2η Εργαστηριακή Άσκηση

Βασίλειος Αϊτσίδης (AEM:10330)

November 21, 2023

Εισαγωγή

Ελαχιστοποίηση συναρτήσεων πολλών μεταβλητών χωρίς περιορισμούς με χρήση παραγώγων

Σε αυτήν την εργασία ασχολούμαστε με το πρόβλημα της ελαχιστοποίησης μιας δοσμένης συνάρτησης πολλών μεταβλητών $f : \mathbb{R}^n \to \mathbb{R}$ χωρίς περιορισμούς. Οι αλγόριθμοι που χρησιμοποιούνται βασίζονται στην ιδέα της επαναληπτικής καθόδου, βάσει της οποίας ξεκινάμε από κάποιο σημείο $x_0 \in \mathbb{R}^n$ και παράγουμε διαδοχικά τα διανύσματα $x_1, x_2, ...$ έτσι ώστε $f(x_{k+1}) < f(x_k), k = 1, 2....$

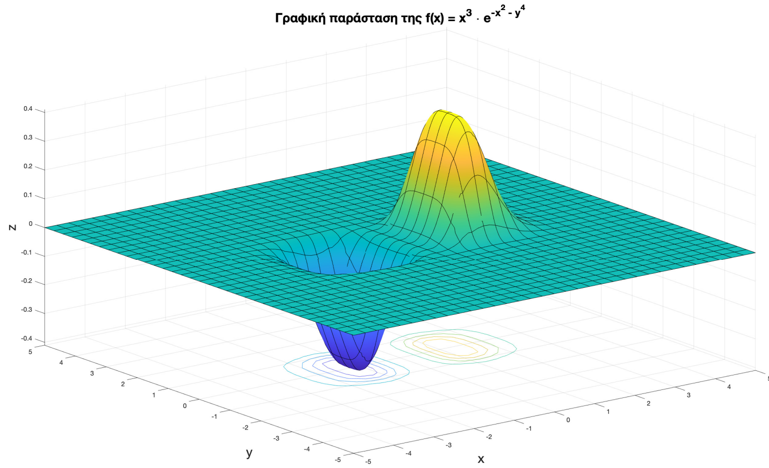
Οι αλγόριθμοι αναζήτησης που μελετώνται είναι :

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

Η αντικειμενική συνάρτηση που θα μελετήσουμε είναι η $f(x, y) = x^3 e^{-x^2 - y^4}$

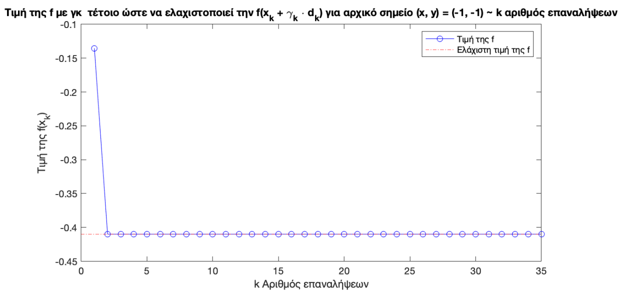
Θέμα 1

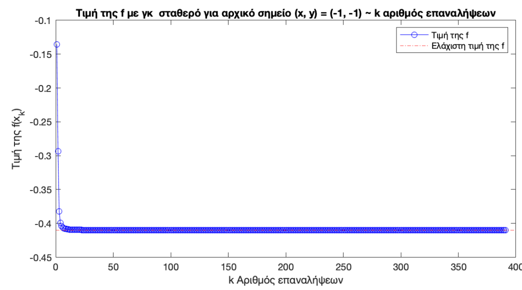
Εδώ παραθέτουμε μια γραφική παράσταση της αντικειμενικής συνάρτησης που καλούμαστε να ελαχιστοποιήσουμε, ώστε να έχουμε μια οπτική εικόνα αυτής. Παρατηρούμε ότι πρόκειται για μια επιφάνεια στον τρισδιάστατο χώρο με δύο βουναλάκια , ένα προς τα πάνω κι ένας προς τα κάτω. Από τη γραφική παράσταση, είναι εμφανές που περίπου βρίσκεται το ελάχιστο (ολικό) της συνάρτησης. Για να φανταστούμε καλύτερα τη συνάρτηση σχεδιάζονται και οι ισοϋψείς καμπύλες, πάνω στις οποίες η τιμή της $f(x, y)$ παραμένει σταθερή.

