



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΤΜΗΜΑ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

Εργαστήριο Δυναμικής Μηχανών

Αριθμητική Βελτιστοποίηση σε Μηχανολογικές Κατασκευές

1^η Εργασία

”Εύρεση ελαχίστου συνάρτησης”

**Χρήστος Γεωργιάδης
ΑΕΜ: 6496**

Περιεχόμενα

1 Παρουσίαση Συνάρτησης	2
2 Μέθοδος Newton	3
2.1 Ο Εσσιανός της συνάρτησης	3
2.2 Εφαρμογή της Newton με χρήση των Wolfe conditions	4
2.2.1 Περίπτωση για $n = 2$ μεταβλητές	5
2.2.2 Περίπτωση για $n = 5$ μεταβλητές	8
3 Μέθοδος Steepest descent	11
3.1 Steepest descent με σταθερό μήκος βήματος	11
3.1.1 Περίπτωση για $n = 2$ μεταβλητές	13
3.2 Steepest descent με inexact line search για το μήκος βήματος	15
3.2.1 Περίπτωση για $n = 2$ μεταβλητές	15
3.2.2 Περίπτωση για $n = 5$ μεταβλητές	17
4 Μέθοδος Conjugate Gradient	18
4.1 Περίπτωση για $n = 2$ μεταβλητές	18
4.2 Περίπτωση για $n = 5$ μεταβλητές	20
5 Συμπεράσματα	22
6 Βιβλιογραφία	23

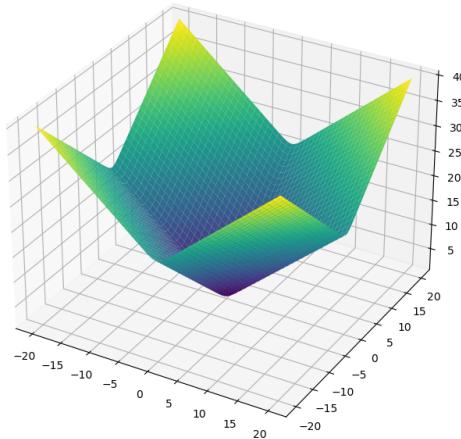
1 Παρουσίαση Συνάρτησης

Στην παρούσα εργασία θα μελετήσουμε και συγκερκιμένα θα προσπαθήσουμε να ελαχιστοποιήσουμε τη συνάρτηση *diagonal 5*, η οποία έχει την ακόλουθη μορφή:

$$f(\vec{x}) = \sum_{i=1}^n \log(e^{x_i} + e^{-x_i})$$

όπου $\vec{x} = [x_1, x_2, \dots, x_n]^T$ και n το πλήθος των μεταβλητών. Συγκεκριμένα, θα εξετάσουμε τις περιπτώσεις για $n = 2$ και $n = 5$ μεταβλητές.

Αρχικά κατασκευάζουμε μια 3D απεικόνιση της συνάρτησης ώστε να δούμε τη μορφή της.



Σχήμα 1: 3D απεικόνιση της συνάρτησης

Όπως φαίνεται, η συνάρτηση παρουσιάζει ελάχιστο και μάλιστα ολικό, οπότε θα προσπαθήσουμε να το βρούμε εφαρμόζοντας τρεις διαφορετικές μεθόδους:

- Μέθοδος Newton
- Μέθοδος Steepest Descent
- Μέθοδος Conjugate Gradient

Πρωτού προχωρήσουμε στην εφαρμογή των μεθόδων, θα υπολογίσουμε τον Ιακωβιανό και Εστιανό πίνακα της συνάρτησης, διότι θα χρειαστούν κατά την υλοποίηση των αλγορίθμων.

Ο Ιακωβιανός πίνακας της συνάρτησης είναι:

$$J = \nabla f = \begin{bmatrix} \frac{e^{x_1} - e^{-x_1}}{e^{x_1} + e^{-x_1}} \\ \vdots \\ \frac{e^{x_n} - e^{-x_n}}{e^{x_n} + e^{-x_n}} \end{bmatrix}$$

Αντίστοιχα, ο Εσσιανός πίνακας είναι:

$$H = \nabla^2 f = \begin{bmatrix} \frac{4}{(e^{x_1} + e^{-x_1})^2} & 0 & \cdots & 0 \\ 0 & \frac{4}{(e^{x_2} + e^{-x_2})^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{4}{(e^{x_n} + e^{-x_n})^2} \end{bmatrix}$$

2 Μέθοδος Newton

2.1 Ο Εσσιανός της συνάρτησης

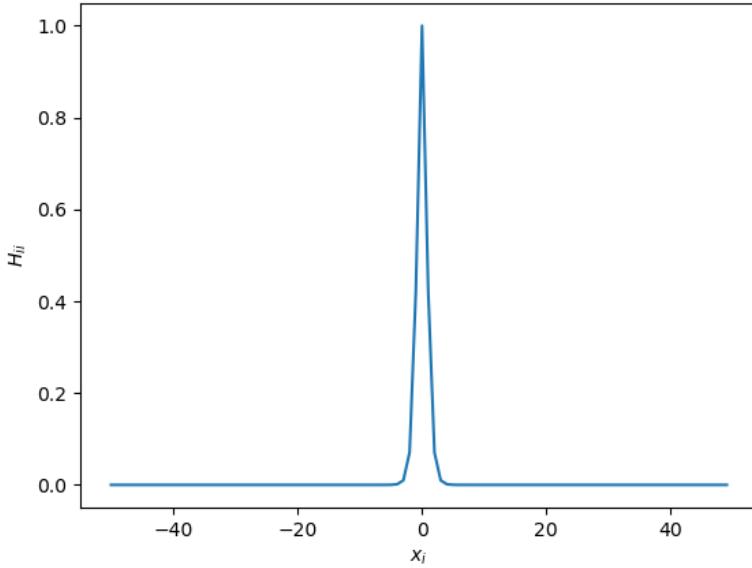
Η εφαρμογή της μεθόδου Newton για την έρευνη του ελαχίστου της συνάρτησης απαιτεί τον υπολογισμό του Εσσιανού πίνακα και την αντιστροφή του.

Ωστόσο, ο Εσσιανός της συγκεκριμένης συνάρτησης παρατηρούμε ότι εμφανίζει μία ιδιαιτερότητα: είναι σχεδόν ιδιάζων. Αυτό σημαίνει ότι καθώς οι τιμές των x_i αυξάνονται τα στοιχεία του πίνακα τείνουν στο μηδέν, δηλαδή:

$$\lim_{x_i \rightarrow \infty} H_{ii} = \lim_{x_i \rightarrow \infty} \frac{4}{(e^{x_i} + e^{-x_i})^2} = 0$$

όπου $i = 1, 2, \dots, n$.

Αυτό φαίνεται επίσης και στο παρακάτω διάγραμμα αν σχεδιάσουμε τη μορφή της σύναρτησης των στοιχείων του εσσιανού πίνακα:



Σχήμα 2: Μορφή της συνάρτησης H_{ii}

Βλέπουμε ότι τα στοιχεία H_{ii} του εσσιανού παίρνουν μη μηδενικές τιμές για ένα μικρό εύρος των x_i . Επομένως, αν ξεκινήσουμε από κάποιο σημείο εκτός αυτού του μικρού εύρους τιμών των x_i ή αν αλγόριθμος σε κάποια επανάληψη οδηγηθεί σε ένα σημείο εκτός του εύρους αυτού, τότε δεν θα μπορέσει να υπολογιστεί ο αντίστροφος του εσσιανού πίνακα και η μέθοδος θα αποτύχει. Καταλαβαίνουμε, λοιπόν, ότι το γεγονός ότι ο εσσιανός πίνακας της συνάρτησης είναι σχεδόν ιδιάζων, καθιστά την μέθοδο Newton, στην απλή της μορφή, εξαιρετικά ασταθή.

2.2 Εφαρμογή της Newton με χρήση των Wolfe conditions

Γνωρίζουμε ότι κατά την απλή μορφή της μεθόδου Newton, η καινούργια προσέγγιση της λύσης υπολογίζεται επαναληπτικά ως:

$$x_{k+1} \leftarrow x_k + \alpha p_k$$

όπου $p_k = -(\nabla^2 f)^{-1} \nabla f$ και $\alpha = 1$.

Ωστόσο, όπως προαναφέρθηκε, όταν ο αλγόριθμος Newton ξεκινάει από ένα μακρινό σημείο, οι τιμές των στοιχείων του αντίστροφου του εσσιανού $(\nabla^2 f)^{-1}$ είναι αρκετά μεγάλες. Μάλιστα, σύμφωνα με τα παραπάνω, χρησιμοποιώντας ως βήμα $\alpha = 1$, δεν επιφέρεται καμία μείωση με αποτέλεσμα το διάνυσμα κατεύθυνσης p_k να παίρνει επίσης μεγάλες τιμές και το νέο σημείο να απομακρύνεται πάρα πολύ από τη λύση.

