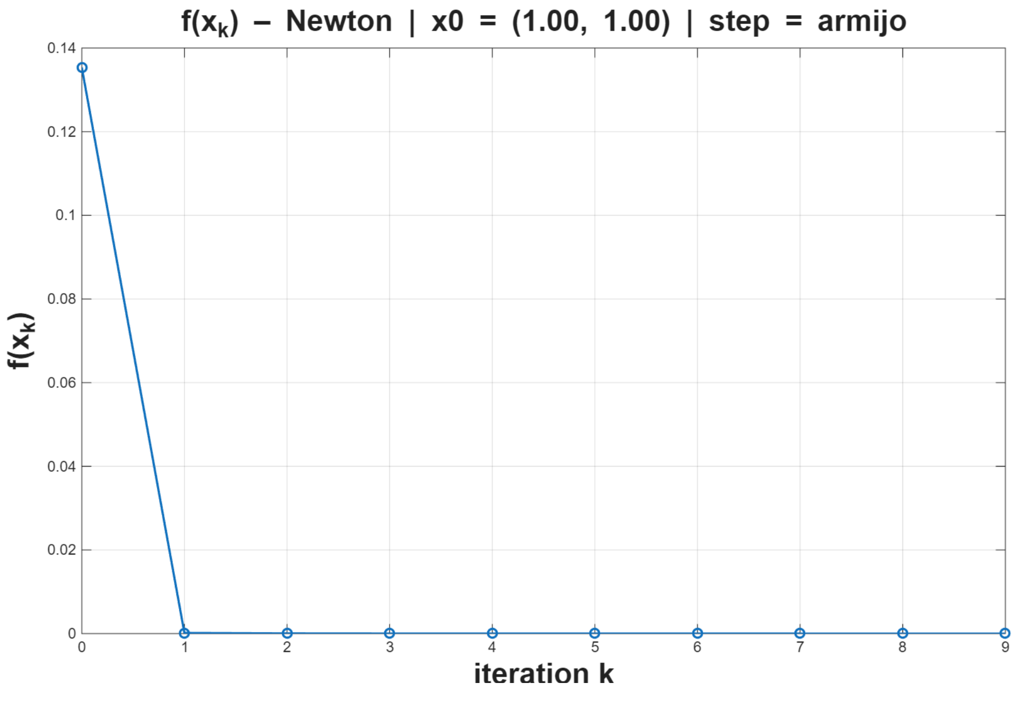
Γ - Βημα βασει του κανονα Armijo

Ιδιες παραμετροι με Θεμα 2









Σχόλια - Παρατηρήσεις

- Οπως βλεπουμε, στο σημειο $x_0 = (-1, -1)$ η α περιπτωση με το σταθερο βημα εκανε βηματα ανοδικα! Αυτο συμβαινει, γιατι στο σημειο αυτο ο εσσιανος πινακας ειναι:

$$H(-1,-1) = \begin{bmatrix} 4e^{-2} & 4e^{-2} \\ 4e^{-2} & -4e^{-2} \end{bmatrix}$$

για τον οποιο ευκολα βλεπουμε (ειτε μεσω της οριζουσας του που ειναι αρνητικη, ειτε υπολογιζοντας τις ιδιοτιμες απο $Hx = \lambda x$ οι οποιες βγαινουν αντιθετες, περιπου ± 0.7656) οτι δεν ειναι θετικα ορισμενος. Η exact δεν εκανε κανενα βημα γιατι για ευρος 5 δεν βρηκε μικροτερη τιμη της f πανω στην γραμμη αναζητησης. Ωστοσο οταν δοκιμασα να βαλω το ευρος ισο με 10, εκανε ενα μεγαλο βημα και εφτασε πολυ κοντα στην σαγματικη γραμμη $x = 0$. Δεν το συμπεριελαβα γιατι μας παει πολυ πιο μακρια απο το πραγματικο ελαχιστο, το οποιο δεν θελουμε. Η Armijo δεν κανει κανενα βημα γιατι, λογω ακριβως του οτι "κοιταει" σε κατευθυνση ανοδου, δεν βρισκει την ικανοποιητικη μειωση της f που θελει (με αρχικο βημα $s = \gamma_0 = 1$).

- Στο σημειο $x_0 = (1, 1)$ εχουμε:

  $$H(1,1) = \begin{bmatrix} -4e^{-2} & -4e^{-2} \\ -4e^{-2} & 4e^{-2} \end{bmatrix}$$

  οπου παλι ευκολα βλεπουμε οτι δεν ειναι θετικα ορισμενος (εχει το ιδιο ζευγαρι αντιθετων ιδιοτιμων). Ωστοσο, η κατευθυνση κινησης ειναι καθοδικη και αυτο μας δειχνει οτι, το να μην ειναι θετικα ορισμενος δεν σημαινει αναγκαστικα οτι η κατευθυνση ειναι ανοδος, γιατι εξαρταται και απο την κατευθυνση του gradient. Μαζι καθοριζουν το τελικο προσημο του $\nabla^T f(x_k) d_k$, το οποιο καθοριζει αν η κατευθυνση θα ειναι ανοδικη, καθοδικη ή σταθερη. Εφοσον βλεπουμε οτι η κατευθυνση ειναι καθοδικη, σημαινει οτι το προσημο ειναι αρνητικο και γι'αυτο και οι 3 περιπτωσεις επιλογης του γ_k κανουν καθοδικα βηματα.

