Aristotle University of Thessaloniki
Department of Electrical and Computer Engineering

Minimization of Multivariable Functions Without Constraints Using Derivative-Based Optimization Methods

Eleni Daskalou, AEM: [redacted]

An assignment submitted as part of the coursework for the Optimization Techniques course at *Aristotle University of Thessaloniki*

November 24, 2024

Περιεχόμενα

List of Figures

List of Algorithms

Κεφάλαιο 1

Εισαγωγή

Η παρούσα εργασία επικεντρώνεται στην ελαχιστοποίηση συναρτήσεων πολλών μεταβλητών χωρίς περιορισμούς, χρησιμοποιώντας μεθόδους βελτιστοποίησης που αξιοποιούν παραγώγους. Εξετάζονται αλγόριθμοι που βασίζονται σε παραμετρικές προσεγγίσεις, με στόχο την αποτελεσματική εύρεση του ελάχιστου σημείου των συναρτήσεων μέσω της σταδιακής προσαρμογής του βήματος και της κατεύθυνσης καθόδου.

Οι μέθοδοι που αναλύονται περιλαμβάνουν τη Μέθοδο της Μέγιστης Καθόδου, τη Μέθοδο του Newton και τη Μέθοδο Levenberg-Marquardt. Αυτοί οι αλγόριθμοι βρίσκουν εφαρμογή σε πολλές περιοχές της επιστήμης και της μηχανικής, όπου η βελτιστοποίηση συναρτήσεων είναι κρίσιμη για την επίλυση σύνθετων προβλημάτων.

1.1 Περιγραφή του Προβλήματος και Μεθοδολογία

Το πρόβλημα που εξετάζεται στην εργασία αυτή αφορά την εύρεση του ελαχίστου σημείου μιας διαφορίσιμης συνάρτησης πολλών μεταβλητών, χωρίς περιορισμούς, χρησιμοποιώντας παραγώγους. Η μεθοδολογία που ακολουθείται βασίζεται στην εφαρμογή διαφορετικών αλγορίθμων βελτιστοποίησης, οι οποίοι υλοποιούνται στο λογισμικό *MATLAB*. Οι αλγόριθμοι συγκρίνονται ως προς την ταχύτητα σύγκλισης και την αποτελεσματικότητά τους στην ελαχιστοποίηση της συνάρτησης. Συγκεκριμένα, Ο σκοπός της αναφοράς αυτής είναι να παρουσιάσει τα αποτελέσματα και τα σχόλια πάνω στην μελέτη ελαχιστοποίησης της συνάρτησης $f(x, y) = x^5 e^{-x^2-y^2}$ χρησιμοποιώντας 3 μεθόδους: την μέθοδο steepest descend, την μέθοδο newton και την μέθοδο levenberg-marquardt.

Η διαδικασία περιλαμβάνει την ανάλυση της αποδοτικότητας κάθε μεθόδου σε σχέση με το πλήθος των επαναλήψεων που απαιτούνται για την επίτευξη σύγκλισης, καθώς και την αξιολόγηση των τελικών αποτελεσμάτων για κάθε μέθοδο.

1.2 Στόχοι της Εργασίας

Ο κύριος στόχος της εργασίας είναι η ανάλυση και η αξιολόγηση των μεθόδων βελτιστοποίησης μέσω παραγώγων για την ελαχιστοποίηση μιας διαφορίσιμης συνάρτησης πολλών μεταβλητών. Η μελέτη εστιάζει στην απόδοση των αλγορίθμων Μέγιστης Καθόδου, Newton και Levenberg-Marquardt, και συγκεκριμένα στην ταχύτητα και την ακρίβεια της σύγκλισης που επιτυγχάνουν.

Επιπλέον, διερευνάται η επίδραση των αρχικών τιμών και των παραμέτρων κάθε μεθόδου στη συνολική απόδοση της διαδικασίας βελτιστοποίησης. Μέσα από τη σύγκριση των μεθόδων, η εργασία αποσκοπεί στην εξαγωγή συμπερασμάτων σχετικά με τη βέλτιστη επιλογή αλγορίθμου ανάλογα με τις ιδιότητες της συνάρτησης και τις απαιτήσεις ακρίβειας του προβλήματος.