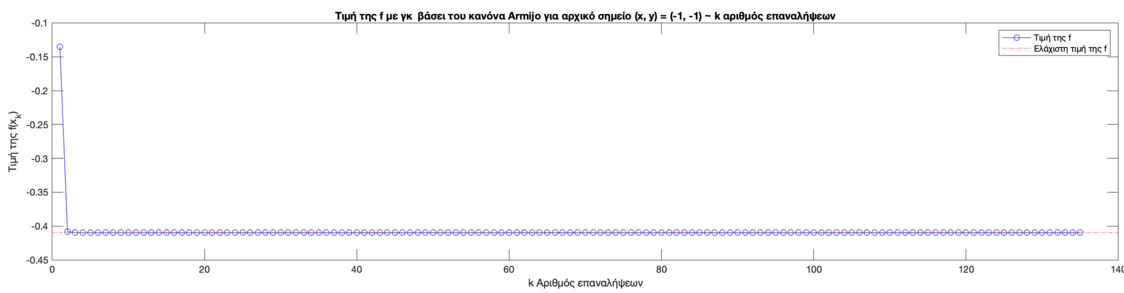


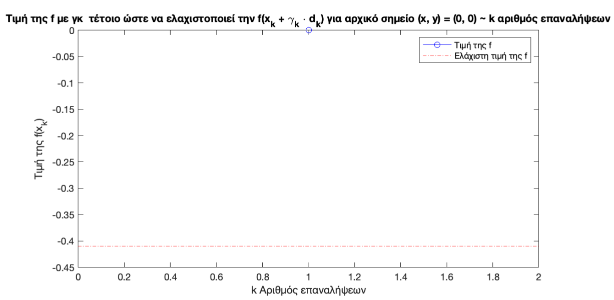
Θέμα 2

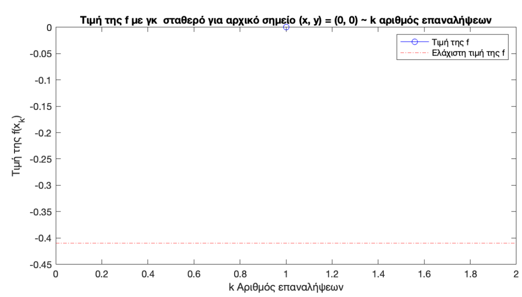
Στο θέμα αυτό, μας ζητείται να ελαχιστοποιήσουμε την f χρησιμοποιώντας τη μέθοδο της **Μέγιστης Καθόδου (Steepest Descent)** επιλέγοντας ως αρχικά σημεία (x_0, y_0) τα i) (0,0), ii) (-1,-1) και iii) (1,1). Το βήμα γ_κ επιλέγεται είτε α) σταθερό (επιλέγουμε κατά την κρίση μας) είτε β) τέτοιο ώστε να ελαχιστοποιεί την $f(x_k + \gamma_\kappa d_k)$ είτε γ) βάσει του κανόνα Armijo. Στη μέθοδο αυτή το διάνυσμα κατεύθυνσης ισούται με την αρνητική κλίση της f υπολογισμένη στο σημείο x_k, δηλαδή $d_k = -\nabla f(x_k)$.Έτσι, λοιπόν, σε κάθε επανάληψη $k = k + 1$ έχουμε $x_{k+1} = x_k - \gamma_k \nabla f(x_k)$. Από γεωμετρικής άποψης , το διάνυσμα που ενώνει το x_k με το x_{k+1} είναι κάθετο στο διάνυσμα που ενώνει το x_{k+1} με το x_{k+2}. Επομένως, η πορεία προς το ελάχιστο είναι μια τεθλασμένη γραμμή, με τα διαδοχικά ευθύγραμμα τμήματα να σχηματίζουν γωνία 90°. Αυτή η *zik-zak* συμπεριφορά , καθιστά τη μέθοδο ως επί το πλείστον αργή και κατά συνέπεια υπολογιστικά αναποτελεσματική. Όπως αντιλαμβανόμαστε, ανάλογα την επιλογή του αρχικού σημείου και του γ_κ προκύπτουν και διαφορετικά αποτελέσματα. Παρακάτω έχουμε τις γραφικές παραστάσεις των τιμών των συναρτήσεων $f(x_k)$ συναρτήσει του αριθμού των επαναλήψεων κ για τα διάφορα αρχικά σημεία και μεθόδους επιλογής του γ_κ.