- Τελος, οσον αφορα τις επαναληψεις του αλγοριθμου, βλεπουμε ξεκαθαρα οτι ο αλγοριθμος ,στην περιπτωση του σταθερου βηματος και του Armijo, τερματιζεται μετα απο μολις 9 επαναληψεις, λογω του κριτηριου με το gradient. Αυτο οφειλεται στο διαφορετικο διανυσμα κατευθυνσης που χρησιμοποιει. Εξαιρεσεις αποτελεσαν η exact που στο $(-1, -1)$ εκανε 200 επαναληψεις προσπαθωντας να βρει καποια μικροτερη τιμη της συναρτησης στη γραμμη και στο διαστημα αναζητησης, ενω στο $(1, 1)$ τερματιστηκε μετα απο μολις μια επαναληψη πολυ κοντα στη σαγματικη γραμμη (προφανως επειδη το gradient εγινε μικροτερο του ε), οπως επισης και ο Armijo στο $(-1, -1)$ που σε καθε επαναληψη μειωνε το γ_k προσπαθωντας να βρει την ικανοποιητικη μειωση, την οποια δεν βρηκε ποτε στις 200 επαναληψεις.

- Αρα βλεπουμε το πως αυτη η μεθοδος ξεπερνα την Steepest Descent σε θεμα ταχυτητας συγκλισης, με σοβαρο μειονεκτημα ομως το κριτηριο θετικης ορισιμοτητας του Hessian, το οποιο εδω προκαλεσε την κινηση σε ανοδικη πορεια!

Θέμα 3 - Μέθοδος Levenberg-Marquardt

Παρουσίαση Μεθόδου

Η μεθοδος αυτη χρησιμοποιει ως διανυσμα κατευθυνσης το εξης:

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

οπου ο μοναδιαιος πινακας και μ_k θετικη σταθερα. Στην ουσια μεσω της προσθεσης του πινακα $\mu_k I$ στον Hessian, προσπαθουμε να κανουμε τον πινακα $\nabla^2 f(x_k) + \mu_k I$ θετικα ορισμενο παντου. Στις σελιδες 138-139 του βιβλιου, αποδεικνυεται οτι αν επιλεξουμε $\mu_k > \bar{\mu_k}$, οπου $\bar{\mu_k}$ η μεγαλυτερη, κατα απολυτη τιμη, ιδιοτιμη του Hessian ο πινακας $\nabla^2 f(x_k) + \mu_k I$ ειναι θετικα ορισμενος.

Αξιζει να σημειωθει οτι αν το μ_k ειναι πολυ μεγαλο, ο παραγοντας $\mu_k I$ υπερισχυει του Hessian, δηλαδη:

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

δηλαδη η μεθοδος λειτουργει σαν την Steepest Descent με $\gamma_k = \frac{1}{\mu_k}$, ενω αν το μ_k ειναι πολυ μικρο, τοτε ο Hessian υπερισχυει του $\mu_k I$ οποτε:

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

και η μεθοδος λειτουργει σαν τη Newton. Γι'αυτο λοιπον η μεθοδος LM λεγεται και τροποποιημενος Newton. Επειδη εκμεταλλευεται την ιδιοτητα του Newton αλλα με θετικα ορισμενο πινακα παντου.

Σημείωση!!

Προσθετοντας τον πινακα $\mu_k I$, στην ουσια προσθετουμε μ στις ιδιοτιμες του Hessian. Προσπαθουμε να τις κανουμε ολες θετικες ωστε να ειναι θετικα ορισμενος. Στο βιβλιο γραφει, πως το $\bar{\mu_k}$ ισουται με την απολυτη τιμη της μεγαλυτερης ιδιοτιμης του Hessian. Νομιζω πως αυτο ειναι λαθος (τουλαχιστον εγω το παρερμηνευσα στην αρχη), γιατι αν ειχαμε πχ τις ιδιοτιμες 3,-2,-7 τοτε η μεγαλυτερη ιδιοτιμη ειναι η 3 και η απολυτη τιμη αυτης ειναι 3. Προσθετοντας 3 ομως, η αρνητικη ιδιοτιμη γινεται -4 αρα ο πινακας παραμενει μη θετικα ορισμενος.

Αυτο που μαλλον εννοει το βιβλιο ειναι "η μεγαλυτερη κατα απολυτη τιμη ιδιοτιμη", δηλαδη να παρουμε πρωτα τις απολυτες τιμες των ιδιοτιμων και μετα το μεγιστο αυτων. Ετσι, στο παραδειγμα μας θα παρουμε τις απολυτες τιμες: 2,3,7 εκ των οποιων η μεγιστη ειναι το 7. Οποτε θα προσθεσουμε 7 συν κατι ακομα για να γινουν ολες οι αρχικες ιδιοτιμες θετικες.

Παρολο που αυτο ειναι μια σιγουρη λυση, δεν ειναι τοσο αποδοτικη γιατι, αν για παραδειγμα εχω μια μεγαλη ιδιοτιμη, π.χ: 130,7,-5, τοτε η μεγιστη κατα απολυτη τιμη ειναι 130, οποτε θα προσθεσουμε 130 συν κατι, ενω θα αρκουσε να προσθεταμε 5 συν κατι. Αυτο μας οδηγει στην επιλογη ενος πολυ μεγαλου μ το οποιο κανει την μεθοδο LM να τεινει προς την Steepest Descent.

