2018

Άσκηση 2 : Εύρεση ελαχίστου σε συνάρτηση δύο μεταβλητών

ΓΙΩΡΓΟΣ ΜΠΑΛΑΟΥΡΑΣ (8861)

Τεχνικές Βελτιστοποίησης

Ονοματεπώνυμο: *Μπαλαούρας Γεώργιος*

AEM: *8861*

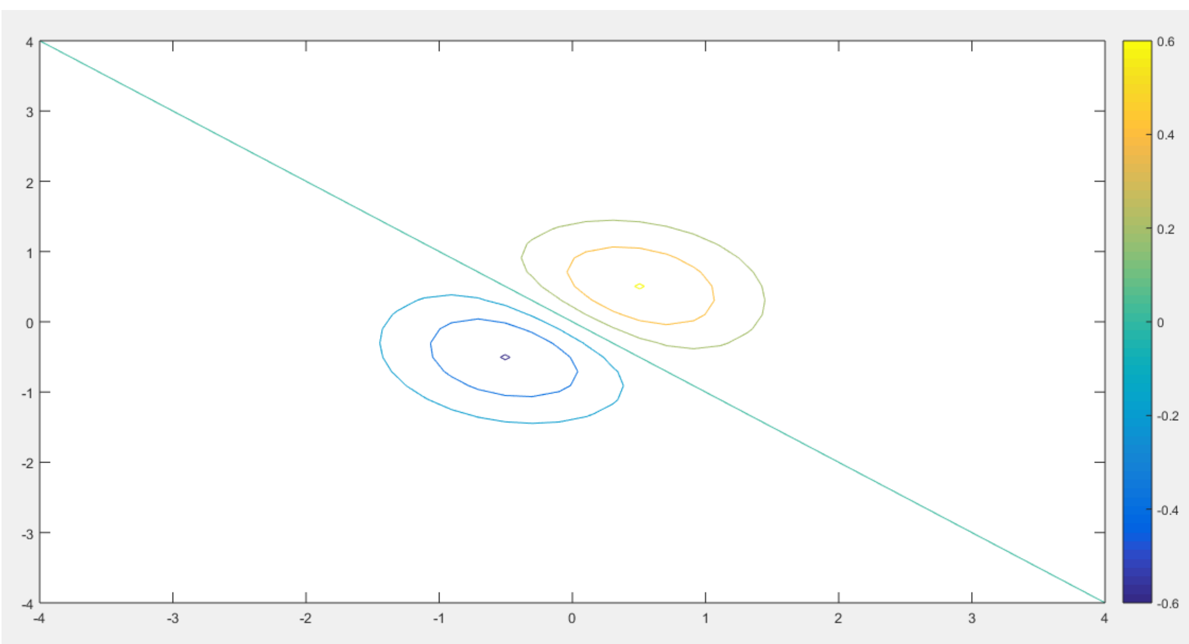
E-mail: *mpalaourg@ece.auth.gr*

Εξάμηνο: *6ο Ηλεκτρονικής*

<u>**Άσκηση 2**</u> **:** <u>*Εύρεση ελαχίστου σε συνάρτηση δύο μεταβλητών*</u>

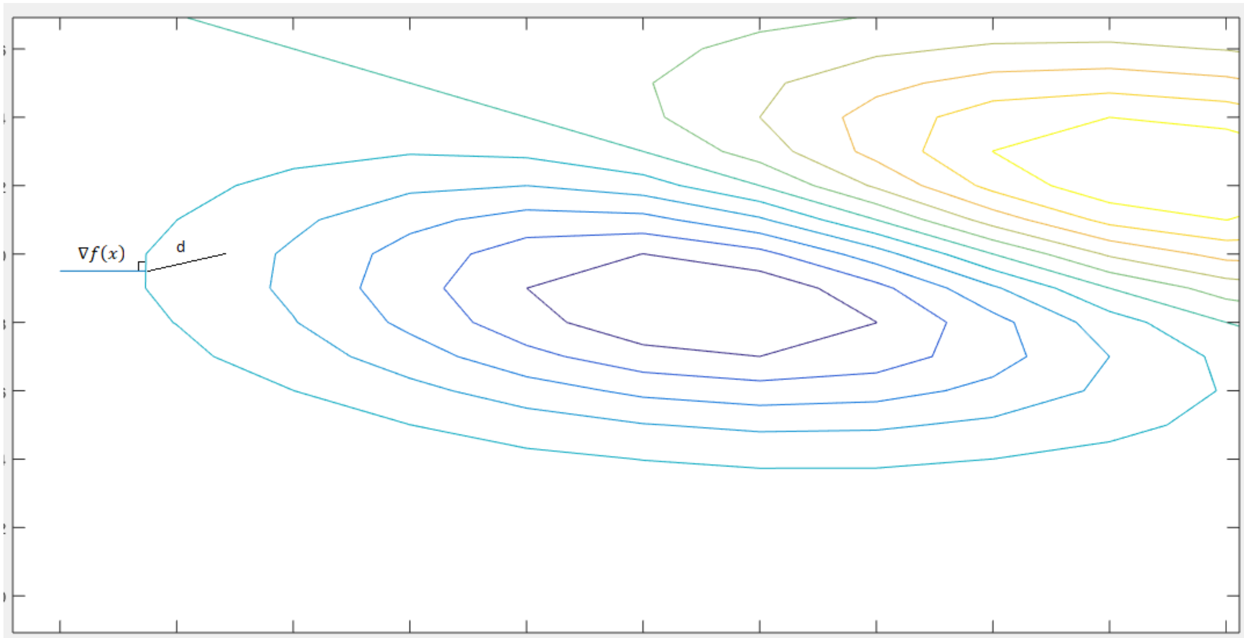
Αντικείμενο αυτής της άσκησης, ήταν η εφαρμογή δύο μεθόδων, της <u>μέγιστης καθόδου</u> και της <u>μεθόδου Newton</u>, σε συνάρτηση f, για την οποία δεν υπήρχε αναλυτικός τύπος, αλλά υπήρχε πρόσβαση στις αριθμητικές τιμές των $f(x), \nabla f(x)$ και $\nabla^2 f(x)$, με $x \in \mathbb{R}^2$.