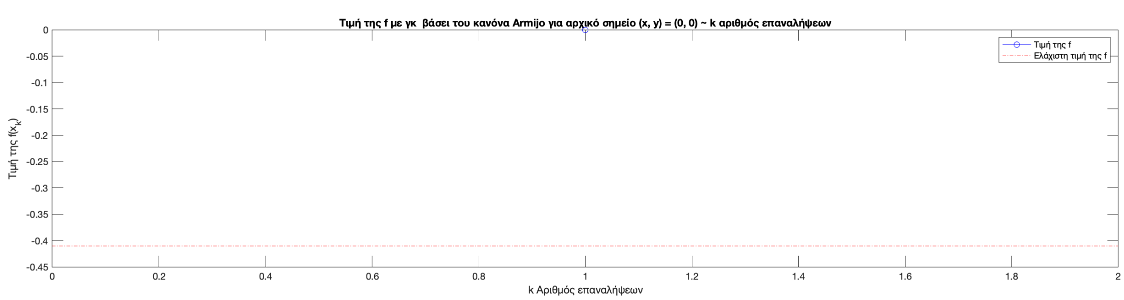

Τιμή της f με γκ τέτοιο ώστε να ελαχιστοποιεί την f(xk + γk · dk) για αρχικό σημείο (x, y) = (-1, -1) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων

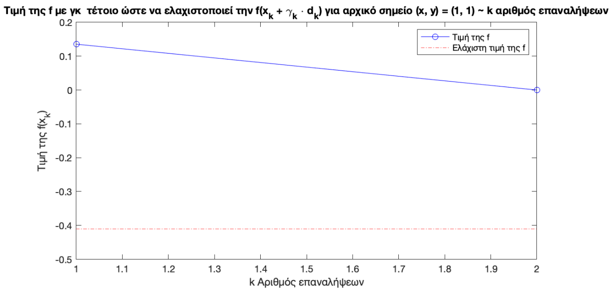

Τιμή της f με γκ σταθερό για αρχικό σημείο (x, y) = (-1, -1) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων

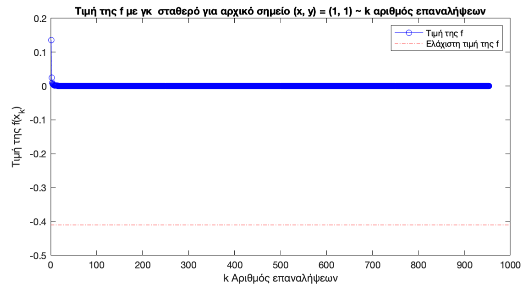

Τιμή της f με γκ βάσει του κανόνα Armijo για αρχικό σημείο (x, y) = (-1, -1) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων

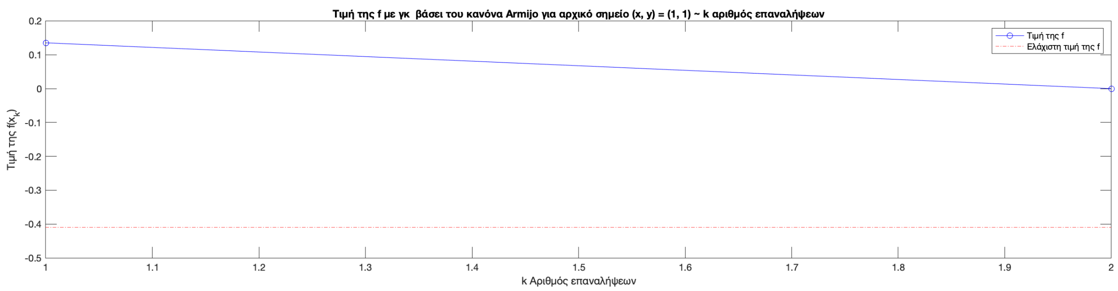

Τιμή της f με γκ τέτοιο ώστε να ελαχιστοποιεί την f(xk + γk · dk) για αρχικό σημείο (x, y) = (0, 0) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων


Τιμή της f με γκ σταθερό για αρχικό σημείο (x, y) = (0, 0) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων


Τιμή της f με γκ βάσει του κανόνα Armijo για αρχικό σημείο (x, y) = (0, 0) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων







Από τα παραπάνω γραφήματα οδηγούμαστε στα παρακάτω συμπεράσματα και σχόλια: Παρατηρούμε ότι όταν επιλέγουμε σημείο εκκίνησης το (0,0) ο αλγόριθμος δεν οδηγεί τη συνάρτηση σε καμια σύγκλιση για καμια επιλογή του γ_κ. Αυτό είναι αναμενόμενο, καθώς από τη γραφική παράσταση της αντικειμενικής συνάρτησης που δόθηκε παραπάνω, βλέπουμε ότι στο (0,0) η συνάρτηση παρουσιάζει τοπικό ελάχιστο και συνεπώς η συνθήκη τερματισμού δηλαδή $|\nabla f(x_k)| < \varepsilon$ τηρείται εξαρχής και συγκεκριμένα η κλίση είναι ακριβώς μηδεν (αφού το σημείο είναι τοπικό ελάχιστο).