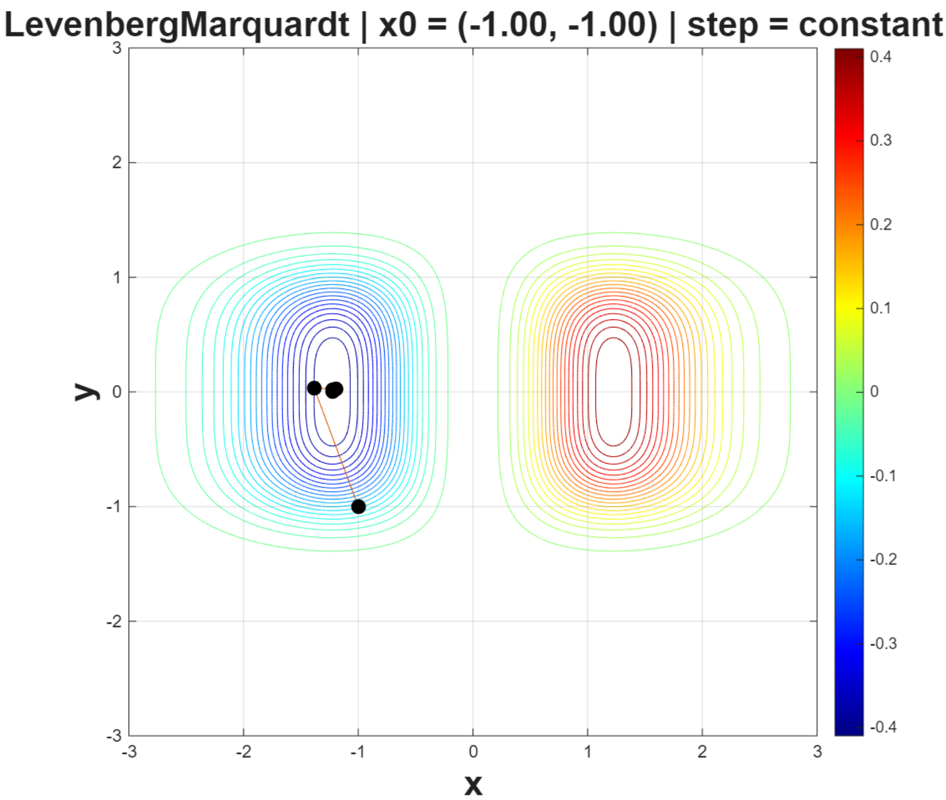
Επειδη θελω να αξιοποιησω την Newton, λογω της ταχυτητας συγκλισης της, θα υλοποιησω τον αλγοριθμο αλλιως. Θα βαλω πρωτα εναν ελεγχο για το αν ο Hessian ειναι θετικα ορισμενος (αν ειναι, θετω $\mu_k = 0$), και στην περιπτωση που δεν ειναι θα θετω: $\bar{\mu}_k = -\lambda_{min}$, οπου λ_{min}. Ετσι, οταν ο Hessian ειναι σιγουρα θετικα ορισμενος, η μεθοδος θα ταυτιζεται με την Newton, και αν δεν ειναι, για παραδειγμα οι ιδιοτιμες ειναι 10,3,-2, θα προσθετω 2 συν κατι ακομα.

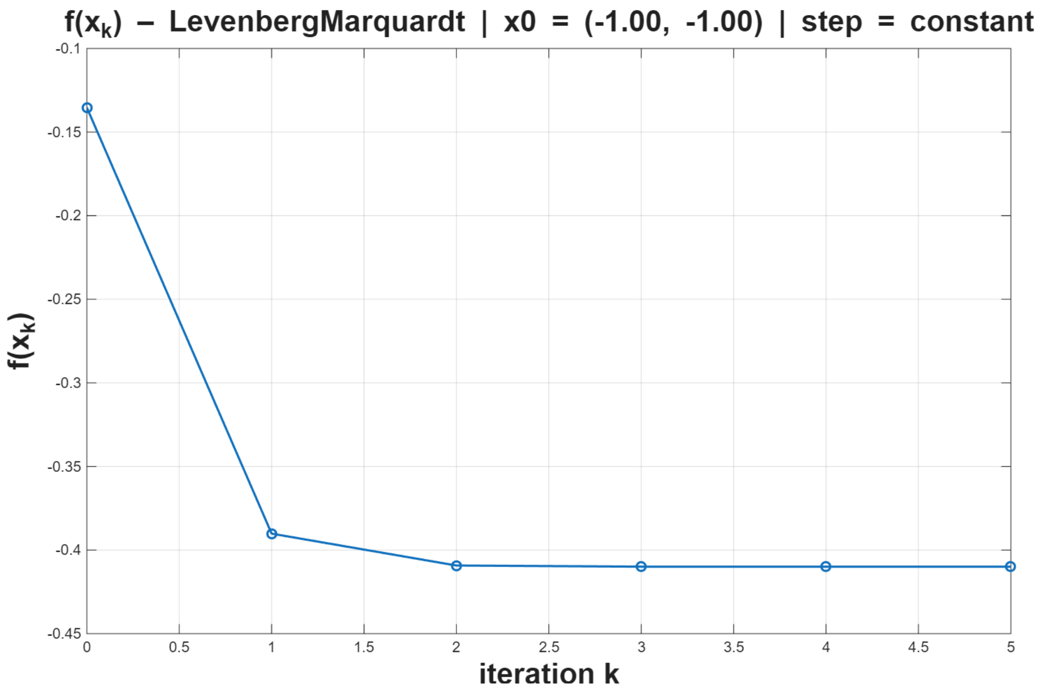
Οπως διαπιστωσα βεβαια, στο προβλημα μας οι 2 υλοποιησεις κανουν ακριβως το ιδιο, γιατι ο Hessian εχει και στα 2 σημεια εκκινησης το ιδιο ζευγαρι αντιθετων ιδιοτιμων.