Κεφάλαιο 2

Υπόβαθρο

2.1 Μέθοδοι Αναζήτησης Ελαχίστου Με Χρήση Παραγώγων

2.1.1 Μέθοδος Μέγιστης Καθόδου (Steepest Descent)

Στόχος

Η ελαχιστοποίηση μιας διαφορίσιμης συνάρτησης $f : \mathbb{R}^n \to \mathbb{R}$, χωρίς περιορισμούς.

Περιγραφή της Μεθόδου

Η μέθοδος Μέγιστης Καθόδου ανήκει στις μεθόδους βελτιστοποίησης που χρησιμοποιούν παραγώγους για την εύρεση τοπικών ελαχίστων. Στόχος της είναι η εύρεση ενός σημείου x^* που ελαχιστοποιεί τη συνάρτηση $f(x)$ ξεκινώντας από ένα αρχικό σημείο και ακολουθώντας σταδιακά την κατεύθυνση της μεγαλύτερης καθόδου. Η κατεύθυνση αυτή καθορίζεται από την αρνητική κλίση της συνάρτησης στο τρέχον σημείο, καθώς η κλίση δείχνει την κατεύθυνση της μέγιστης αύξησης της συνάρτησης. Η διαδικασία συνεχίζεται μέχρι η κλίση να πλησιάσει το μηδέν ή να ικανοποιηθεί κάποιο άλλο κριτήριο σύγκλισης, επιτυγχάνοντας έτσι τη σύγκλιση σε ένα τοπικό ελάχιστο.

Βασικά Βήματα

Έστω ότι θέλουμε να ελαχιστοποιήσουμε μια διαφορίσιμη συνάρτηση $f : \mathbb{R}^n \to \mathbb{R}$, όπου $f(x)$ είναι συνεχώς διαφορίσιμη. Η μέθοδος ξεκινά από ένα αρχικό σημείο x_0 και ακολουθεί τα εξής βήματα.

Βήμα 1: Υπολογισμός της Κλίσης $\nabla f(x_k)$
Στο σημείο x_k, υπολογίζουμε την κλίση της συνάρτησης:

$$\nabla f(x_k) = \left(\frac{\partial f}{\partial x_1}(x_k), \frac{\partial f}{\partial x_2}(x_k), \ldots, \frac{\partial f}{\partial x_n}(x_k) \right)^T$$

Η κλίση $\nabla f(x_k)$ δείχνει την κατεύθυνση της μέγιστης αύξησης της συνάρτησης στο σημείο x_k. Η κατεύθυνση καθόδου ορίζεται ως η αντίθετη της κλίσης:

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

Βήμα 2: Επιλογή του Βήματος γ_k

Το βήμα γ_k είναι ένας παράγοντας κλιμάκωσης που καθορίζει την απόσταση που θα κινηθούμε κατά μήκος της κατεύθυνσης d_k. Η επιλογή του γ_k μπορεί να γίνει με τους εξής τρόπους:

1. *Σταθερό Βήμα*: Επιλέγουμε μια σταθερή τιμή για γ_k (π.χ. $\gamma_k = \gamma$) σε κάθε επανάληψη. Η τιμή αυτή είναι συνήθως μικρή για να διασφαλίζεται η σταθερότητα της μεθόδου.

2. *Ελαχιστοποίηση κατά μήκος της κατεύθυνσης καθόδου*:

$$\gamma_k = \arg\min_{\gamma>0} f(x_k + \gamma d_k)$$

Στην περίπτωση αυτή, υπολογίζουμε την τιμή του γ_k που ελαχιστοποιεί τη συνάρτηση κατά μήκος της κατεύθυνσης καθόδου. Αυτό απαιτεί επίλυση ενός προβλήματος ελαχιστοποίησης μίας μεταβλητής.