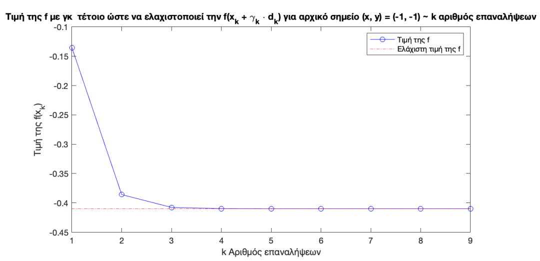
Όσον αφορά το σημείο (1,1), παρατηρούμε ότι για γ_κ σταθερό και ίσο με 0.5, υπάρχει σύγκλιση της f σε μια τιμή , ωστόσο η τιμή αυτή είναι διαφορετική από την πραγματική ελάχιστη (η οποία υπολογίζεται με τη βοήθεια έτοιμης συνάρτησης από το MATLAB). Συγκεκριμένα , παρατηρούμε ότι η τιμή της f σταθεροποιείται στην τιμή 0, ενώ το πραγματικό ελάχιστο βρίσκεται κάπου στο -0.4. Επίσης , να σημειώσουμε ότι ο αλγόριθμος τερματίζει επειδή φτάνει το μέγιστο αριθμό επαναλήψεων που έχει επιλεγεί αυθαίρετα απο μας. Για γ_κ που επιλέγεται με τις άλλες δυο μεθόδους, βλέπουμε ότι η τιμή της f μειώνεται , αλλά στις δύο κιόλας επαναλήψεις ο αλγόριθμος τερματίζει περίπου στην τιμή 0 (πάλι εγκλωβίζεται σε τοπικό ελάχιστο).

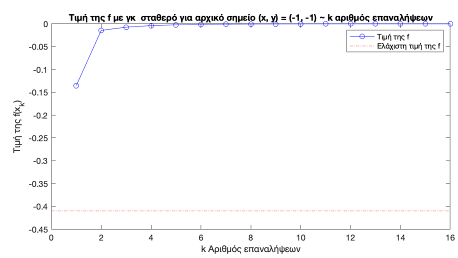
Για το σημείο (-1,-1) , φαίνεται και στις τρεις περιπτώσεις (γ_κ σταθερό,βέλτιστο και με μέθοδο Armijo), ότι υπάρχει σύγκλιση της τιμής της f και μάλιστα συγκλίνει στο πραγματικό ελάχιστο -0.4. Στη συνέχεια παρατίθεται κι ένας πίνακας που παρουσιάζει συγκριτικά τον αριθμό των επαναλήψεων k για κάθε σημείο και γ_k.

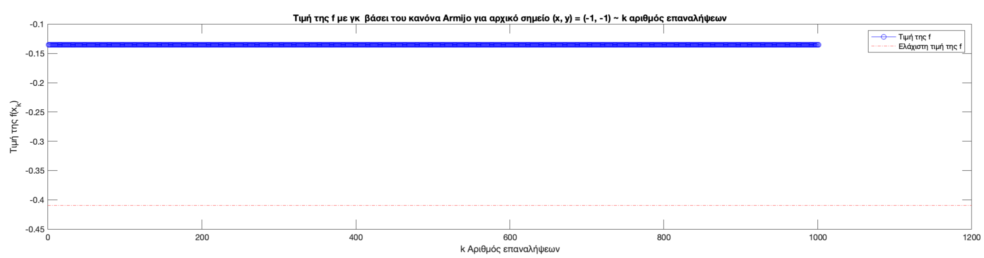
σημείο	γ=0.5	γ που ελαχιστοποεί $f(x_k + \gamma_\kappa d_k)$	Armijo
(1,1)	954	2	2
(-1,-1)	391	35	135
(0,0)	1	1	1