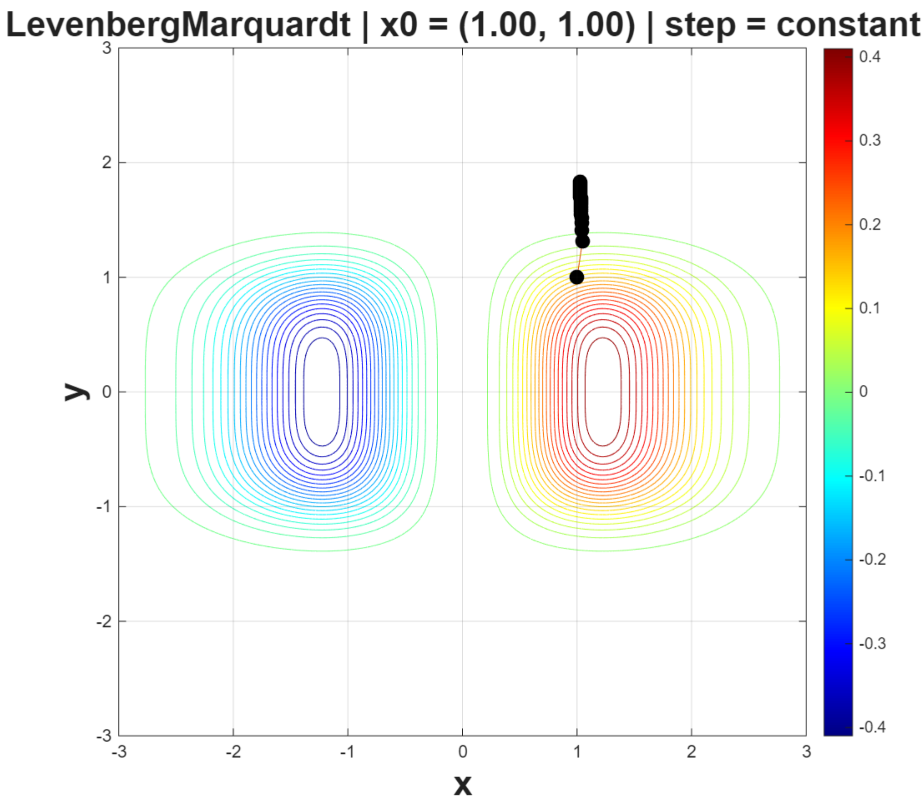
Πειραματικά Αποτελέσματα:

A - Σταθερο βημα γ_k

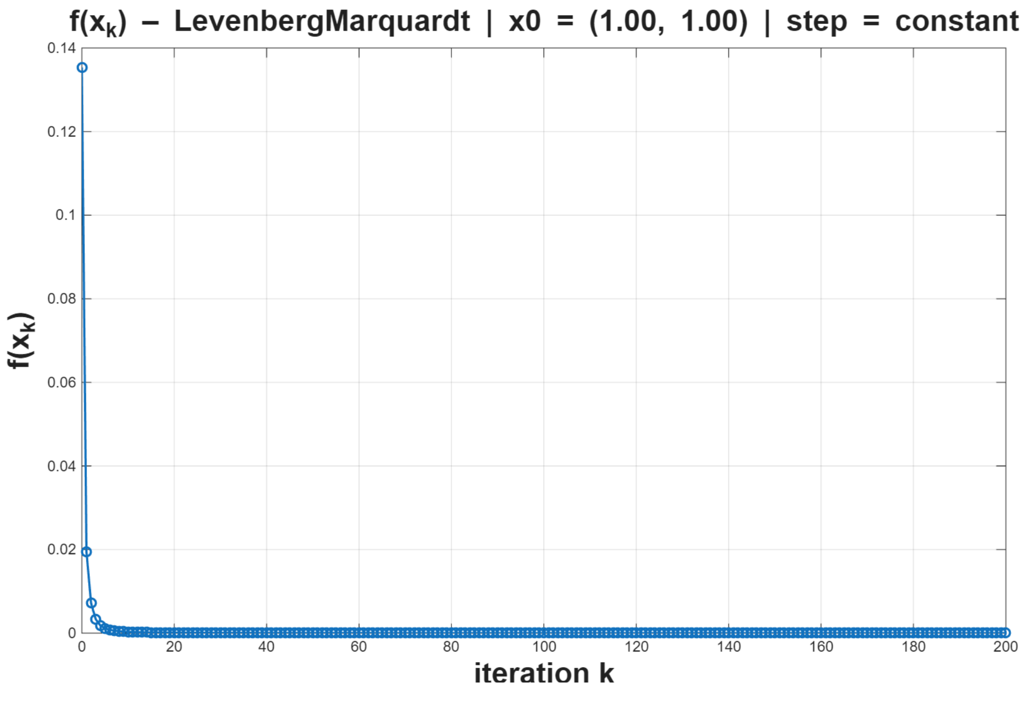
Παλι επελεξα $\gamma_k = 1$.