Ως μία λύση στο παραπάνω πρόβλημα, χρησιμοποιήθηκε ένα *inexact line search* με συνθήκες τερματισμού τα *Wolfe conditions* για την εύρεση ενός κατάλληλου μήκους βήματος α . Ουσιαστικά, βρίσκει ένα πολύ μικρό βήμα ώστε να επιφέρει μειώση στους όρους του γινομένου $\alpha(\nabla^2 f)^{-1} \nabla f$.

Μάλιστα, το συγκεκριμένο line search δέχεται σε κάθε επανάληψη ως αρχικό μήκος βήματος $\alpha = 1$, ελέγχει αν ικανοποιεί τα Wolfe conditions και αν όχι τότε ψάχνει για άλλο. Αυτό το κάνουμε διότι εξ' ορισμού ο αλγόριθμος Newton έχει μήκος βήματος μονάδα, επομένως όπου αυτό είναι εφικτό θέλουμε να το επιλέξουμε.

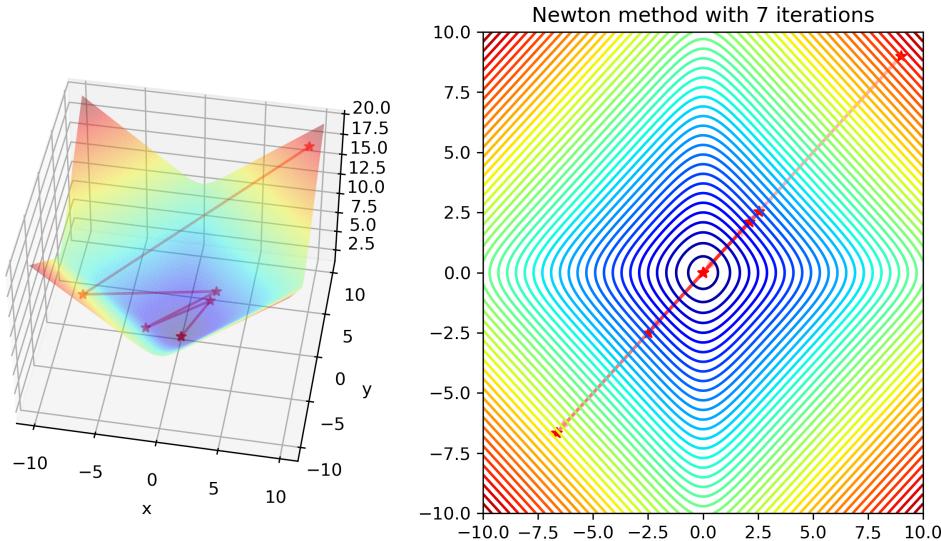
Παρόλα αυτά, ακόμη και αυτή η παραλλαγή της μεθόδου δεν λειτουργεί σε όλες τις περιπτώσεις. Αν το σημείο είναι αρκετά μακριά δυσκολεύεται να βρει ένα κατάλληλο, πολύ μικρό μήκος βήματος και αποτυγχάνει.

Στη συνέχεια ακολούθοιν τα αποτελέσματα της μεθόδου και η οπτικοποίηση τους για διάφορες αρχικές συνθήκες και για αριθμό μεταβλητών $n = 2$ και $n = 5$.

2.2.1 Περίπτωση για $n = 2$ μεταβλητές

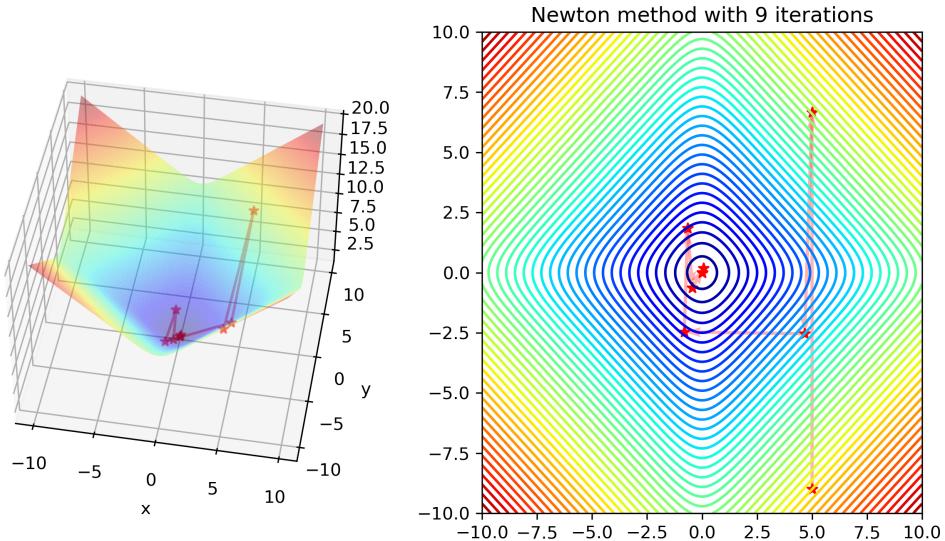
Δοκιμάστηκαν διάφορα αρχικά σημεία, κάποια κοντά στη βέλτιστη λύση και κάποια μακριά από αυτήν. Παρακάτω παρατίθενται μερικά από αυτά.