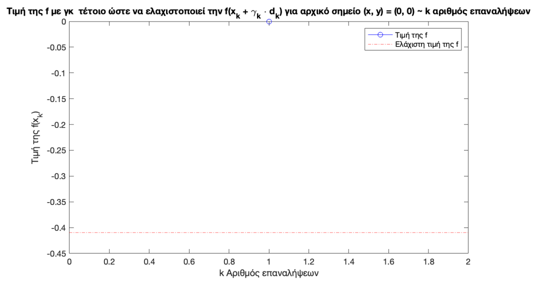
Θέμα 3

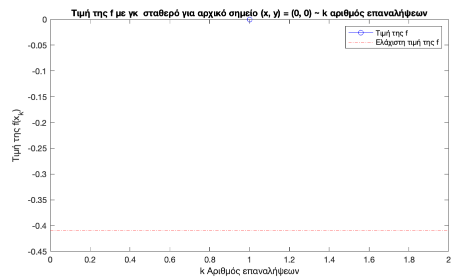
Εδώ καλούμαστε να ελαχιστοποιήσουμε την f χρησιμοποιώντας τη μέθοδο **Newton** για τις ίδιες περιπτώσεις που αναλύθηκαν και με την μέθοδο του θέματος 2. Τα αντίστοιχα γραφήματα δίνονται παρακάτω. Στη μέθοδο του **Newton** αυτό που αλλάζει είναι το διάνυσμα κατεύθυνσης το οποίο πλέον ισούται με $d_k = -[\nabla^2 f(x_k)]^{-1}\nabla f(x_k)$. Θεωρητική προϋπόθεση αποτελεί ο $\nabla^2 f(x_k)$ να είναι θετικά ορισμένος.

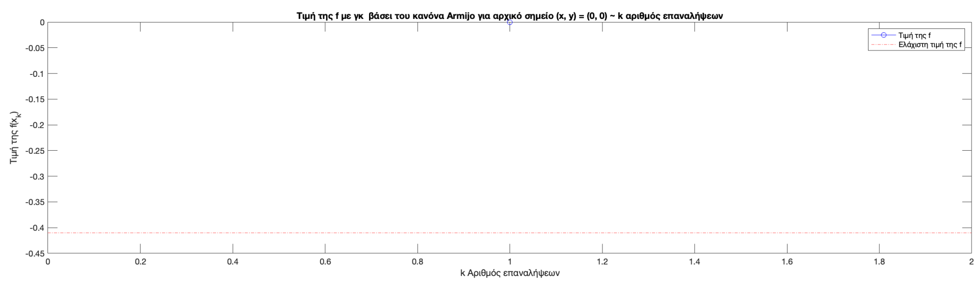


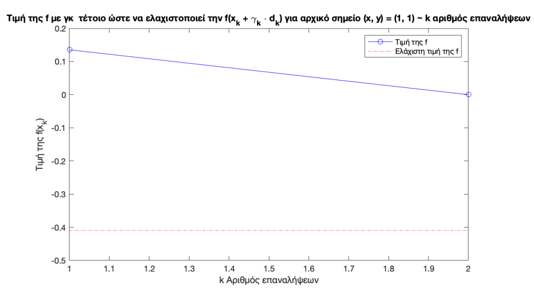


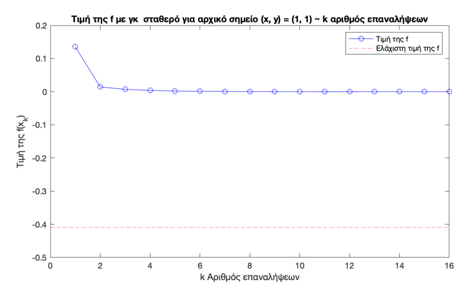


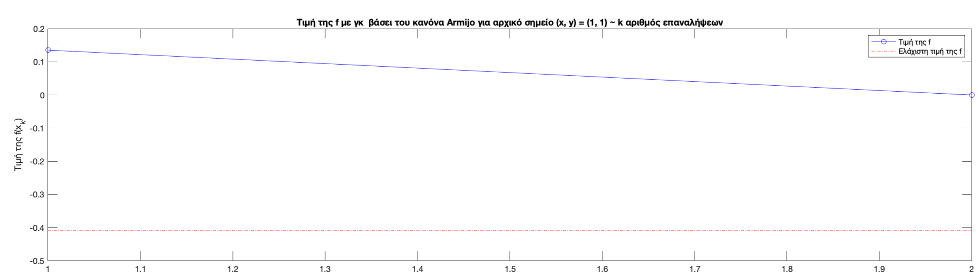

Τιμή της f με γκ τέτοιο ώστε να ελαχιστοποιεί την f(xk + γk · dk) για αρχικό σημείο (x, y) = (0, 0) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων


Τιμή της f με γκ σταθερό για αρχικό σημείο (x, y) = (0, 0) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων


Τιμή της f με γκ βάσει του κανόνα Armijo για αρχικό σημείο (x, y) = (0, 0) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων


Τιμή της f με γκ τέτοιο ώστε να ελαχιστοποιεί την f(xk + γk · dk) για αρχικό σημείο (x, y) = (1, 1) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων


Τιμή της f με γκ σταθερό για αρχικό σημείο (x, y) = (1, 1) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων

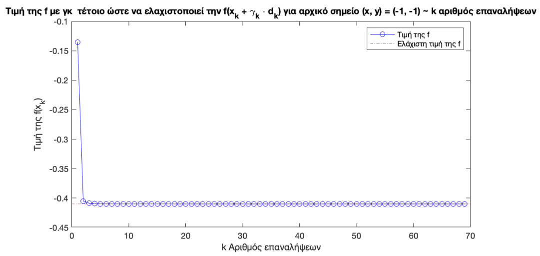

Τιμή της f με γκ βάσει του κανόνα Armijo για αρχικό σημείο (x, y) = (1, 1) ~ k αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f
Τιμή της f(xk)
k Αριθμός επαναλήψεων

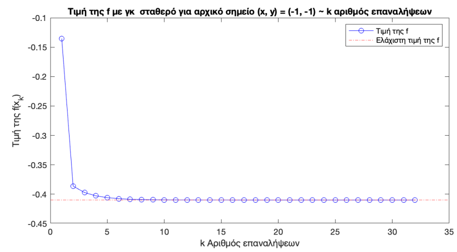
Για αρχικά σημεία (0,0) και (1,1) καταλήγουμε σε ακριβώς ίδια συμπεράσματα με την παραπάνω μέθοδο (με διαφορές στον αριθμό των επαναλήψεων).
Για το σημείο (-1,-1), υπάρχει σύγκλιση στην σωστή τιμή όταν χρησιμοποιούμε τη μέθοδο που ελαχιστοποεί την $f(x_k + \gamma_\kappa d_k)$, στις 9 επαναλήψεις.Ωστόσο, για γ_κ σταθερό και ίσο με 0.5, η τιμή της f ξεκινά από μια τιμή κοντά στο -0.15 και αυξάνεται έως ότου σταθεροποιηθεί στο 0 (μακριά από το πραγματικό ελάχιστο), τερματίζοντας στις 16 επαναλήψεις. Ουσιαστικά, σε αυτή την περίπτωση ο $\nabla^2 f(x_k)$ είναι αρνητικά ορισμένος και συνεπώς ο αλγόριθμος κάνει αναζήτηση ολικού μεγίστου (σταματώντας σε τοπικό μέγιστο).Ο αλγόριθμος δε κάνει διαχωρισμό μεταξύ σημείων ελαχίστου και σημείων μεγίστου. Τέλος, με τη μέθοδο Armijo , η τιμή $f(x_k)$ παραμένει από την αρχή σταθερή σε μια τιμή διαφορετική του ελαχίστου και τερματίζει λόγω μέγιστου αριθμού επαναλήψεων. Παρακάτω δίνεται πάλι το πινακάκι για τα διαφορετικά k σε κάθε περίπτωση.

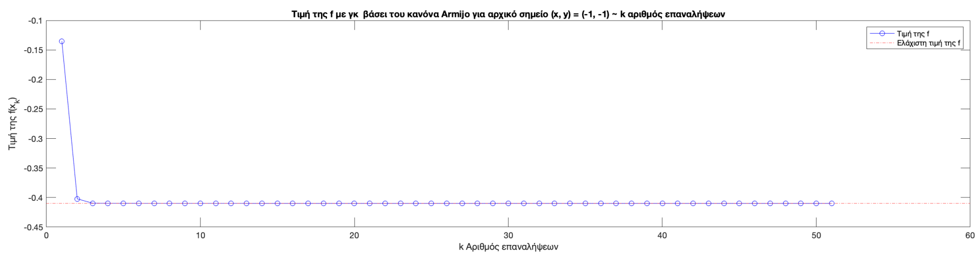
σημείο	γ=0.5	γ που ελαχιστοποεί $f(x_k + \gamma_\kappa d_k)$	Armijo
(1,1)	16	2	2
(-1,-1)	16	9	1001
(0,0)	1	1	1