Οι δύο αυτοί μέθοδοι, ανήκουν στις μεθόδους κλίσης. Βασικό χαρακτηριστικό τους είναι η αξιοποίηση της καθετότητας του διανύσματος κλίσης, $\nabla f(x)$, στην ισοβαρή καμπύλη της f.



Σχήμα 1: Ισοβαρείς καμπύλες της f

Γνωρίζοντας, ότι οι μέθοδοι στηρίζονται στη ιδέα της επαναληπτικής διαδικασίας , δηλαδή ισχύει $f(x_{k+1}) < f(x_k)$, $k = 0, 1, 2, \dots$, ορίζουμε το $\bar{x} = x + \gamma d$, με $\gamma \geq 0$ και διάνυσμα κατεύθυνσης $d \in \mathbb{R}^n$ (εδώ $d \in \mathbb{R}^2$). Για να μειωθεί στην επόμενη επανάληψη το $\bar{x}$, πρέπει το διάνυσμα κατεύθυνσης να σχηματίζει αμβλεία γωνία με το διάνυσμα κλίσης, δηλαδή $\nabla f^T(x)d < 0$, έτσι ώστε να «βλέπει» προς μικρότερη ισοβαρή καμπύλη.

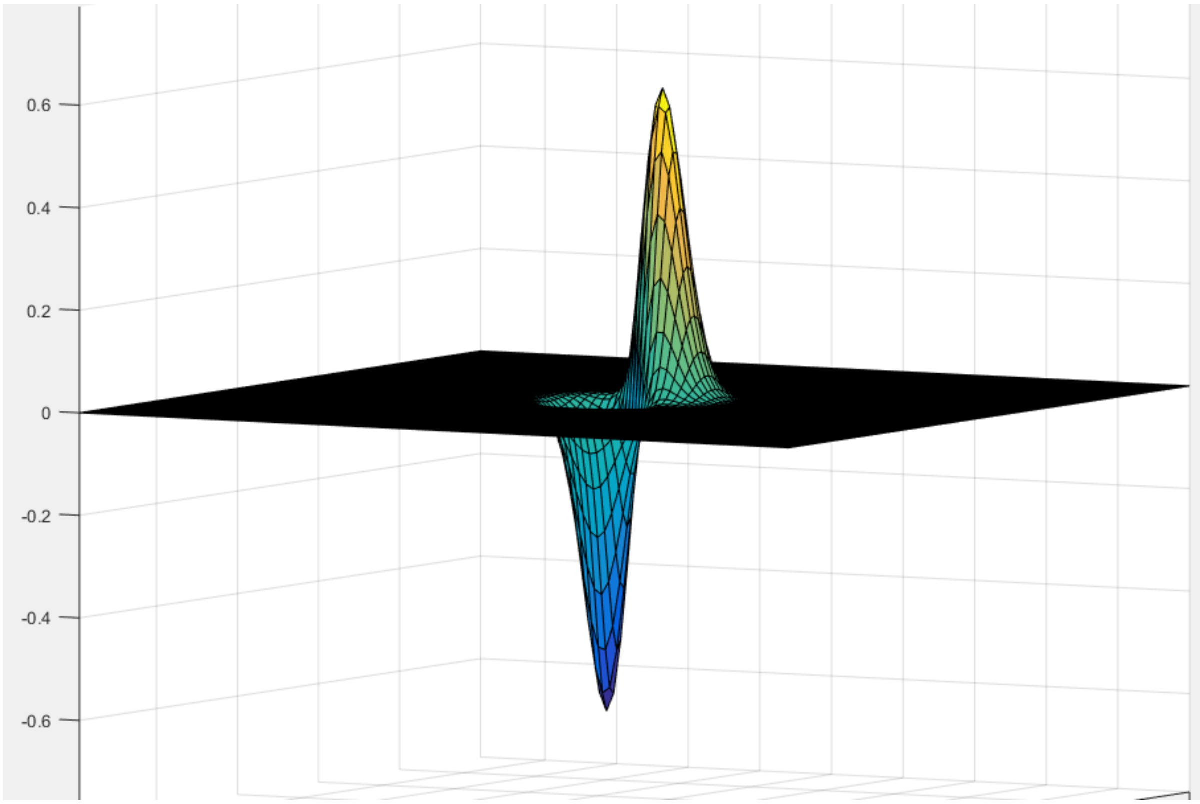


Σχήμα 2: Μεγεθυμένη η φωτογραφία του Σχήματος 1

Στο σχήμα 2, φαίνεται η αναγκαιότητα τα δύο διανύσματα να σχηματίζουν αμβλεία γωνία. Επίσης, γίνεται σαφές πως η επιλογή μικρού γ θα έχει ως αποτέλεσμα αργή σύγκλιση στο ζητούμενο σημείο, ενώ η επιλογή μεγάλου γ, μπορεί να οδηγήσει σε ταλαντώσεις ή σε ισοβαρείς καμπύλες με μεγαλύτερη τιμή.