1. $x_0 = [9, 9]^T$



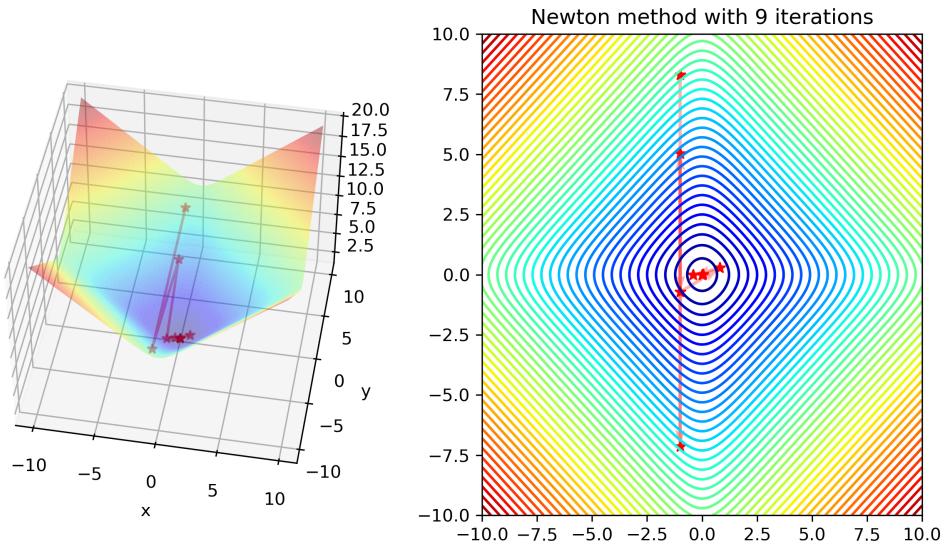
Σχήμα 3

2. $x_0 = [5, -9]^T$



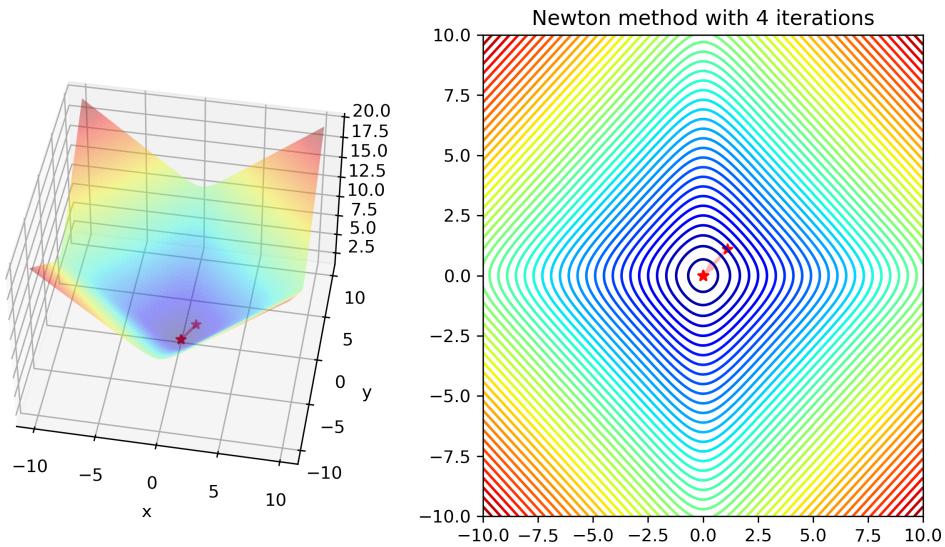
$\Sigma\chi\eta\mu\alpha 4$

3. $x_0 = [-1, 8.3]^T$

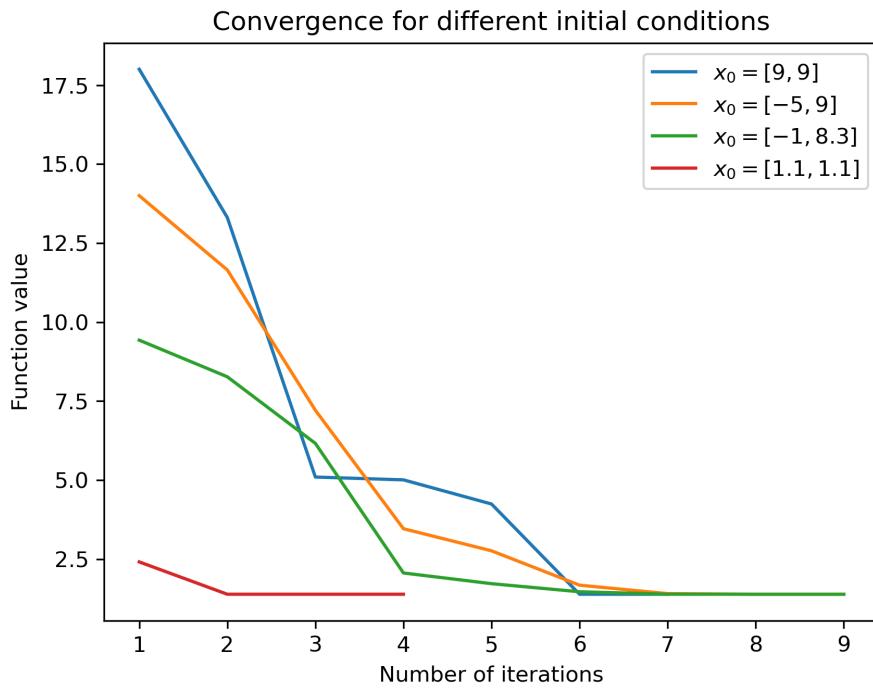


$\Sigma\chi\eta\mu\alpha 5$

4. $x_0 = [1.1, 1.1]^T$



Σχήμα 6

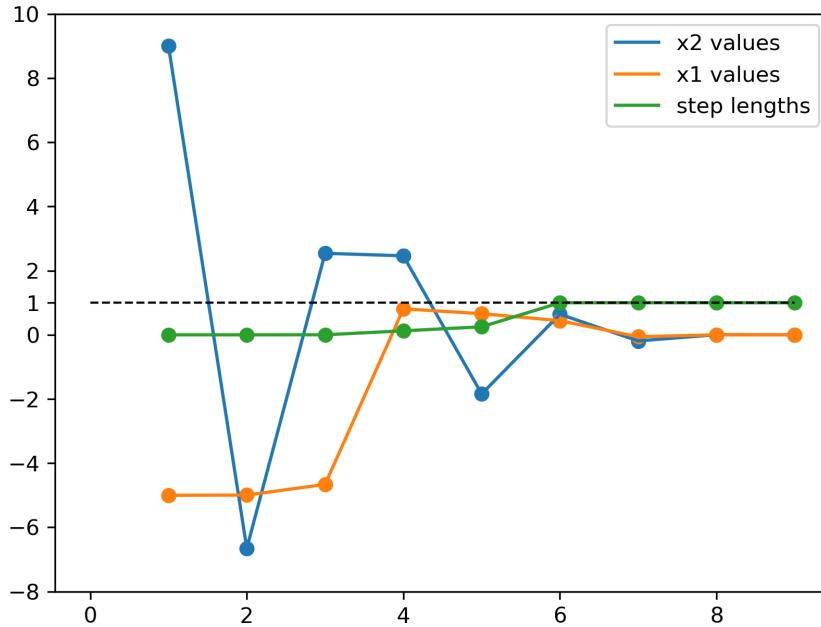


Σχήμα 7

Στο τελευταίο διάγραμμα (Σχήμα 7) παρατηρούμε την ταχύτητα σύγκλισης της μεθόδου όταν αυτή ξεκινάει από διάφορα αρχικά σημεία. Όπως είναι αναμενόμενο, η ταχύτητα μειώνεται όσο πιο μακριά από τη βέλτιστη λύση είναι οι αρχικές συνθήκες. Αξίζει να σημειωθεί ότι, αν οι

συντεταγμένες του σημείου λάβουν τιμές μεγαλύτερες του 9, η μέθοδος αποκλίνει διότι δεν μπορεί να βρεθεί ένα κατάλληλο μήκος βήματος α .

Επίσης, στο παρακάτω γράφημα φαίνεται για την περιπτώση του $x_0 = [5, -9]^T$, η επιλογή του μήκους βήματος που κάνει ο αλγόριθμος σε κάθε επανάληψη (iteration).

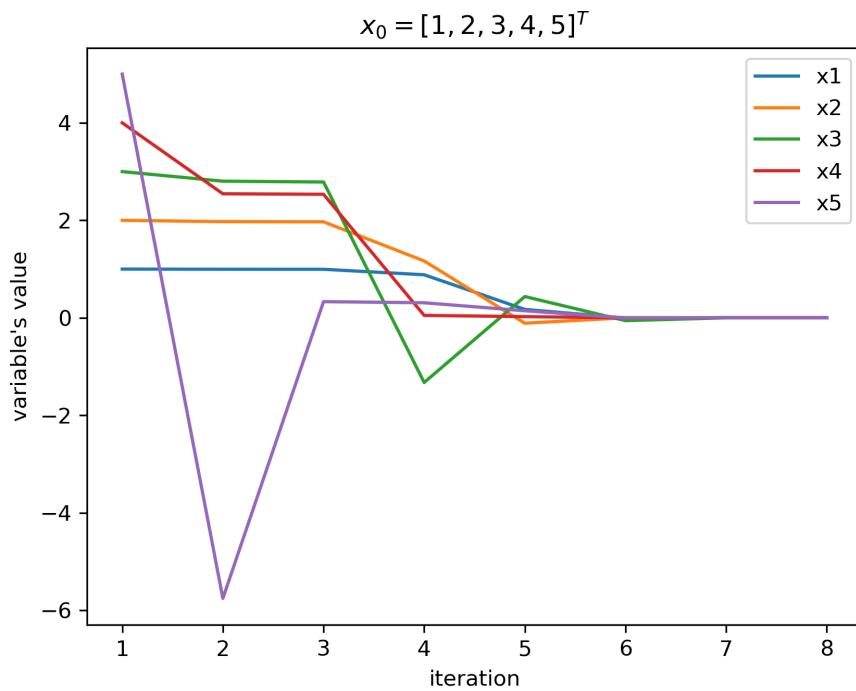


Σχήμα 8

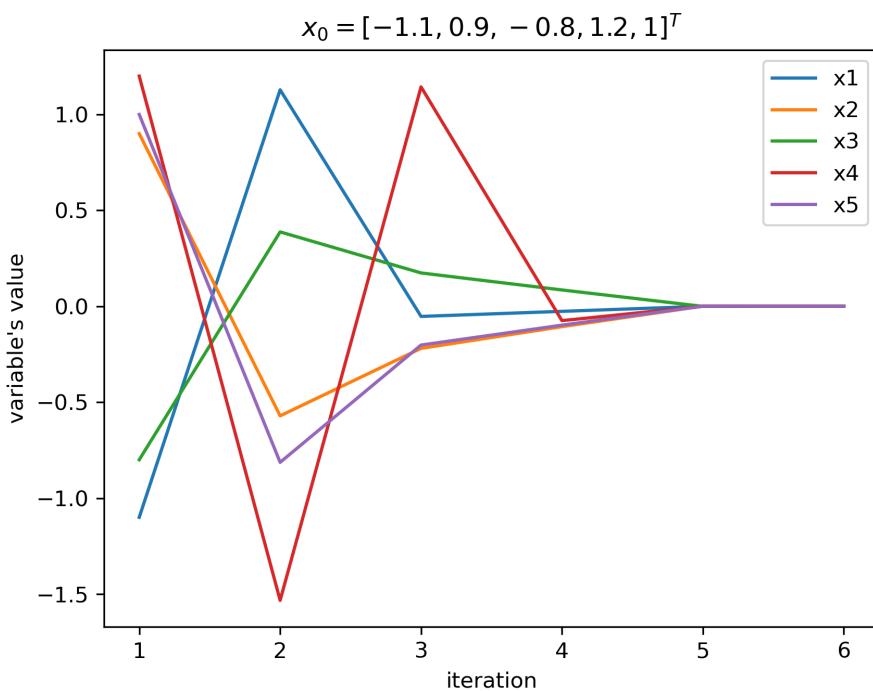
Παρατηρούμε, αυτό που ανεφέρθηκε και παραπάνω. Αρχικά, επιλέγονται πολύ μικρά μήκη βήματος, κόντα στο μηδέν, και ύστερα από ένα σημείο επιλέγεται ως μήκος βήματος η μονάδα.