3. *Κανόνας Armijo*: Επιλέγουμε το γ_k έτσι ώστε να ικανοποιείται η ανισότητα:

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

όπου $\sigma \in (0, 1)$ είναι μια σταθερά (συνήθως κοντά στο 0.1). Αυτό εξασφαλίζει ότι η μείωση της συνάρτησης είναι επαρκής.

Βήμα 3: Υπολογισμός της Νέας Θέσης x_{k+1}

Με βάση την τιμή του γ_k που επιλέξαμε, η νέα θέση x_{k+1} υπολογίζεται ως εξής:

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

ή

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

Αυτό το βήμα μετακινεί το σημείο x_k κατά μήκος της κατεύθυνσης της μέγιστης καθόδου, δηλαδή της αρνητικής κλίσης.

Βήμα 4: Κριτήριο Τερματισμού

Η επαναληπτική διαδικασία συνεχίζεται μέχρι να ικανοποιηθεί κάποιο κριτήριο σύγκλισης, όπως:

1. *Μικρό μέγεθος της κλίσης*:

$$\|\nabla f(x_k)\| \leq \epsilon$$

όπου ϵ είναι μια προκαθορισμένη μικρή θετική σταθερά.

2. *Μικρή αλλαγή στη θέση*:

$$\|x_{k+1} - x_k\| \leq \delta$$

όπου δ είναι μια προκαθορισμένη μικρή θετική σταθερά.

3. *Μέγιστος αριθμός επαναλήψεων*: Η διαδικασία τερματίζεται αν φτάσει σε προκαθορισμένο αριθμό επαναλήψεων, εάν κανένα από τα παραπάνω κριτήρια δεν έχει ικανοποιηθεί.

Η μέθοδος της Μέγιστης Καθόδου βασίζεται στην ιδιότητα της κατεύθυνσης της κλίσης να οδηγεί στη μέγιστη μείωση της συνάρτησης για κάθε σημείο εκκίνησης. Συγκεκριμένα, η αρνητική κλίση της συνάρτησης $f(x)$ υποδεικνύει την κατεύθυνση κατά την οποία η τιμή της συνάρτησης μειώνεται με τον ταχύτερο δυνατό ρυθμό.

Έτσι, η μέθοδος εφαρμόζει επαναληπτικά την κίνηση προς την κατεύθυνση της μέγιστης καθόδου με ένα κατάλληλο βήμα γ_k, που είτε είναι σταθερό είτε επιλέγεται με βάση κριτήρια όπως η ελαχιστοποίηση κατά μήκος της κατεύθυνσης ή ο κανόνας Armijo. Σε κάθε επανάληψη, η μέθοδος συγκλίνει προς το σημείο ελαχίστου x^*, όπου η κλίση προσεγγίζει το μηδέν, επιτυγχάνοντας έτσι ένα τοπικό ελάχιστο της συνάρτησης.

Αλγόριθμος

Steepest Descent Method for Minimization

Input: Objective function $f(x)$, initial point x_0, tolerance ϵ, maximum iterations N
Output: Approximate minimizer x^*
Initialize $x = x_0$
for $k = 1$ to N **do**
    Compute gradient $g = \nabla f(x)$
    if $\|g\| \leq \epsilon$ **then**
        return x
    end if
    Choose step size γ_k (e.g., fixed, line search, or Armijo rule)
    Update $x = x - \gamma_k \cdot g$
end for
return x

Συμπεράσματα

Η μέθοδος Μέγιστης Καθόδου είναι μια απλή και αποτελεσματική τεχνική για την εύρεση τοπικών ελαχίστων σε διαφορίσιμες συναρτήσεις. Βασίζεται στην κατεύθυνση της αρνητικής κλίσης, που οδηγεί στη μέγιστη μείωση της συνάρτησης σε κάθε επανάληψη. Ωστόσο, η μέθοδος μπορεί να είναι αργή στη σύγκλιση όταν η συνάρτηση είναι έντονα κυρτή ή έχει μεγάλες διαφορές στην κλίση στις διάφορες κατευθύνσεις.