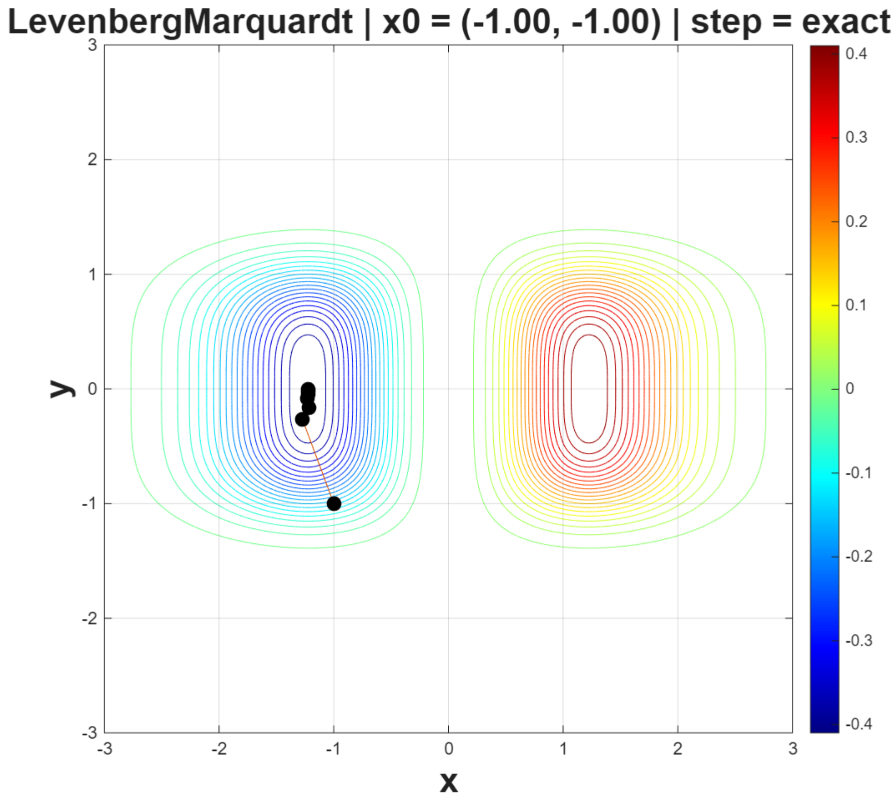


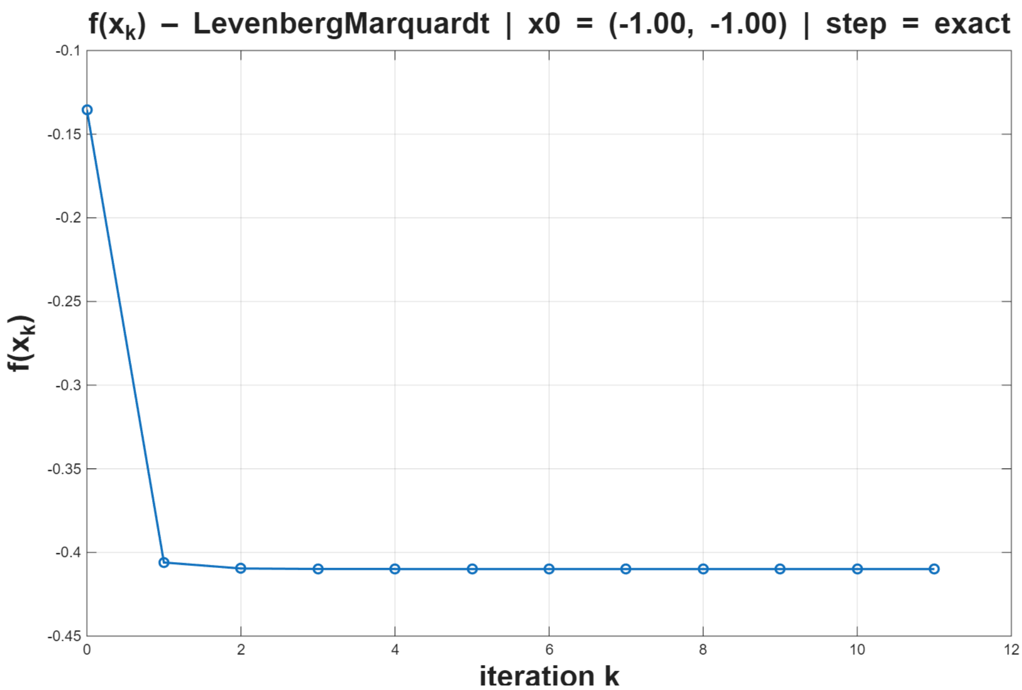
LevenbergMarquardt | x0 = (1.00, 1.00) | step = constant
y
x