2.2.2 Περίπτωση για $n = 5$ μεταβλητές

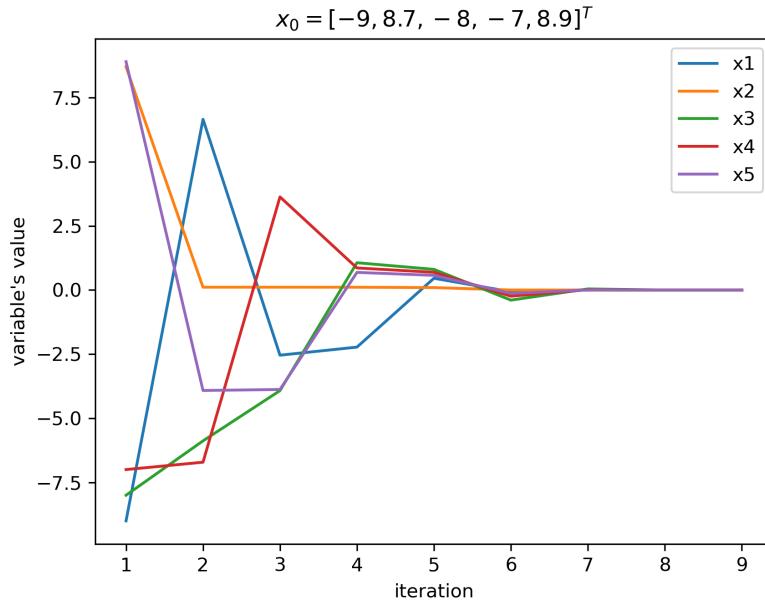
Η μέθοδος δίνει αντίστοιχα αποτελέσματα και για την περιπτώση με περισσότερες μεταβλητές ($n = 5$). Παρακάτω παρατίθενται ενδεικτικά κάποια διαγραμμάτα που απεικονίζουν τη σύγκλιση των μεταβλητών από διαφορετικές αρχικές συνθήκες.



$\Sigma\chi\eta\mu\alpha\ 9$

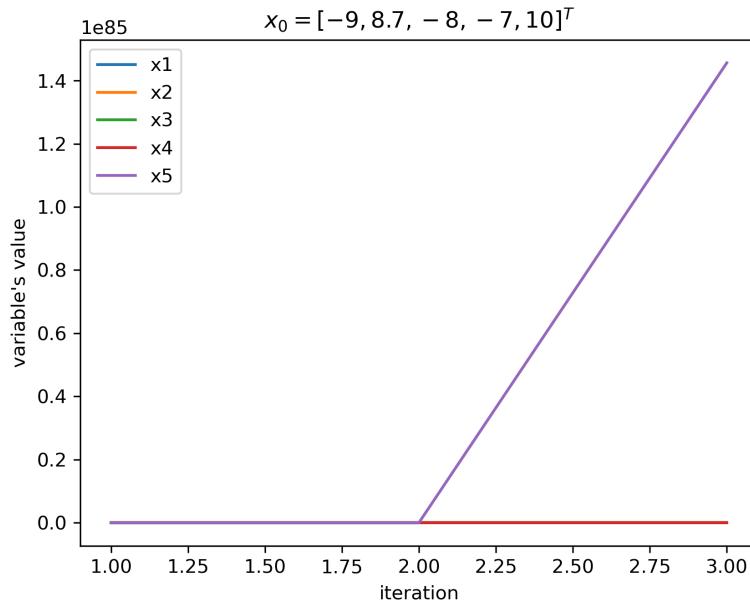


$\Sigma\chi\eta\mu\alpha\ 10$



Σχήμα 11

Παρατηρούμε ότι όλες οι μεταβλητές συγκλίνουν στη λύση σε κάθε περίπτωση αρχικής συνθήκης. Ωστόσο, όπως ίσχυε και προηγουμένως, αν κάποια από τις μεταβλητές λάβει τιμή μεγαλύτερη του 9, η μέθοδος αποκλίνει. Αυτό μπορούμε να το δούμε στο Σχήμα 12, όπου ως αρχικό σημείο δίνεται το ίδιο με αυτό του Σχήματος 11, με τη μόνη διαφορά ότι η αρχική τιμή της μεταβλητής x_5 τίθεται ίση με 10.



Σχήμα 12

Στο παραπάνω διάγραμμα φαίνεται ότι η μεταβλητή x_5 στην τρίτη επανάληψη αποκλίνει στο άπειρο. Ουσιαστικά, η μέθοδος αναζήτησης με τα Wolfe conditions αποτυγχάνει να βρει ένα κατάλληλο μήκος βήματος.

3 Μέθοδος Steepest descent

Η μέθοδος steepest descent εφαρμόστηκε με δύο τρόπους: α) χρησιμοποιώντας ένα σταθερό, μικρό μήκος βήματος a και β) χρησιμοποιώντας ένα inexact line search για την εύρεση του βέλτιστου μήκους βήματος σε κάθε επανάληψη.

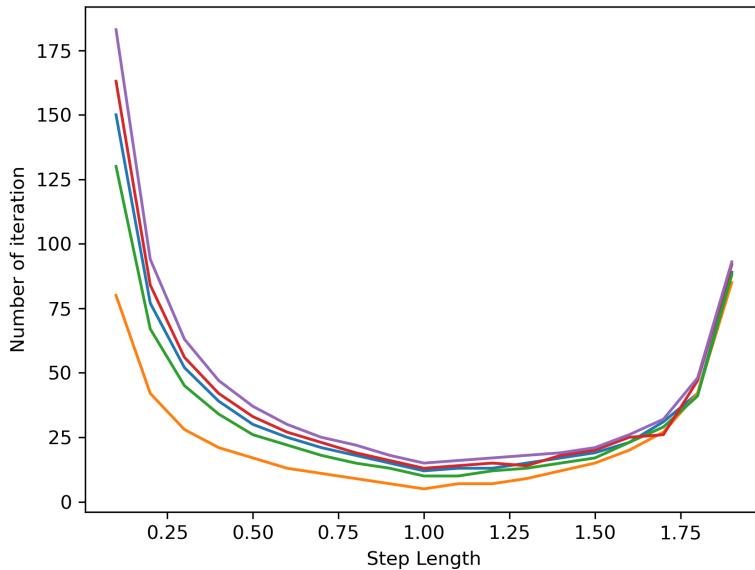
Ως συνθήκες τερματισμού της μεθόδου ορίστηκαν οι εξής:

- Απαιτούμενη ακρίβεια: $\epsilon = 10^{-4}$
- Μέγιστος αριθμός επαναλήψεων: 10000

Οπώς είναι γνωστό η μέθοδος steepest descent βασίζεται μόνο στον υπολογισμό του διανύσματος κλίσης της συνάρτησης και δεν απαιτεί τον υπολογισμό του εσσιανού πίνακα. Επομένως, δεν έχουμε το πρόβλημα που είχαμε με τη μέθοδο Newton.

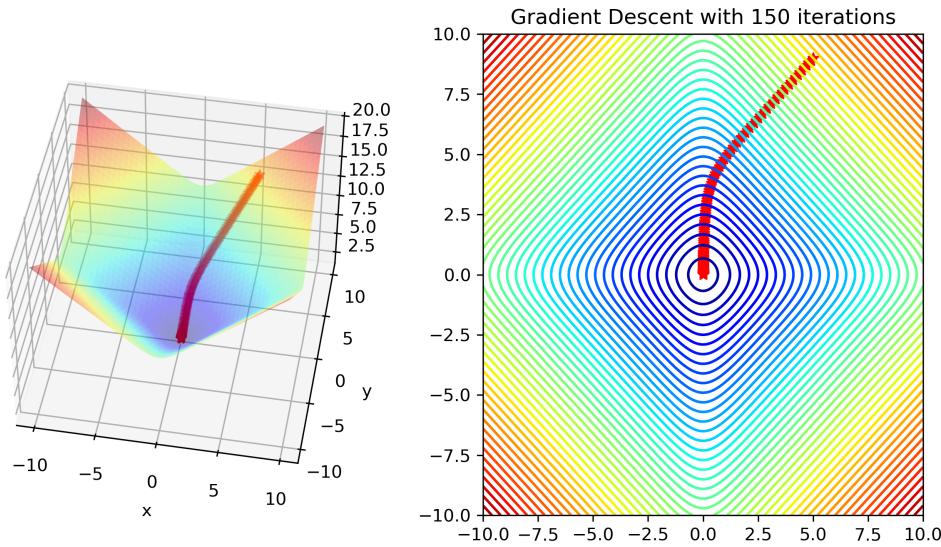
3.1 Steepest descent με σταθερό μήκος βήματος

Η συγκεκριμένη μορφή της μεθόδου δοκιμάστηκε για διάφορα αρχικά σημεία και για διάφορα μήκη βήματος. Παρακάτω παρουσιάζεται ένα συγκεντρωτικό διάγραμμα, όπου κάθε καμπύλη ανταποκρίνεται σε διαφορετικό αρχικό σημείο. Η κάθε καμπύλη παριστάνει τον αριθμό των επαναλήψεων που χρειάστηκαν σε συνάρτηση με διαφορετικά μήκη βήματος.