Ένα ακόμα, χαρακτηριστικό (αρκετών) αλγορίθμων κλίσης είναι η σχέση που δίνει το επόμενο διάνυσμα $x_{\kappa+1}$, δηλαδή $x_{\kappa+1} = x_\kappa - \gamma_\kappa \Delta_\kappa \nabla f(x_\kappa)$, όπου $\Delta_\kappa = {\Delta_\kappa}^T > 0$.   (1)

Επιπλέον, για να έχω μια γενική ιδέα της αντικειμενικής συνάρτησης χρησιμοποιώ το `plot_f.m`, όπου εκτός της γραφικής παράστασης της f, υπολογίζεται (στο περίπου) το $max\{f\} \cong 0.6065$ και το $\min\{f\} \cong -0.6065$



Σχήμα 3: Γραφική παράσταση της f

<u>Εφαρμογή των μεθόδων</u>:

Στο πρώτο σημείο που μας δόθηκε, $(-0.5, -0.5)$:

Υπολογίζοντας το `f(-0.5,-0.5)` $= -0.6065$,

Μέσω του `gradf(-0.5,-0.5)` $= \begin{bmatrix} 0 \\ 0 \end{bmatrix}$,

<u>και</u> `hessianf(-0.5,-0.5)` $= \begin{bmatrix} 1.8196 & 0.6065 \\ 0.6065 & 1.8196 \end{bmatrix}$, με ιδιοτιμές $\lambda = \begin{bmatrix} 1.2131 \\ 2.4261 \end{bmatrix}$, άρα θετικά ορισμένος. Δηλαδή, το σημείο είναι το (ολικό) ελάχιστο της f, άρα καμία μέθοδος δεν θα εκτελεστεί μ' αυτό το αρχικό σημείο.

Ομοίως, στο δεύτερο σημείο που μας δόθηκε, $(0.5, 0.5)$:

`f(0.5,0.5)` $= 0.6065$,

Μέσω του `gradf(0.5,0.5)` $= \begin{bmatrix} 0 \\ 0 \end{bmatrix}$,

<u>και</u> `hessianf(0.5,0.5)` $= \begin{bmatrix} -1.8196 & -0.6065 \\ -0.6065 & -1.8196 \end{bmatrix}$, με ιδιοτιμές $\lambda = \begin{bmatrix} -2.4261 \\ -1.2131 \end{bmatrix}$, άρα αρνητικά ορισμένος. Δηλαδή, το σημείο είναι το (ολικό) μέγιστο της f, άρα και για αυτό το αρχικό σημείο, οι μέθοδοι δεν θα εκτελεστούν.

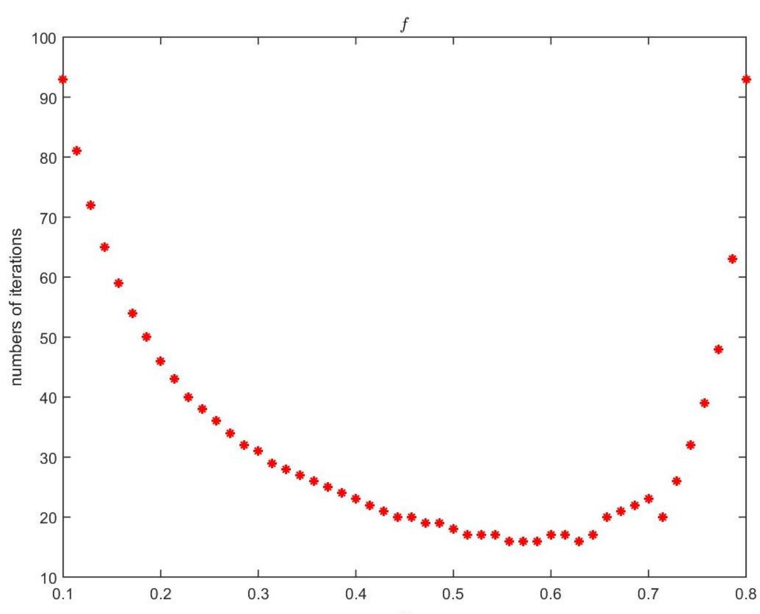
<u>**Μέθοδος της Μέγιστης Καθόδου**</u>: Η γενική ιδέα της μεθόδου είναι, ότι η μεγαλύτερη αύξηση, $||dx||$, της μεταβλητής x, θα υπάρξει όταν το διάνυσμα dx είναι συγγραμικό με το διάνυσμα κλίσης, ∇f, της f.

Συγκεκριμένα, αν στην σχέση (1), όπου Δ_κ θέσουμε τον μοναδιαίο πίνακα I, θα έχουμε $x_{\kappa+1} = x_\kappa - \gamma_\kappa \nabla f(x_\kappa)$.

- Αρχικό σημείο $(-2,-1)$

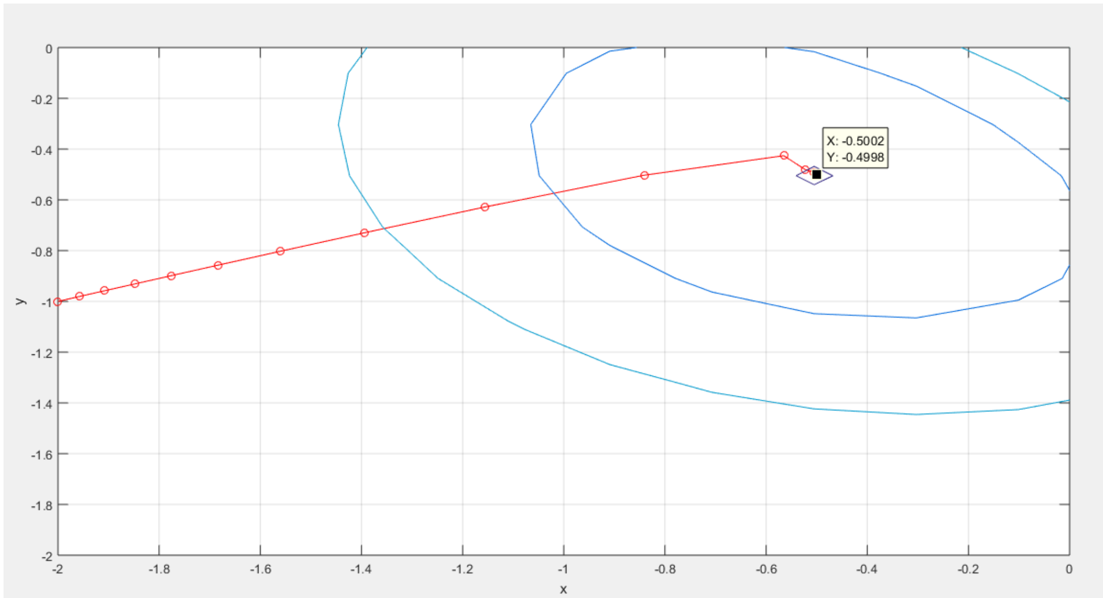
Υπολογίζουμε `f(-2,-1)` $= -0.0202$ και `gradf(-2,-1)` $= [-0.0741, -0.0337]$, με |`gradf(-2,-1)`| $= 0.8014$.