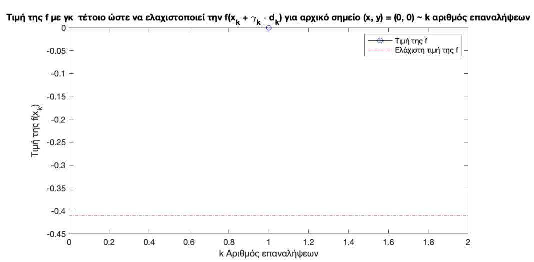
Θέμα 4

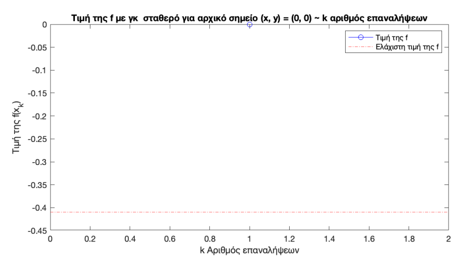
Στο 4ο και τελευταίο θέμα χρησιμοποείται η μέθοδος **Levenberg-Marquardt**. Η μέθοδος αποτελεί ουσιαστικά μια τροποποιημένη μορφή του Newton και χρησιμοποιείται όταν ο $\nabla^2 f(x_k)$ δεν είναι θετικά ορισμένος $\forall x \in \mathbb{R}^n$.Σε αυτήν την περίπτωση, υπολογίζουμε το μ_k έτσι ώστε ο $\nabla^2 f(x_k) + \mu_k I$ να είναι θετικά ορισμένος. Στη συνέχεια , ορίζουμε το διάνυσμα κατεύθυνσης ίσο με $d_k = -[\nabla^2 f(x_k) + \mu_k I]^{-1} \nabla f(x_k)$.Παρατηρείται ότι όταν το μ_k είναι αρκετά μεγάλο, ο παράγοντας $\mu_k I$ κυριαρχεί σε σχέση με τον $\nabla^2 f(x_k)$. Κάτα συνέπεια, η μέθοδος *Levenberg-Manquardt* θα λειτουργεί σχεδόν όπως η μέθοδος της μέγιστης καθόδου. Αν όμως το μ_k είναι μικρό, τότε ο $\nabla^2 f(x_k)$ υπερισχύει και η μέθοδος συμπεριφέρεται σα τη μέθοδο *Newton.*

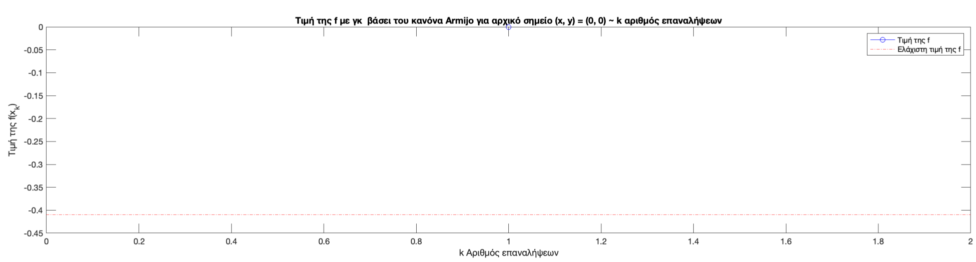

Τιμή της f με γκ τέτοιο ώστε να ελαχιστοποιεί την f(xk + γk · dk) για αρχικό σημείο (x, y) = (-1, -1) ~ k αριθμός επαναλήψεων
Τιμή της f(xk)
k Αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f

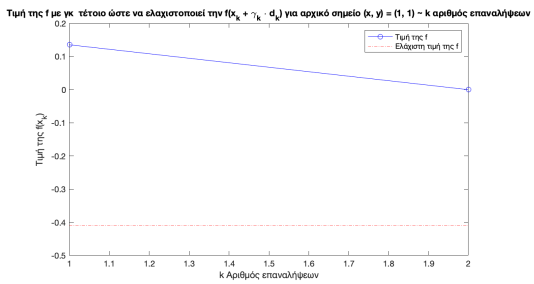

Τιμή της f με γκ σταθερό για αρχικό σημείο (x, y) = (-1, -1) ~ k αριθμός επαναλήψεων
Τιμή της f(xk)
k Αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f

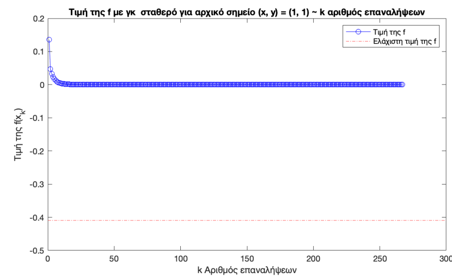

Τιμή της f με γκ βάσει του κανόνα Armijo για αρχικό σημείο (x, y) = (-1, -1) ~ k αριθμός επαναλήψεων
Τιμή της f(xk)
k Αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f