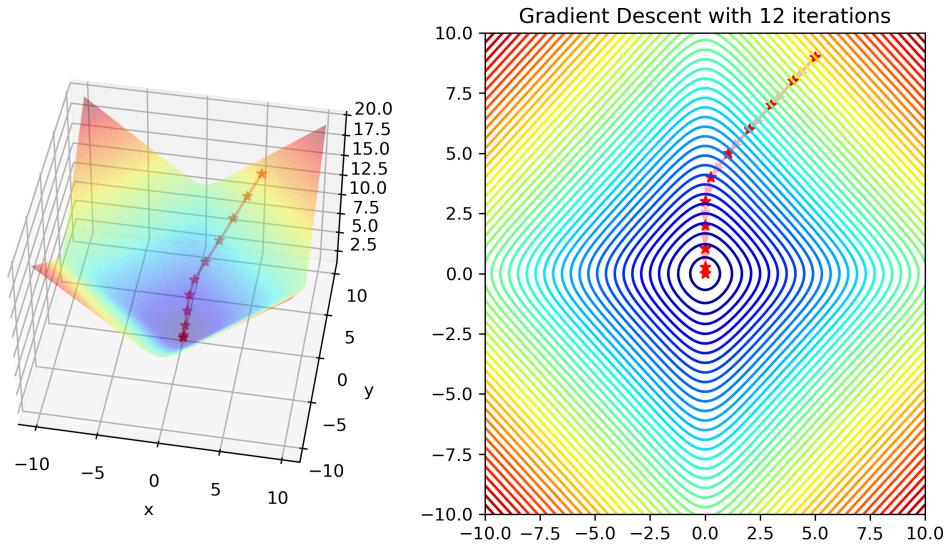


Σχήμα 13

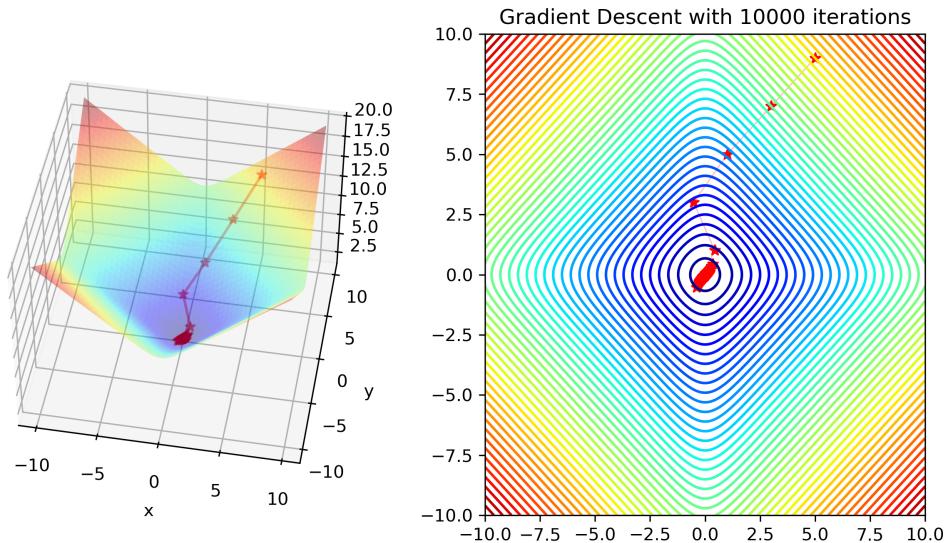
Παρατηρούμε ότι σε κάθε περίπτωση, δηλαδή για οποιδήποτε αρχικό σημείο, το βέλτιστο μήκος βήματος είναι $\alpha = 1$. Για μήκη βήματος μικρότερα από αυτό η μεθόδος συγκλίνει πιο αργά, ενώ για μεγαλύτερα βήματα η μέθοδος παρουσιάζει μία ασταθή συμπεριφορά και αυξάνεται ο αρθιμός των επαναλήψεων. Αν αυξήσουμε ακόμη περισσότερο το μήκος βήματος, η μέθοδος αποκλίνει. Ουσιαστικά, φτάνει σε ένα σημείο κοντά στη λύση και έπειτα "ταλαντώνεται" γύρω από αυτό μέχρις ότου φτάνει τον μέγιστο αριθμό επαναλήψεων. Παρακάτω παρατίθενται, ενδεικτικά, τρία διαγράμματα που απεικονίζουν αυτές τις τρεις περιπτώσεις.



Σχήμα 14: Περίπτωση αργής σύγκλισης με μήκος βήματος $\alpha = 0.1$



Σχήμα 15: Περίπτωση ταχύτερης σύγκλισης με μήκος βήματος $\alpha = 1$

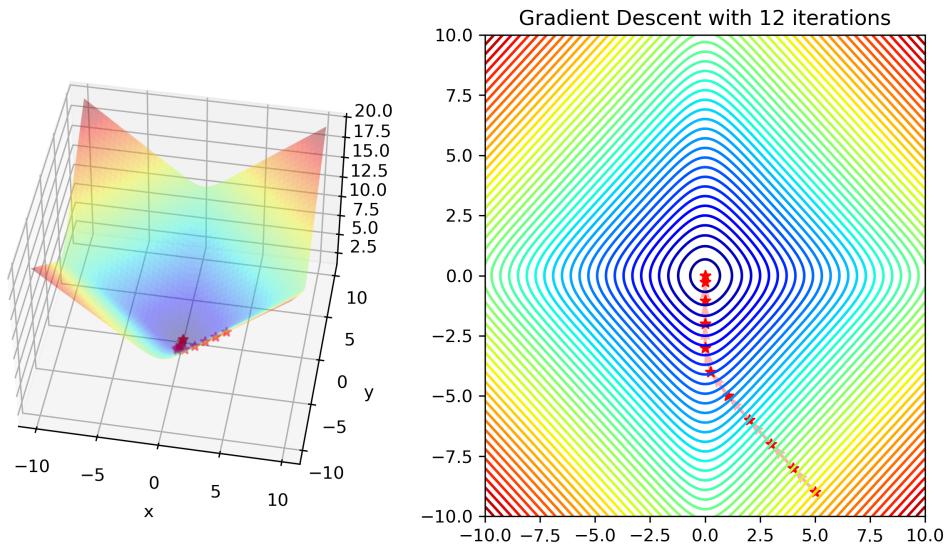


Σχήμα 16: Περίπτωση όπου η μέθοδος αποκλίνει με μήκος βήματος $\alpha = 2$

3.1.1 Περίπτωση για $n = 2$ μεταβλητές

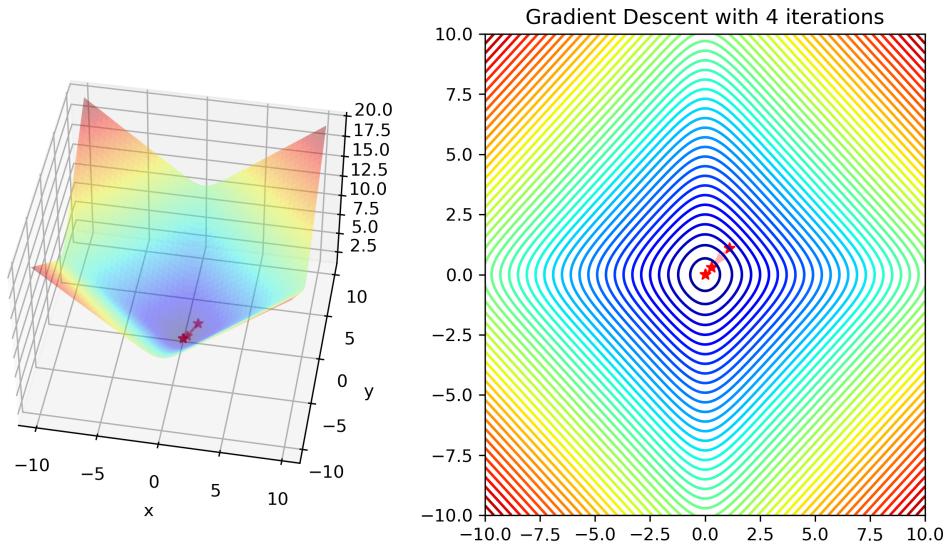
Στη συνέχεια παρουσιάζονται μερικά διαγράμματα της μεθόδου για την περίπτωση με δύο μεταβλητές. Επιλέγεται μήκος βήματος $\alpha=1$ και δύο αρχικά σημεία που χρησιμοποιήθηκαν και στη μέθοδο Newton για λόγους σύγκρισης των μεθόδων.

$$1. \quad x_0 = [5, -9]^T$$



Σχήμα 17

$$2. \quad x_0 = [1.1, 1.1]^T$$



Σχήμα 18

Συγκρίνοντας τα διαγράμματα με αυτά της μεθόδου Newton, παρατηρούμε ότι η steepest descent είναι πιο αργή.