Επιλέγοντας $\varepsilon = 0.001$, εκτελώ το `steepest_descent_plot_for_different_g` και παρατηρώ, για τιμές κοντά στο 0 , η συνάρτηση αργεί να οδηγηθεί στο ελάχιστο και για τιμές κοντά στο $\gamma = 0.81$, η μέθοδος ταλαντώνει.



Σχήμα 4: Αριθμός επαναλήψεων για διάφορα γ [Μέγιστη Κάθοδος]

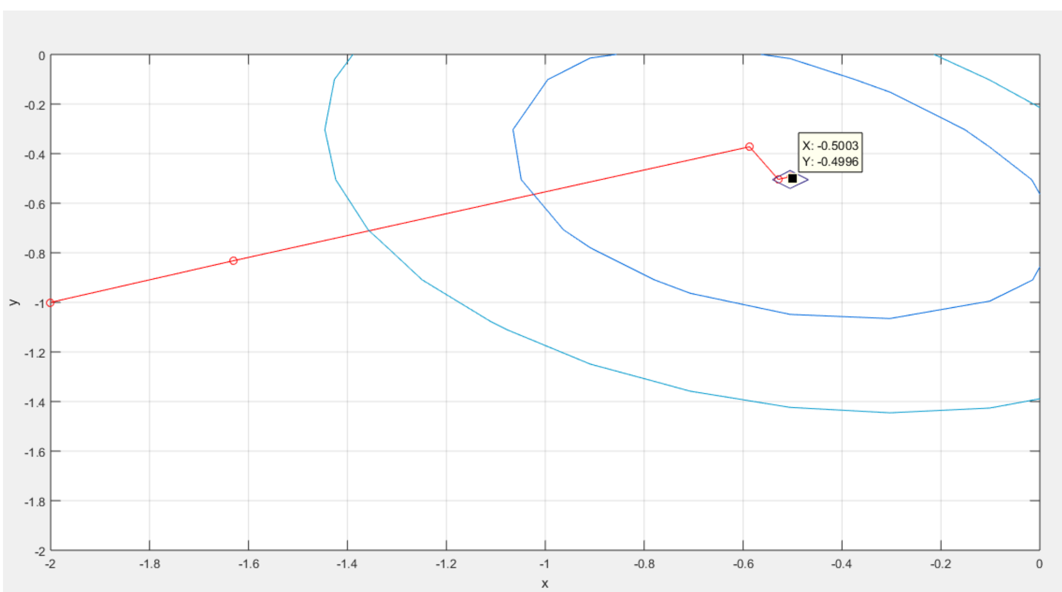
a) Παρατηρώντας, την παραπάνω γραφική παράσταση και έχοντας ως κριτήριο τις λιγότερες επαναλήψεις επιλέγω $\gamma = \sigma\tau\alpha\theta\varepsilon\rho\acute{o} = 0.5714$.



Σχήμα 5: Σύγκλιση του ζεύγους (x, y) για σταθερό γ [Μέγιστη Κάθοδος]

b) Για να ελαχιστοποιηθεί η συνάρτηση $h(\gamma) = f(x_k + \gamma_\kappa d_k)$, αφού είναι συνάρτηση μιας μεταβλητής χρησιμοποιώ την `golden_section.m` (από την προηγούμενη εργασία).

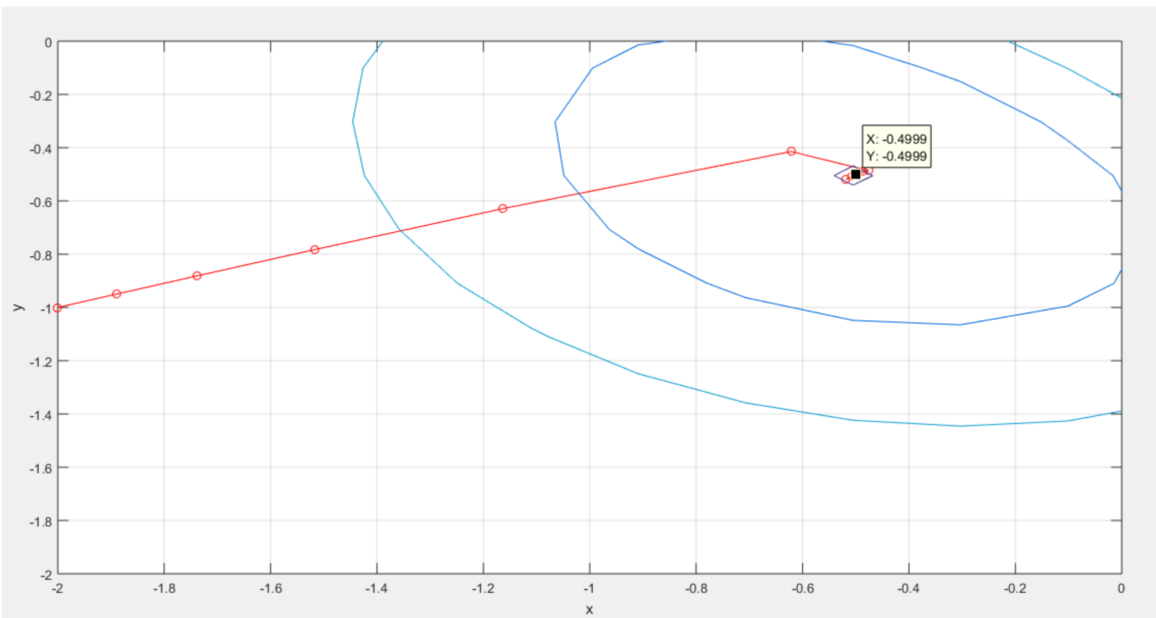
<u>Σημείωση</u>: Ο κώδικας της `golden_section.m`, προσαρμόστηκε έτσι ώστε να μην επιστρέφει διάστημα, αλλά σημείο.