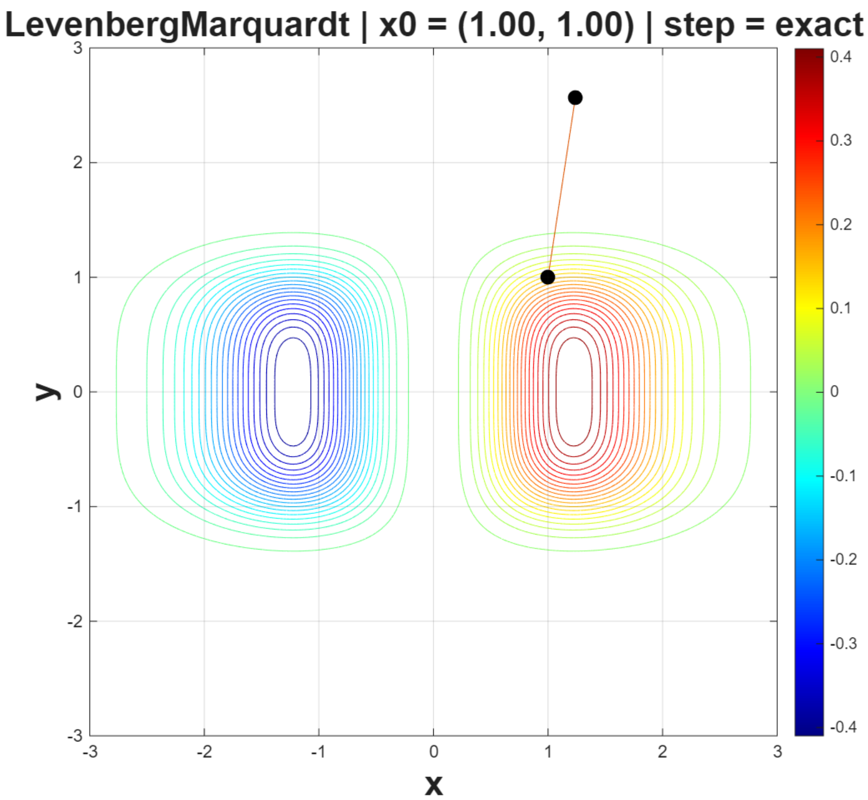


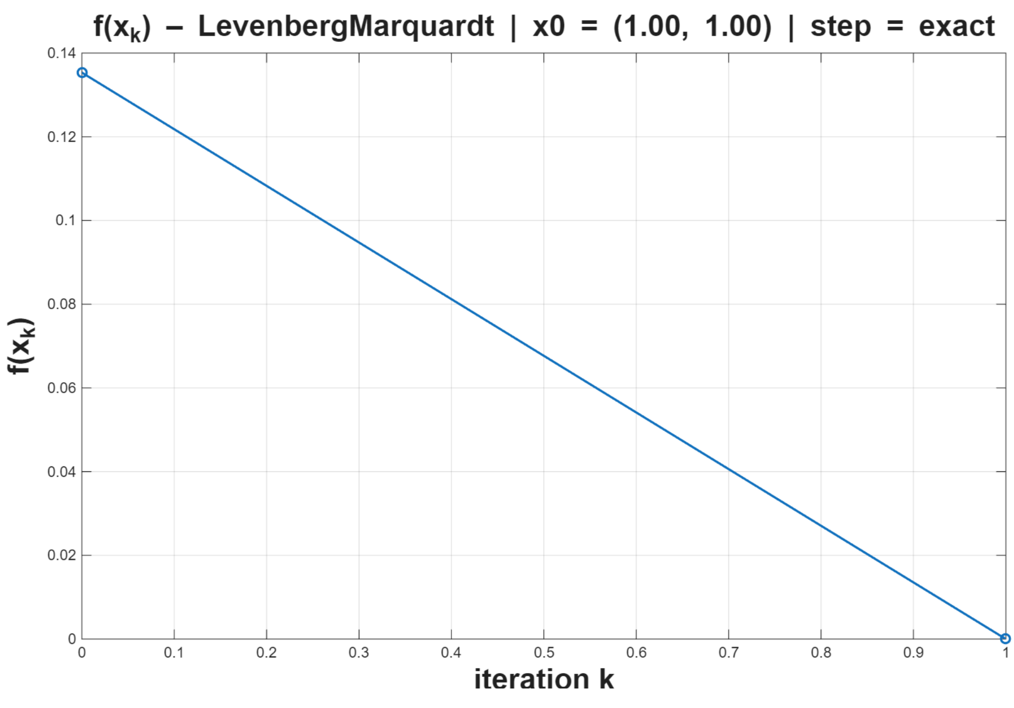
f(xk) – LevenbergMarquardt | x0 = (1.00, 1.00) | step = constant
f(xk)
iteration k