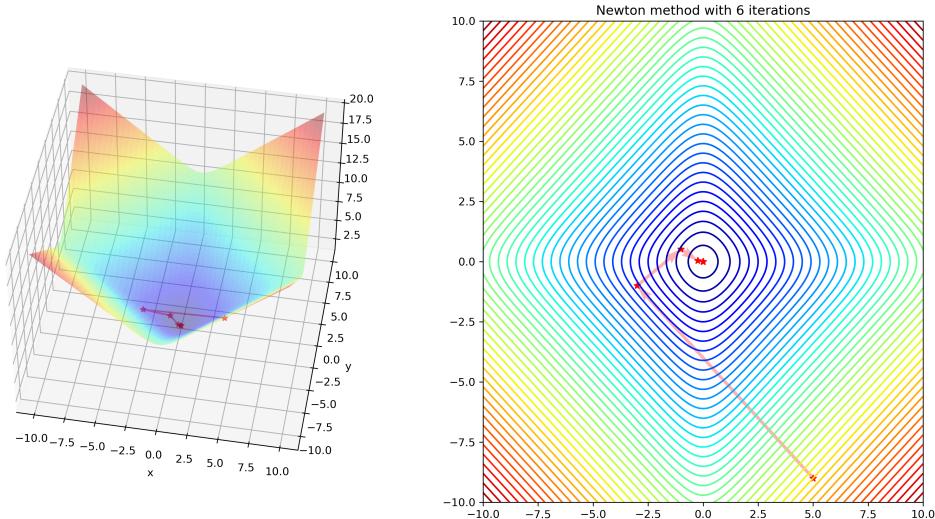
3.2 Steepest descent με inexact line search για το μήκος βήματος

Για να βελτιώσουμε την απόδοση της μεθόδου steepest descent, αντί για σταθερό μήκος βήματος, χρησιμοποιούμε ένα inexact line search το οποίο σε κάθε επανάληψη βρίσκει το βέλτιστο α με κριτήριο και πάλι τα Wolfe conditions.

3.2.1 Περίπτωση για $n = 2$ μεταβλητές

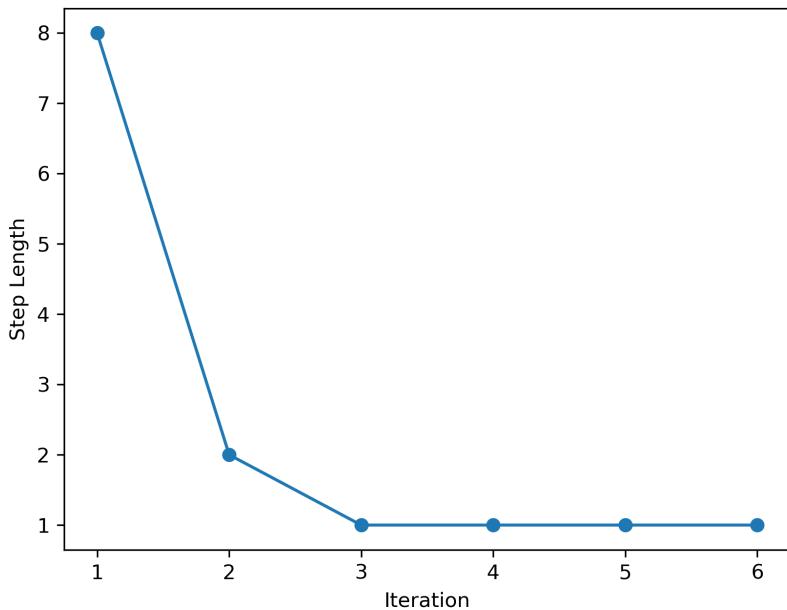
Παρακάτω βλέπουμε εφαρμογές αυτής της παραλλαγής της μεθόδου με δύο μεταβλητές, ξεκινώντας από διάφορα αρχικά σημεία. Για λόγους σύγκρισης χρησιμοποιούμε τα ίδια με αυτά που χρησιμοποιήσαμε στη μέθοδο Newton.

1. $x_0 = [5, -9]^T$



Σχήμα 19

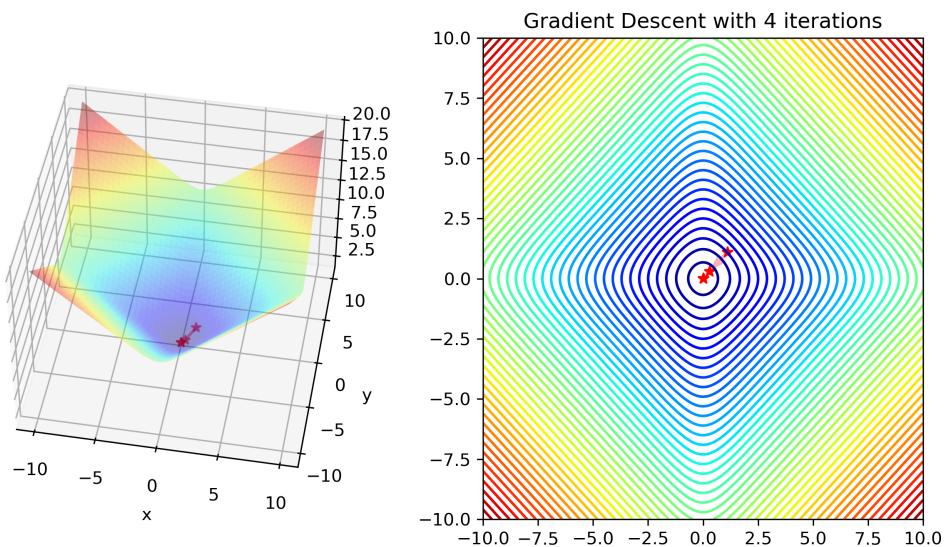
Μπορούμε να δούμε ότι η μέθοδος συγκλίνει πιο γρήγορα σε σχέση με πριν που είχαμε σταθερό μήκος βήματος, αφού επιλέγει πιο μεγάλο μήκος βήματος όταν είναι μακριά από τη λύση και μικρότερο καθώς πλησιάζει σε αυτήν. Για τη συγκεκριμένη περίπτωση μποορύμε να δούμε ακριβώς τι μήκος βήματος επιλέγεται σε κάθε επανάληψη στο Σχήμα 20.



Σχήμα 20

Παρατηρούμε ότι καθώς φτάνει κοντά στη λύση το μήκος βήματος $\alpha=1$ ικανοποιεί τα conditions και επιλέγεται ως βέλτιστο.