2.1.2 Μέθοδος Newton

Στόχος

Η μέθοδος του Newton χρησιμοποιείται για την εύρεση του τοπικού ελαχίστου μιας δύο φορές διαφορίσιμης συνάρτησης $f(x)$. Η μέθοδος βασίζεται στην προσέγγιση της συνάρτησης $f(x)$ με ένα παραβολικό πολυώνυμο δεύτερου βαθμού γύρω από το τρέχον σημείο x_k, χρησιμοποιώντας τη δεύτερη παράγωγο (Hessian) για να εκτιμήσει την καμπυλότητα της συνάρτησης. Η μέθοδος έχει ως στόχο την ταχύτερη σύγκλιση στο τοπικό ελάχιστο σε σχέση με άλλες μεθόδους καθόδου, όπως η μέθοδος της μέγιστης καθόδου, ειδικά όταν η συνάρτηση είναι καλά προσεγγισμένη από ένα τετραγωνικό πολυώνυμο.

Θεωρητική Ανάλυση της Μεθόδου Newton

Η μέθοδος του Newton ελαχιστοποιεί μια δύο φορές διαφορίσιμη συνάρτηση $f(x)$ χρησιμοποιώντας πληροφορία για την κλίση και την καμπυλότητα της συνάρτησης στο τρέχον σημείο x_k. Σε κάθε επανάληψη, η μέθοδος υπολογίζει το επόμενο σημείο x_{k+1} χρησιμοποιώντας τόσο την πρώτη παράγωγο όσο και τη δεύτερη παράγωγο (Hessian) της συνάρτησης, με στόχο την ταχύτερη σύγκλιση στο τοπικό ελάχιστο. Η διαδικασία συνεχίζεται μέχρι η αλλαγή στη θέση ή η τιμή της κλίσης να γίνει αρκετά μικρή, υποδεικνύοντας ότι έχει βρεθεί το σημείο ελαχίστου x^*.

Βασικά Βήματα

Η μέθοδος ξεκινά από ένα αρχικό σημείο x_0 και σε κάθε επανάληψη προσδιορίζει το επόμενο σημείο x_{k+1} λύνοντας το πρόβλημα βελτιστοποίησης μιας τετραγωνικής προσέγγισης της $f(x)$ γύρω από το x_k. Η κατεύθυνση καθόδου προκύπτει από την επίλυση της εξίσωσης:

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

όπου $\nabla f(x_k)$ είναι η κλίση της συνάρτησης στο σημείο x_k και $\nabla^2 f(x_k)$ είναι το Hessian, δηλαδή η δεύτερη παράγωγος της συνάρτησης, η οποία καθορίζει την καμπυλότητα της $f(x)$ στο σημείο αυτό.

1. Υπολογισμός της κλίσης και του Hessian: Στο σημείο x_k, υπολογίζουμε την κλίση $\nabla f(x_k)$ και τον πίνακα Hessian $\nabla^2 f(x_k)$.

2. Υπολογισμός της κατεύθυνσης καθόδου: Η κατεύθυνση καθόδου d_k προσδιορίζεται από τη σχέση:
$$d_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

3. Υπολογισμός του νέου σημείου: Το νέο σημείο x_{k+1} υπολογίζεται από τη σχέση:
$$x_{k+1} = x_k + d_k = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

4. Έλεγχος σύγκλισης: Ελέγχουμε αν η κλίση $\|\nabla f(x_{k+1})\|$ είναι μικρότερη από ένα προκαθορισμένο όριο ϵ. Εάν ισχύει, τερματίζουμε τη διαδικασία, διαφορετικά επαναλαμβάνουμε τα βήματα.