B - Βημα γ_k που ελαχιστοποιει την $\phi(\gamma_k) = f(x_k + \gamma_k d_k)$



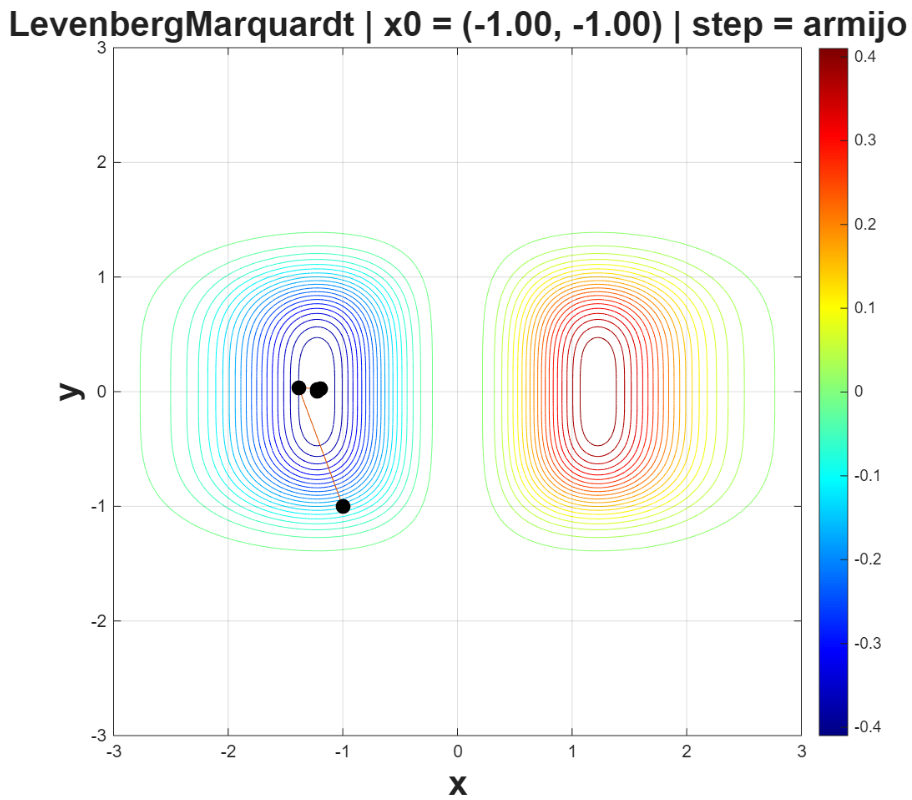


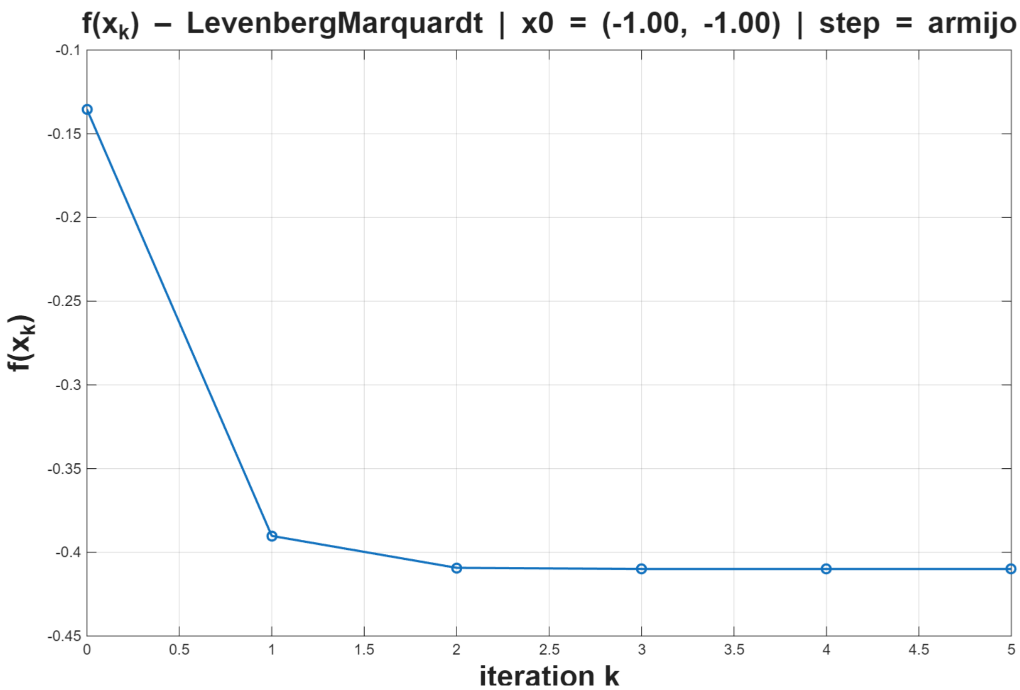


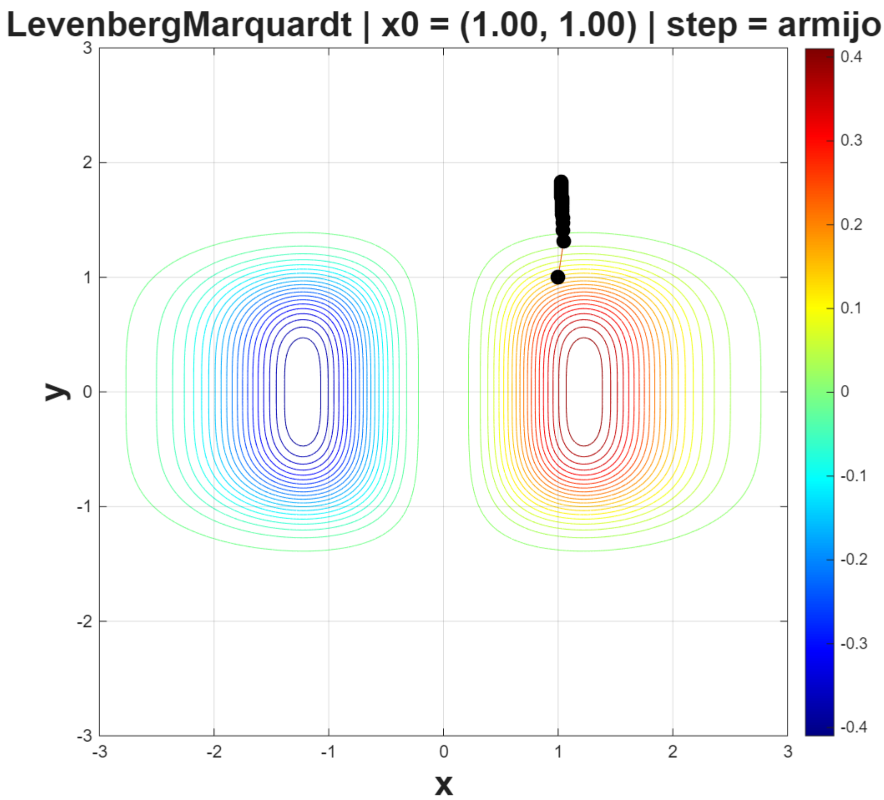


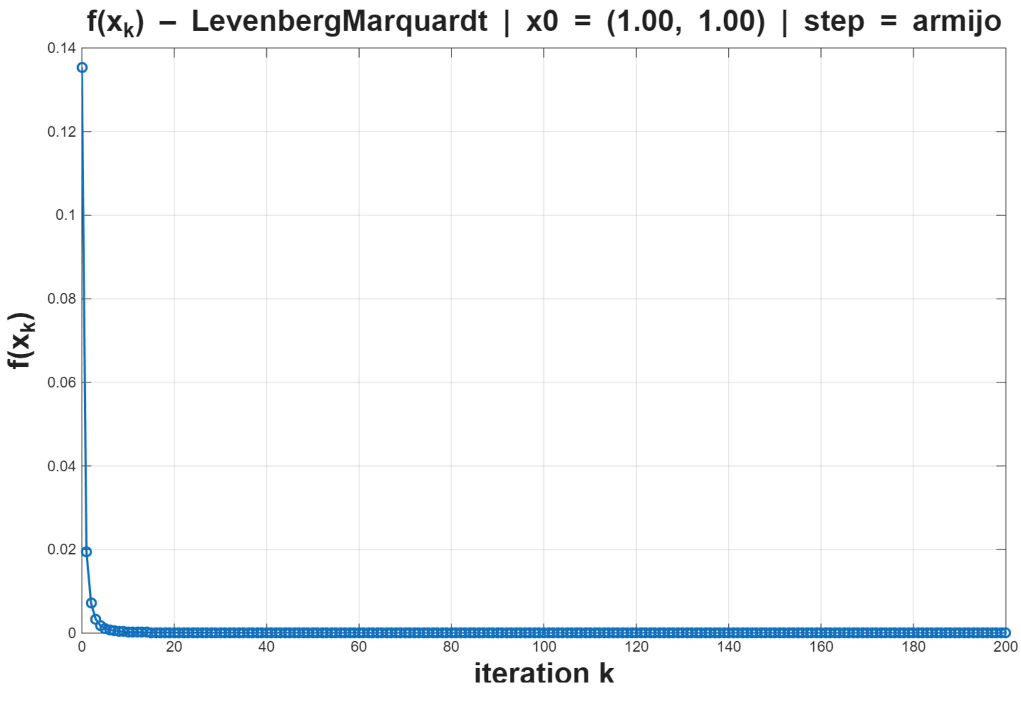
Γ - Βημα βασει του κανονα Armijo

Ιδιες παραμετρους με τα προηγουμενα Θεματα.


LevenbergMarquardt | x0 = (-1.00, -1.00) | step = armijo
x
y
3
2
1
0
-1
-2
-3
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4


f(x_k) – LevenbergMarquardt | x0 = (-1.00, -1.00) | step = armijo
f(x_k)
iteration k
-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





Σχόλια - Παρατηρήσεις

- Στην υλοποιηση του αλγοριθμου χρησιμοποιησα: $\mu_k = \bar{\mu}_k + 0.5$ οπου $\bar{\mu}_k = -\lambda_{min}$ και λ_{min} η μικροτερη αλγεβρικα ιδιοτιμη του Hessian. Μετα απο δοκιμες ειδα οτι με αυτη την επιλογη, ειχα τα καλυτερα αποτελεσματα.

- Βλεπουμε πως τωρα και οι 3 περιπτωσεις επιλογης βηματος γ_k εκτελουν βηματα απο το σημειο $(-1,-1)$, και μαλιστα φτανουν πολυ κοντα στο ελαχιστο πολυ νωρις. Αυτο γινεται γιατι με τον τροποποιημενο Hessian ξεπερασαμε το προβλημα της θετικης ορισιμοτητας.

- Η μεθοδος αυτη, οπως ειπαμε, ειναι ενας καλος συνδυασμος Newton + SD. Ετσι, χρησιμοποιει τη σταθεροτητα της SD αλλα και την ταχυτητα της Newton. Συγκεκριμενα, για το σημειο εκκινησης (1,1) βλεπουμε οτι κραταει μια συμπεριφορα Steepest Descent και εγκλωβιζεται στην περιοχη της "σελας", το οποιο μαλλον συμβαινει επειδη σε εκεινη την περιοχη ο Hessian δεν ειναι ποτε θετικα ορισμενος οποτε η μεθοδος δεν εκμεταλλευεται