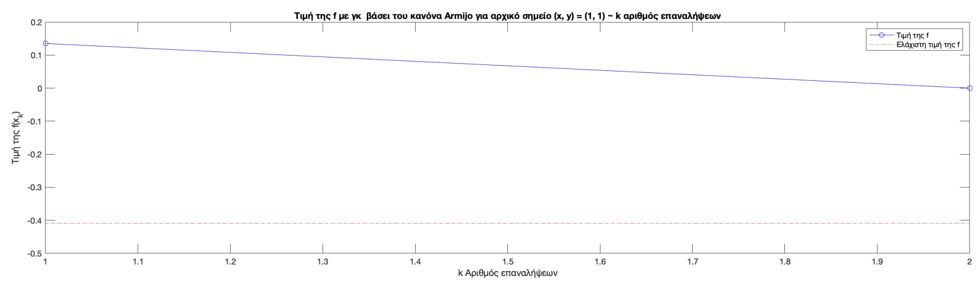

Τιμή της f με γκ τέτοιο ώστε να ελαχιστοποιεί την f(xk + γk · dk) για αρχικό σημείο (x, y) = (0, 0) ~ k αριθμός επαναλήψεων
Τιμή της f(xk)
k Αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f


Τιμή της f με γκ σταθερό για αρχικό σημείο (x, y) = (0, 0) ~ k αριθμός επαναλήψεων
Τιμή της f(xk)
k Αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f


Τιμή της f με γκ βάσει του κανόνα Armijo για αρχικό σημείο (x, y) = (0, 0) ~ k αριθμός επαναλήψεων
Τιμή της f(xk)
k Αριθμός επαναλήψεων
Τιμή της f
Ελάχιστη τιμή της f







Τα συμπεράσματα είναι και πάλι τα ίδια με αυτά που σχολιάσαμε στη μέθοδο της **Μέγιστης Καθόδου** με διαφορές στον αριθμό των επαναλήψεων. Στο (-1,-1) βλέπουμε πάλι ότι για οποιαδήποτε επιλογή γ_κ, η τιμή $f(x_k)$ συγκλίνει στη σωστή.

σημείο	γ=0.5	γ που ελαχιστοποεί $f(x_k + \gamma_\kappa d_k)$	Armijo
(1,1)	267	2	2
(-1,-1)	32	69	51
(0,0)	1	1	1

Σύνοψη

Από τις παραπάνω παρατηρήσεις γίνεται εμφανές, ότι η σύγκλιση της f στο σημείο ελαχίστου εξαρτάται έντονα τόσο από την επιλογή του αρχικού σημείο εκκίνησης $x_0 \in \mathbb{R}^n$, όσο και από τις διάφορες τιμές του γ_κ. Με άλλα λόγια, η σύγκλιση στο σωστό σημείο αποτελεί ένα πολυπαραγοντικό ζήτημα. Έτσι, όπως και στα περισσότερα προβλήματα βελτιστοποίησης αναγκαζόμαστε να συμβιβαστούμε. Η επιλογή που κάνουμε κάθε φορά περιορίζεται είτε από τα δεδομένα του προβλήματος, είτε από τους ίδιους τους αλγορίθμους που χρησιμοποιούμε, είτε από άλλες συνθήκες. Συνεπώς, η βέλτιστη λύση που αναζητούμε δεν είναι μια τέλεια λύση, απλά είναι η καλύτερη δυνατή δεδομένων όλων των περιορισμών που μας επιβάλλονται. Το σημείο (0,0) αποτελεί γενικά μια κακή επιλογή σε κάθε περίπτωση, καθώς στο σημείο αυτό η κλίση είναι εξαρχής μηδενική και ο αλγόριθμος περατώνεται. Με σημείο εκκίνησης το (1,1), η f συγκλίνει κάθε φορά στην περίπτωση που το γ είναι σταθερό. Για το σημείο (-1,-1) παρατηρούμε σύγκλιση για όλες τις τιμές γ_κ για τις μεθόδους **Μέγιστης Καθόδου** και **Levenberg-Marquardt** ενώ με τη μέθοδο **Newton** υπάρχει σύγκλιση μόνο στην περίπτωση που το γ_κ είναι τέτοιο ώστε να ελαχιστοποεί την $f(x_k + \gamma_\kappa d_k)$. Γενικά, η καλύτερη επιλογή όσον αφορά τη σύγκλιση στο σωστό σημείο φαίνεται να είναι αυτή με σημείο εκκίνησης το (-1,-1). Έπειτα, οι λιγότερες επαναλήψεις παρατηρούνται με τα γ_k εκείνα που ελαχιστοποιούν την $f(x_k+\gamma_\kappa d_k)$. Άρα ίσως είναι βάσιμο να επιλέξουμε συνδυασμό του σημείου (-1,-1) με αυτή τη μέθοδο επιλογής γ_k.