Σύγκλιση

Η μέθοδος του Newton παρουσιάζει ταχεία σύγκλιση κοντά στο σημείο ελαχίστου όταν η συνάρτηση $f(x)$ είναι επαρκώς ομαλή και ο πίνακας Hessian $\nabla^2 f(x)$ είναι θετικά ορισμένος σε αυτή την περιοχή. Λόγω της χρήσης του Hessian, η μέθοδος Newton συγκλίνει πιο γρήγορα από άλλες μεθόδους όπως η μέθοδος της μέγιστης καθόδου, ειδικά όταν η συνάρτηση μπορεί να προσεγγιστεί καλά από ένα τετραγωνικό πολυώνυμο. Ωστόσο, αν ο Hessian δεν είναι θετικά ορισμένος, μπορεί να χρειαστεί τροποποίηση της μεθόδου για να εξασφαλιστεί η σύγκλιση σε ελάχιστο.

Αλγόριθμος

Newton's Method for Minimization

Input: Objective function $f(x)$, initial point x_0, tolerance ϵ, maximum iterations N
Output: Approximate minimizer x^*
Initialize $x = x_0$
for $k = 1$ to N **do**
    Compute gradient $g = \nabla f(x)$ and Hessian $H = \nabla^2 f(x)$
    if $\|g\| \leq \epsilon$ **then**
        return x
    end if
    Compute direction $d = -H^{-1} g$
    Update $x = x + d$
end for
return x

Συμπεράσματα

Η μέθοδος του Newton είναι μια ισχυρή και γρήγορη μέθοδος βελτιστοποίησης, ειδικά όταν η συνάρτηση $f(x)$ μπορεί να προσεγγιστεί καλά από ένα τετραγωνικό πολυώνυμο κοντά στο σημείο ελαχίστου. Χάρη στη χρήση του πίνακα Hessian, η μέθοδος συγκλίνει συνήθως γρήγορα κοντά στο ελάχιστο, παρουσιάζοντας τετραγωνική σύγκλιση όταν βρισκόμαστε σε γειτονιά του σημείου ελαχίστου. Ωστόσο, απαιτεί τον υπολογισμό του Hessian, γεγονός που μπορεί να είναι υπολογιστικά ακριβό για υψηλών διαστάσεων προβλήματα. Επιπλέον, η μέθοδος δεν εγγυάται σύγκλιση σε ελάχιστο αν ο Hessian δεν είναι θετικά ορισμένος, και σε αυτές τις περιπτώσεις μπορεί να χρειαστεί τροποποίηση της μεθόδου (π.χ., μέθοδος Levenberg-Marquardt) για να διασφαλιστεί η σύγκλιση.

2.1.3 Μέθοδος Levenberg-Marquardt

Στόχος

Η μέθοδος Levenberg-Marquardt χρησιμοποιείται για την εύρεση του τοπικού ελαχίστου μιας συνάρτησης, συνδυάζοντας στοιχεία από τη μέθοδο Newton και τη μέθοδο της μέγιστης καθόδου. Είναι ιδιαίτερα χρήσιμη για προβλήματα μη γραμμικής βελτιστοποίησης και ειδικά για την προσαρμογή παραμέτρων σε μη γραμμικά μοντέλα. Ο στόχος της είναι να εξασφαλίσει τη σύγκλιση ακόμα και όταν ο πίνακας Hessian δεν είναι θετικά ορισμένος, ρυθμίζοντας την ταχύτητα και τη σταθερότητα της μεθόδου σε δύσκολα προβλήματα βελτιστοποίησης.

Περιγραφή της Μεθόδου

Η μέθοδος Levenberg-Marquardt συνδυάζει τη μέθοδο Newton, που βασίζεται στην κλίση και τον πίνακα Hessian, με μια ρυθμιστική παράμετρο για να αντιμετωπίσει περιπτώσεις όπου ο Hessian δεν είναι θετικά ορισμένος. Σε κάθε επανάληψη, η μέθοδος υπολογίζει μια τροποποιημένη εκδοχή του Hessian προσθέτοντας μια σταθερά μ στον διαγώνιο, η οποία βοηθά στη διασφάλιση θετικής οριστικότητας. Όταν η παράμετρος μ είναι μεγάλη, η μέθοδος προσεγγίζει τη μέθοδο της μέγιστης καθόδου, ενώ για μικρές τιμές της μ, η μέθοδος προσεγγίζει τη μέθοδο Newton, εξισορροπώντας έτσι ταχύτητα και σταθερότητα.