Σχήμα 6: Σύγκλιση του ζεύγους (x, y) για ελαχιστοποιημένο γ [Μέγιστη Κάθοδος]

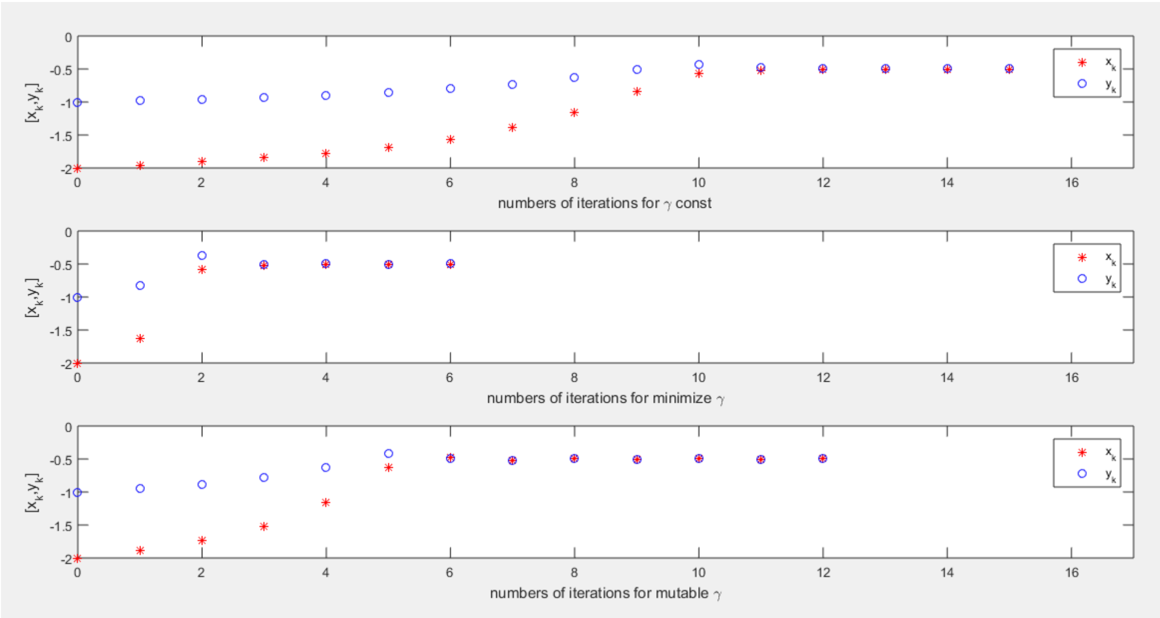
c) Για το μεταβλητό γ, η ιδέα ήταν να ξεκινήσω από μια μεγάλη τιμή για το γ, $\gamma = 1.5$ και όσο προχωράω σε «μικρότερες» ισοβαρείς, να μειώνω το γ, κατά 10%. Η ιδέα, βασίστηκε στο γεγονός, πως αν είμαι μακριά από το σημείο θέλω να κάνω μεγαλύτερα βήματα και όσο πλησιάζω στο ζητούμενο σημείο, να μικραίνει το βήμα μου. Για να αποφύγω τα πολύ μικρά βήματα, ελέγχω αν το γ, πέσει κάτω από μια τιμή και αν γίνει κάτι τέτοιο, το τριπλασιάζω.

<u>Σημείωση</u>: Ο έλεγχος στο γ, ενεργοποιείται -κυρίως- για μικρά ε.



Σχήμα 7: Σύγκλιση του ζεύγους (x, y) για μεταβλητό γ [Μέγιστη Κάθοδος]

Για να γίνει πιο σαφής η σύγκριση μεταξύ των διαφορετικών «μορφών» της Μέγιστης Καθόδου, χρησιμοποιώ το `steepest_descent_plot_x_y_compare.m`, όπου το ζεύγος (x, y), για κάθε μια από τις παραπάνω μορφές, σχεδιάζεται συναρτήσει των αριθμών επαναλήψεων.



Σχήμα 8: Σύγκλιση των (x_κ, y_κ) [Μέγιστη Κάθοδος]

- Αρχικό σημείο $(-3,3)$

Υπολογίζουμε `f(-3,3)` $= 0$ και `gradf(-3,3)` $= [0.1523,\ 0.1523] * 10^{-7}$, όποτε το μέτρο του, |`gradf(-2,-1)`| $= 2.1538 * 10^{-8}$.

Για να καταφέρουμε τον αλγόριθμο, έστω να ξεκινήσει τις επαναλήψεις πρέπει να θέσουμε ένα αρκούντως μικρό ε. Ακόμα κι έτσι, ο αλγόριθμος φτάνει στο άνω όριο βημάτων (5000 επαναλήψεις) και παρατηρούμε ότι η μέθοδος για αυτό το σημείο δεν μπορεί να οδηγηθεί στο ελάχιστο, αφού δεν μπορεί να «ξεφύγει» από την μικρή κλίση της περιοχής.

Μέθοδος Newton: Σ’ αυτήν την μέθοδο, προσεγγίζουμε την αντικειμενική συνάρτηση, με την τετραγωνική της μορφή, με σκοπό την σύγκλιση στο σημείο ελαχίστου σε ένα και μόνο βήμα.

Συγκεκριμένα, αν στην σχέση (1), όπου Δ_κ θέσουμε τον αντίστροφο εσσιανό πίνακα $[\nabla^2 f(x_k)]^{-1}$, θα έχουμε $x_{\kappa+1} = x_\kappa - \gamma_\kappa [\nabla^2 f(x_k)]^{-1} \nabla f(x_\kappa)$.