ποτε την Newton πολυ. Ωστοσο, στο κοντινο σημειο εκκινησης (-1,-1) βλεπουμε πως το σταθερο βημα και ο Armijo συγκλινουν σε μολις 5 βηματα, ενω η exact σε 11. Αυτο σημαινει οτι μετα το σημειο εκκινησης, σε καποια σημεια ο Hessian ηταν θετικα ορισμενος οποτε δεν εγινε τροποποιηση και η μεθοδος ηταν αποκλειστικα Newton, προσφεροντας την μεγαλη ταχυτητα συγκλισης της.

Συμπεράσματα

- Παρατηρουμε οτι και οι τρεις μεθοδοι, οταν ξεκινουν απο το (1,1), δεν συγκλινουν στο ολικο ελαχιστο. Αυτο γινεται λογω της γεωμετριας της συναρτησης στο σημειο αυτο, η οποια ειναι τετοια, που "στελνει" τους αλγοριθμους σε μια περιοχη οπου η f σχηματιζει μια επιφανεια "σελας", αντι στο ελαχιστο. Σε αυτη την περιοχη το gradient σχεδον μηδενιζεται ενω δεν υπαρχει τοπικο ακροτατο — προκειται ουσιαστικα για saddle region. Επομενως οι μεθοδοι “παγωνουν” εκει, καθως το μικρο $\|\nabla f\|$ τις κανει να θεωρουν οτι εχουν φτασει στο ελαχιστο.

- Οταν ομως ξεκινουν απο το (-1,-1) οι SD-LM πλησιαζουν ικανοποιητικα το ολικο ελαχιστο της συναρτησης, ενω η Newton κρασαρει επειδη ο Hessian δεν ειναι θετικα ορισμενος εκει.

- Συνολικα, η Steepest Descent ειναι σταθερη αλλα αργη, η Newton πολυ γρηγορη οταν ο Hessian ειναι θετικα ορισμένος αλλα μπορει να αποτυχει σε μη κυρτές περιοχες, ενω η Levenberg–Marquardt συνδυαζει σταθεροτητα και ταχυτητα, διορθωνοντας τα προβληματα της Newton. Ωστοσο, οι 2 τελευταιες απαιτουν τον υπολογισμο του Hessian, το οποιο περαν των απαιτησεων που βαζει (αντιστρεψιμοτητα), ειναι πολυ ακριβο υπολογιστικα ακριβο και πολλες φορες αδυνατο.