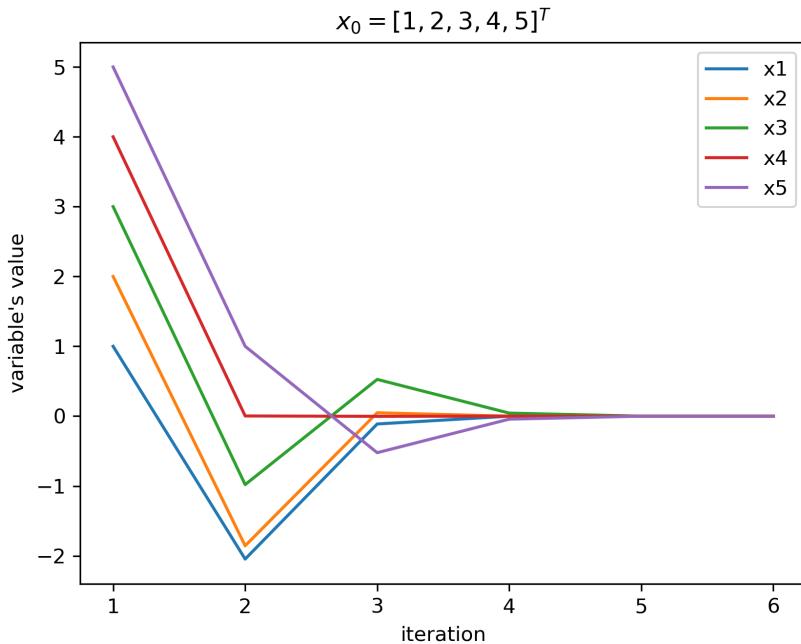
2. $x_0 = [1.1, 1.1]^T$



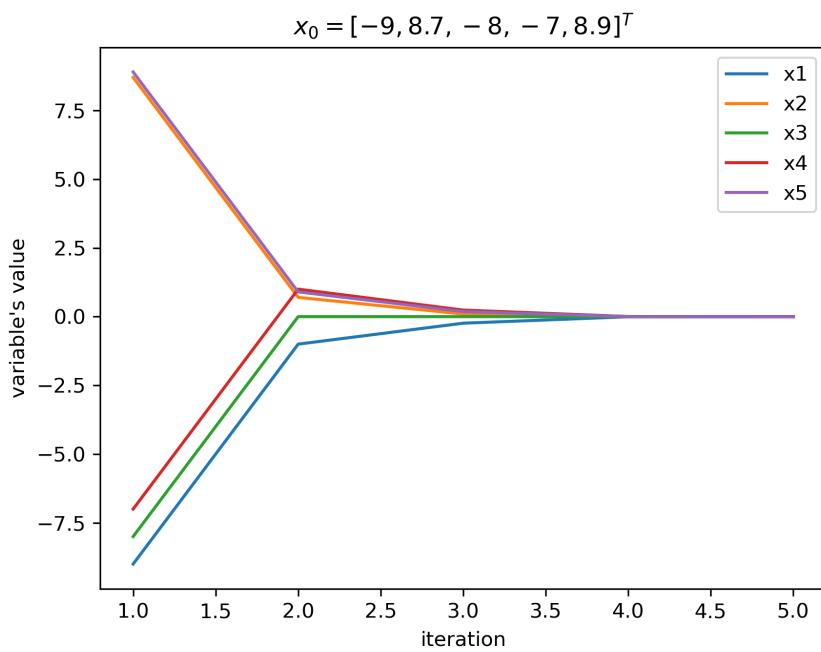
Σχήμα 21

3.2.2 Περίπτωση για $n = 5$ μεταβλητές

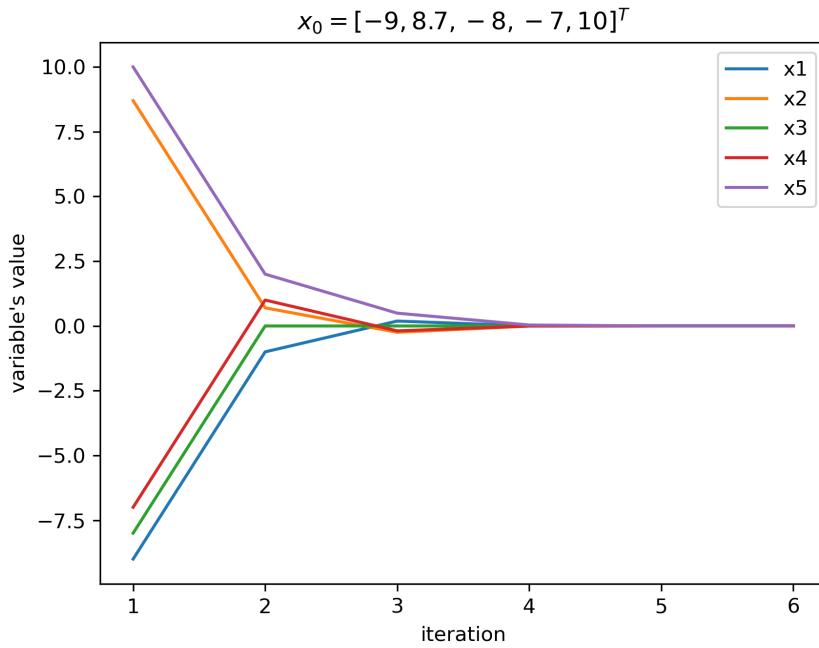
Εφαρμόζοντας τη μέθοδο για πέντε μεταβλητές προκύπτουν τα ακόλουθα διαγράμματα.



Σχήμα 22



Σχήμα 23



Σχήμα 24

Συγκρίνοντας αυτά τα διαγράμματα με τα αντίστοιχα της μεθόδου Newton, θα παρατηρήσουμε ότι αυτή η μέθοδος συγκλίνει πιο γρήγορα, σε κάθε περίπτωση. Επίσης, δεν εμφανίζει το πρόβλημα της Newton και οι μεταβλητές μπορούν να λάβουν οποιαδήποτε τιμή.

4 Μέθοδος Conjugate Gradient

Τέλος, εφαρμόζεται η μέθοδος conjugate gradient, και συγκεκριμένα ο αλγόριθμος *Fletcher - Reeves* επειδή η συνάρτηση είναι μη γραμμική.

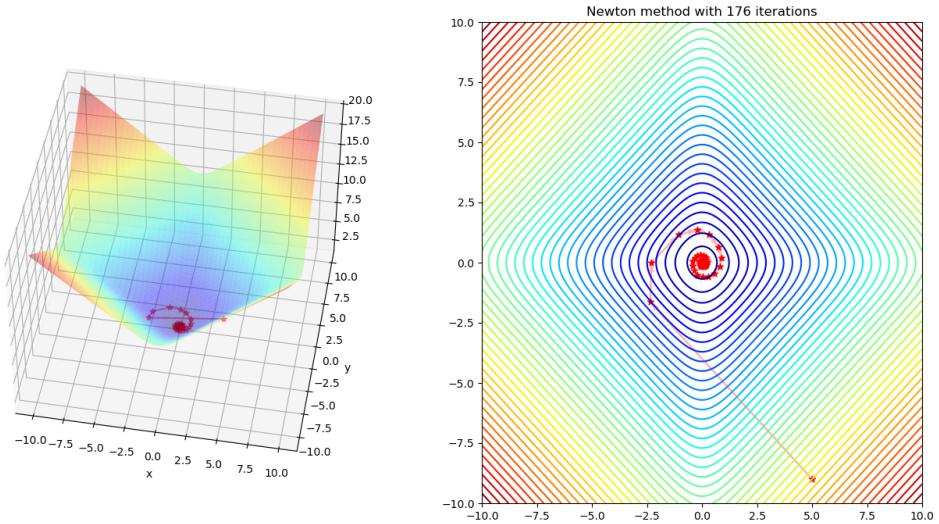
Οι συνθήκες τερματισμού ορίζονται οι εξής:

- Απαιτούμενη ακρίβεια: $\epsilon = 10^{-4}$
- Μέγιστος αριθμός επαναλήψεων: 200

Σύμφωνα με αυτόν τον αλγόριθμο, χρησιμοποείται και πάλι ένα line search για την εύρεση του βέλτιστου μήκους βήματος σε κάθε επανάληψη.