Η μέθοδος Newton, για να λειτουργήσει απαιτεί ο $\nabla^2 f(x_k)$ να είναι θετικά ορισμένος και να αντιστρέφεται.

- Αρχικό σημείο $(-2,-1)$

Υπολογίζοντας τον εσσιανό στο συγκεκριμένο σημείο:

`hessianf(-2,-1)` $= \begin{bmatrix} -0.2291 & -0.1213 \\ -0.1213 & -0.0135 \end{bmatrix}$, παρατηρούμε ότι οι ελάσσονες ορίζουσες είναι αρνητικές, $D_1 = -0.2291 < 0$ και $D_2 = \det\left(\begin{bmatrix} -0.2291 & -0.1213 \\ -0.1213 & -0.0135 \end{bmatrix}\right) = -0.0116 < 0$. Άρα, ο εσσιανός αρνητικά ορισμένος και η μέθοδος <u>δεν</u> θα λειτουργήσει.

- Αρχικό σημείο $(-3,3)$

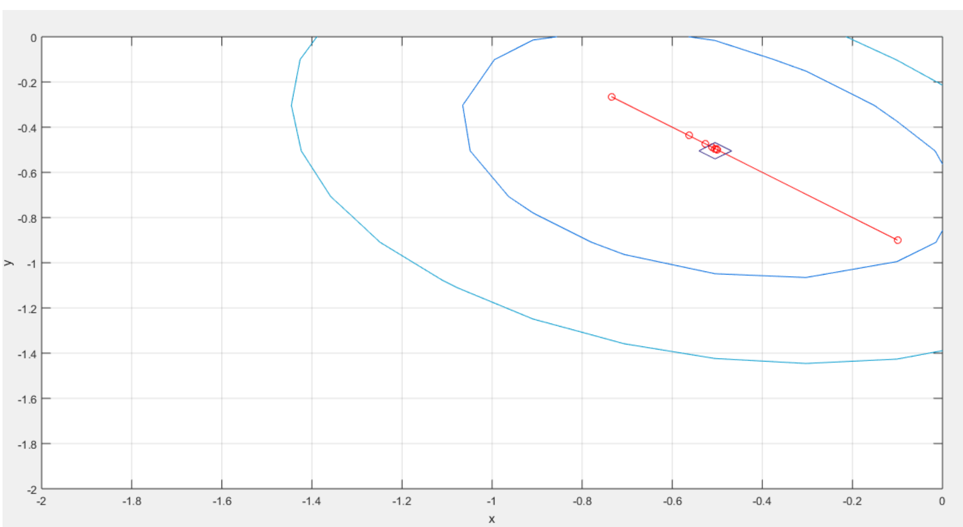
Με το ίδιο σκεπτικό, υπολογίζω:

`hessianf(-3,3)` $= \begin{bmatrix} 0.1828 & 0 \\ 0 & -0.1828 \end{bmatrix} * 10^{-6}$, παρατηρούμε ότι οι ελάσσονες ορίζουσες, $D_1 = 0.1828 * 10^{-6} > 0$ και $D_2 = \det\left(\begin{bmatrix} 0.1828 & 0 \\ 0 & -0.1828 \end{bmatrix} * 10^{-6}\right) = -3.34 * 10^{-14} < 0$. Άρα, ο εσσιανός είναι αόριστος και πάλι το κριτήριο δεν τηρείται.

Για να αποδειχθεί, ότι ο αλγόριθμος και η μέθοδος δουλεύει επιλέγω το σημείο $(-0.1,-0.9)$.

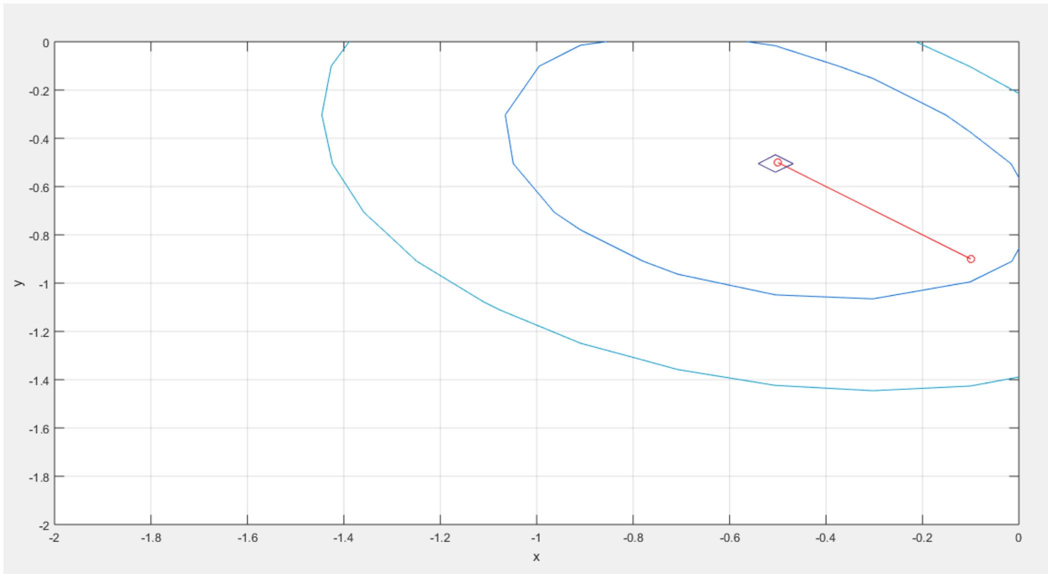
`hessianf(-0.1,-0.9)` $= \begin{bmatrix} 1.0394 & 0.7223 \\ 0.7223 & 1.0394 \end{bmatrix}$, παρατηρούμε ότι οι ελάσσονες ορίζουσες, $D_1 = 1.0394 > 0$ και $D_2 = \det\left(\begin{bmatrix} 1.0394 & 0.7223 \\ 0.7223 & 1.0394 \end{bmatrix}\right) = 1.7617 > 0$. Άρα, ο εσσιανός θετικά ορισμένος και η μέθοδος μπορεί να δοκιμαστεί.

I. Κρατώντας σταθερό το $\varepsilon = 0.001$ και επιλέγοντας το ίδιο γ με πριν, $\gamma = 0.5714$, έχω:



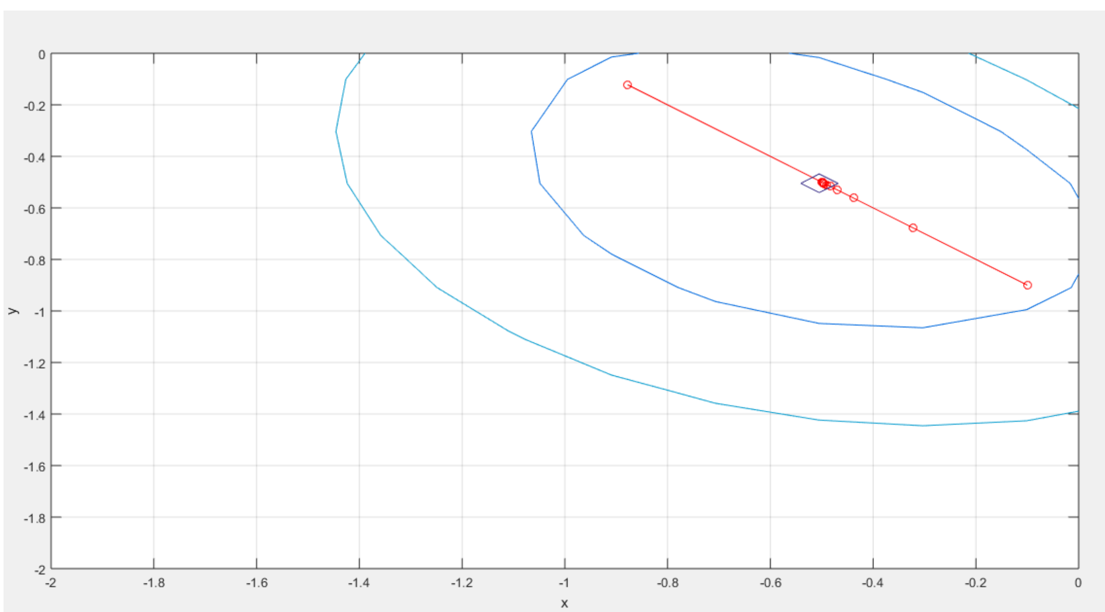
Σχήμα 9: Σύγκλιση του ζεύγους (x, y) για σταθερό γ [Newton]

II. Εφαρμόζοντας την ελαχιστοποίηση για το γ, πετυχαίνουμε τον στόχο μας. Συγκλίνουμε στο ελάχιστο, σε μία και μόνο επανάληψη.



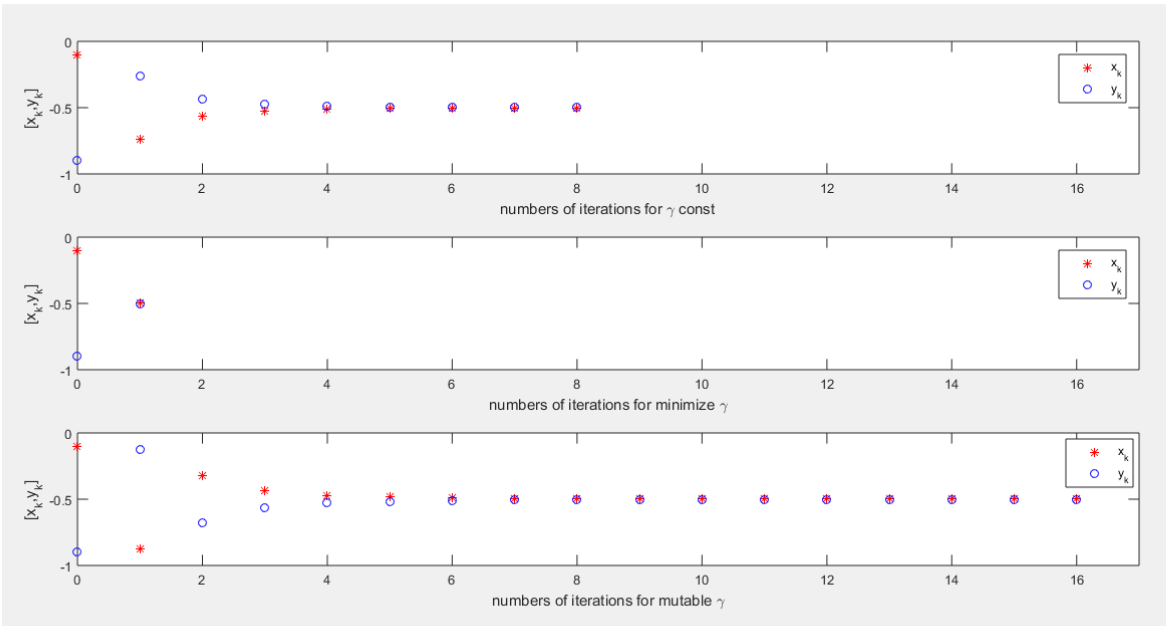
Σχήμα 10: Σύγκλιση του ζεύγους (x, y) για ελαχιστοποιημένο γ [Newton]

III. Και, τέλος ο ευριστικός κανόνας,



Σχήμα 11: Σύγκλιση του ζεύγους (x, y) για μεταβλητό γ [Newton]

Για να γίνει πιο σαφής η σύγκριση μεταξύ των διαφορετικών «μορφών» της μεθόδου Newton, χρησιμοποιώ το `newton_plot_x_y_compare.m`, όπου το ζεύγος (x, y), για κάθε μια από τις παραπάνω μορφές, σχεδιάζεται συναρτήσει των αριθμών επαναλήψεων.