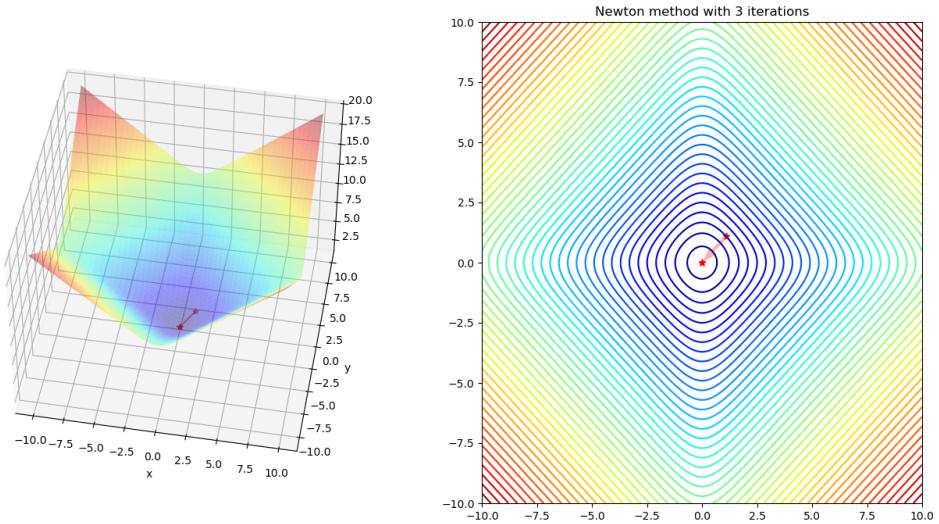
4.1 Περίπτωση για $n = 2$ μεταβλητές

1. $x_0 = [5, -9]^T$



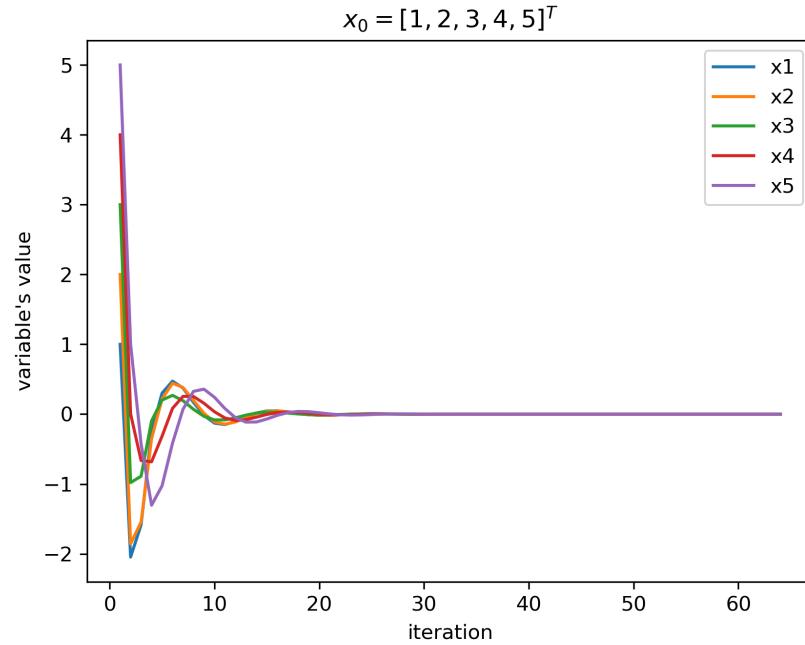
Σχήμα 25

2. $x_0 = [1.1, 1.1]^T$

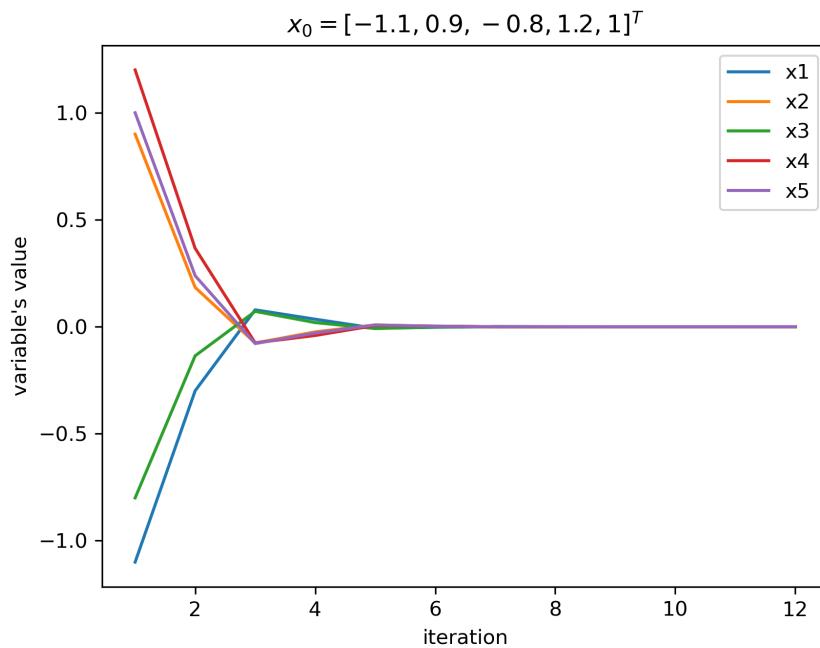


Σχήμα 26

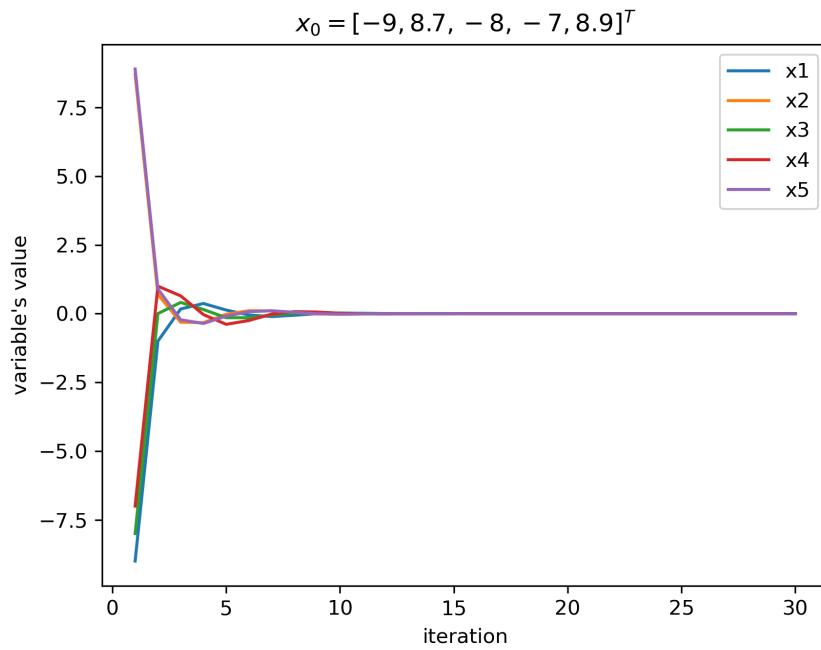
4.2 Περίπτωση για $n = 5$ μεταβλητές



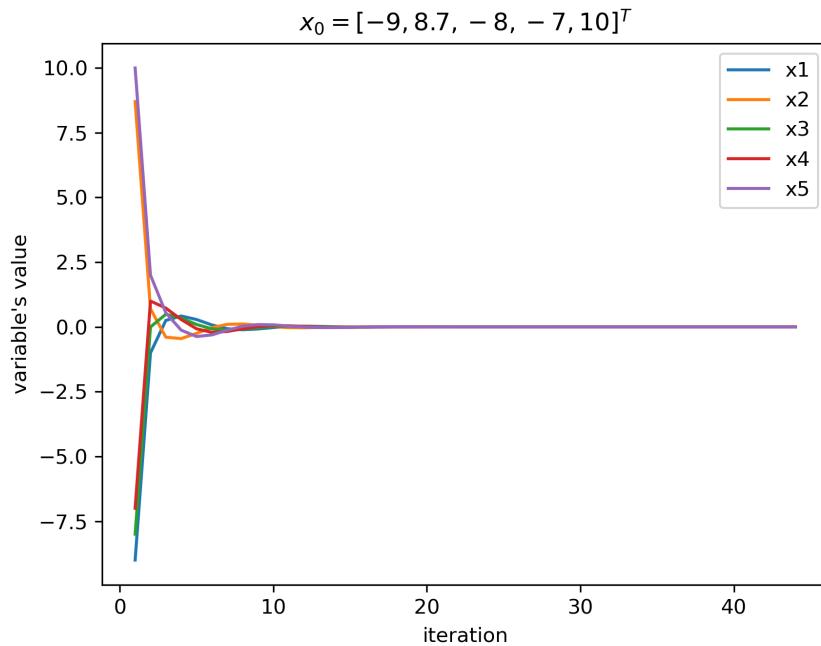
Σχήμα 27



Σχήμα 28



Σχήμα 29



Σχήμα 30

Παρατηρούμε ότι η συγκεκριμένη μέθοδος παρουσιάζει την πιο αργή σύγκλιση σε σχέση με τις άλλες δύο, σε κάθε μία από τις εξεταζόμενες περιπτώσεις.

5 Συμπεράσματα

Συνοψίζοντας, η μέθοδος steepest descent φαίνεται να είναι η καταλληλότερη για την ελαχιστοποίηση της συγκεκριμένης συνάρτησης. Δοκιμάζοντας τις τρεις μεθόδους, είδαμε ότι η μέθοδος Newton είναι αρκετά ασταθής λόγω του ιδιάζοντος εσσιανού πίνακα, ακόμα και στην βελτιωμένη της μορφή με την εφαρμογή του line search. Όσον αναφορά την μέθοδο conjugate gradient παρατηρούμε ότι έχει αρκετά αργή σύγκλιση συγκριτικά με τις άλλες δύο, ωστόσο φαίνεται να είναι ευσταθής. Από την άλλη πλευρά η steepest descent με τη χρήση του line search έχει την ταχύτερη σύγκλιση και ταυτόχρονα είναι η πιο ευσταθής και πιο αξιόπιστη, αφού εγγυάται σε κάθε περίπτωση την έρευνη της λύσης.

6 Βιβλιογραφία

1. George Nocedal, Stephen J. Wright, "*Numerical Optimization*", 2nd edition, Springer, 2006
2. Neculai Andrei, "*An Unconstrained Optimization Test Functions Collection*", 2008