Σχήμα 12: Σύγκλιση των (x_κ, y_κ) [Newton]

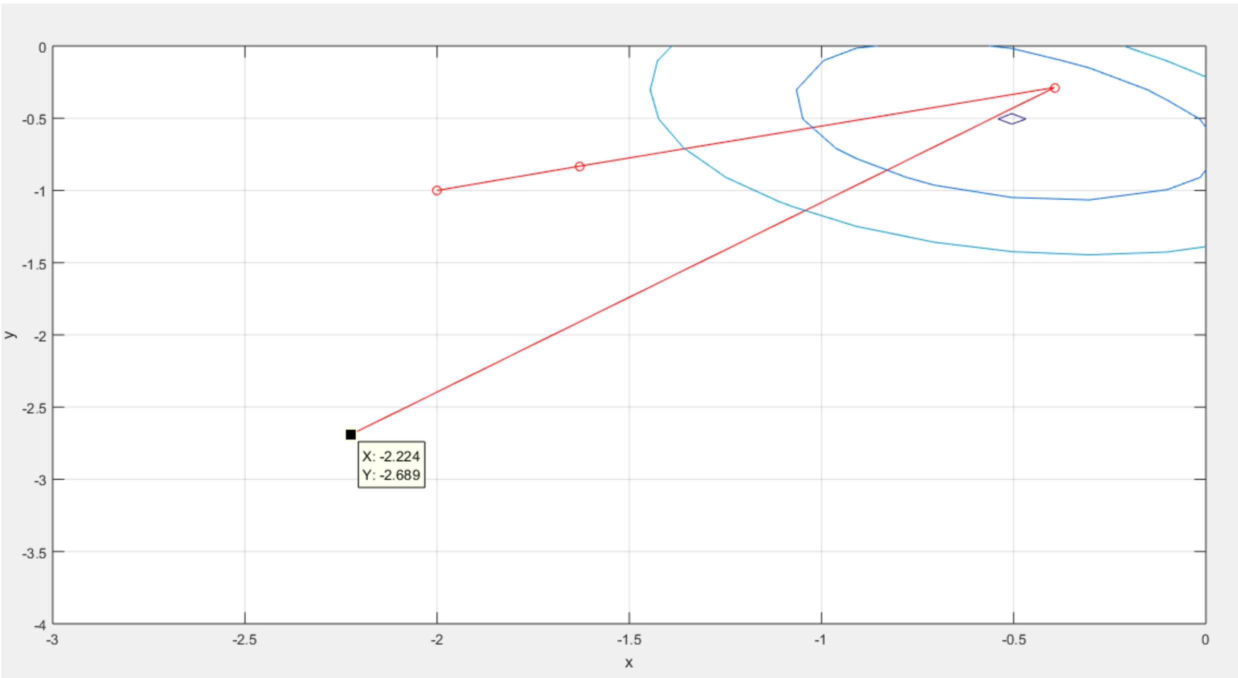
Συγκρίσεις – Συμπεράσματα:

Παρατηρώντας, τα παραπάνω διαγράμματα η μέθοδος Newton είναι σε γενικές γραμμές καλύτερη από την μέθοδο της μέγιστης καθόδου. Ωστόσο, και οι δύο μέθοδοι έχουν σημαντικές αδυναμίες. Από την μία στην μέγιστη κάθοδο, ο αλγόριθμος καθυστερεί στην σύγκλιση του και από την άλλη η Newton, χρειάζεται θετικά ορισμένο τον εσσιανό πίνακα, κάτι που δεν μπορεί να εγγυηθεί πως θα ισχύει σε κάθε επανάληψη. Τέλος και οι δύο μέθοδοι έχουν τοπικό χαρακτήρα, εγκλωβίζονται στην γειτονιά του ελαχίστου και έτσι δεν υπάρχει εγγύηση ότι το ελάχιστο που βρέθηκε είναι το ολικό.

Επιπλέον, οι δύο μέθοδοι για σταθερό γ στα παραδείγματα που παρουσιάστηκαν, οδηγήθηκαν στο ελάχιστο. Ωστόσο, επιλέχθηκαν έτσι ώστε να συμβεί κάτι τέτοιο. Για παράδειγμα, στην μέγιστη κάθοδο, για $\gamma = 1.2$, ο αλγόριθμος καταλήγει -μετά από 2500 βήματα- στο

$(-0.4999, -0.4999)$ ενώ για $\gamma = 0.83$ ο αλγόριθμος ταλαντώνεται μεταξύ των σημείων, $(-0.4605, -0.4605)$ - $(-0.5429, -0.5429)$ και τερματίζει λόγω του ανώτατου ορίου βημάτων.

Τέλος, για $\gamma = 5$, ο αλγόριθμος αποκλίνει:



Σχήμα 13: Απόκλιση για σταθερό γ =5 (Μέγιστη Κάθοδος)

Ο ευριστικός κανόνας, βασίζεται στο γεγονός ότι μακριά από το σημείο ελαχίστου, θέλω να κάνω μεγάλα βήματα. Ωστόσο, κι εδώ υπάρχουν «προβλήματα» καθώς πρέπει να επιλεγεί μια αρχική τιμή για το γ, που μπορεί να οδηγήσει σε μεγαλύτερη ισοβαρή καμπύλη , άρα πιο αργή σύγκλιση- όπως για παράδειγμα συμβαίνει στο σχήμα 11, για την μέθοδο Newton, που αρχικά οδηγούμε μακριά από το ελάχιστο- ή ακόμα και απόκλιση από το ζητούμενο σημείο.

Τέλος, η ελαχιστοποίηση του γ, οδηγεί στα καλύτερα αποτελέσματα. Ειδικά στην μέθοδο Newton, όπου η σύγκλιση έγινε σε μόλις ένα βήμα.