Βασικά Βήματα

1. Ορισμός του αρχικού σημείου x_0, της ακρίβειας ϵ, και της αρχικής τιμής της παραμέτρου μ.

2. Υπολογισμός της κλίσης και του Hessian: Στο σημείο x_k, υπολογίζουμε την κλίση $\nabla f(x_k)$ και τον πίνακα Hessian $\nabla^2 f(x_k)$.

3. Προσθήκη παραμέτρου στη διαγώνιο του Hessian:

$$H_k = \nabla^2 f(x_k) + \mu I$$

   όπου μ είναι μια θετική παράμετρος και I ο μοναδιαίος πίνακας. Αυτή η προσθήκη διασφαλίζει ότι ο πίνακας H_k είναι θετικά ορισμένος.

4. Υπολογισμός της κατεύθυνσης καθόδου:

$$d_k = -H_k^{-1} \nabla f(x_k)$$

5. **Ενημέρωση του σημείου**:

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

6. Έλεγχος σύγκλισης: Εάν $\|\nabla f(x_{k+1})\| \leq \epsilon$, τερματίζουμε τη διαδικασία. Διαφορετικά, προσαρμόζουμε την παράμετρο μ ανάλογα: - Μειώνουμε μ αν η ενημέρωση x_{k+1} μειώνει επιτυχώς τη συνάρτηση $f(x)$. - Αυξάνουμε μ αν η ενημέρωση x_{k+1} δεν μειώνει τη συνάρτηση $f(x)$, ώστε να ενισχύσουμε τη σταθερότητα.

7. Επανάληψη της διαδικασίας: Θέτουμε $x_k = x_{k+1}$ και επαναλαμβάνουμε από το Βήμα 2 έως ότου επιτευχθεί η σύγκλιση.

Αλγόριθμος

Levenberg-Marquardt Method for Minimization

Input: Objective function $f(x)$, initial point x_0, tolerance ϵ, initial μ, maximum iterations N
Output: Approximate minimizer x^*
Initialize $x = x_0$
for $k = 1$ to N **do**
  Compute gradient $g = \nabla f(x)$ and Hessian $H = \nabla^2 f(x)$
  if $\|g\| \leq \epsilon$ **then**
    return x
  end if
  Modify Hessian: $H_k = H + \mu I$
  Compute direction $d = -H_k^{-1} g$
  Update $x_{\text{new}} = x + d$
  if $f(x_{\text{new}}) < f(x)$ **then**
    $x = x_{\text{new}}$
    Reduce μ
  else
    Increase μ
  end if
end for
return x

Συμπεράσματα

Η μέθοδος Levenberg-Marquardt είναι ιδανική για την επίλυση προβλημάτων μη γραμμικής βελτιστοποίησης, συνδυάζοντας τη σταθερότητα της μεθόδου της μέγιστης καθόδου με την ταχύτητα της μεθόδου Newton. Η προσαρμοστικότητα της παραμέτρου μ επιτρέπει τη μετάβαση από μια πιο σταθερή σε μια πιο ταχεία μέθοδο, ανάλογα με τις ανάγκες κάθε επανάληψης. Ωστόσο, η επιλογή της αρχικής τιμής της παραμέτρου μ και ο τρόπος προσαρμογής της επηρεάζουν την απόδοση της μεθόδου και μπορεί να χρειαστεί κάποια πειραματική ρύθμιση. Συνολικά, η μέθοδος είναι πολύ αποτελεσματική για βελτιστοποίηση παραμέτρων σε μη γραμμικά μοντέλα και παρουσιάζει ταχεία σύγκλιση σε πολλές πρακτικές εφαρμογές.

Κεφάλαιο 3

Εφαρμογή Αλγορίθμων Βελτιστοποίησης

3.1 Θέμα 1

Εκφώνηση

Η αντικειμενική συνάρτηση που θα μελετήσουμε είναι η:

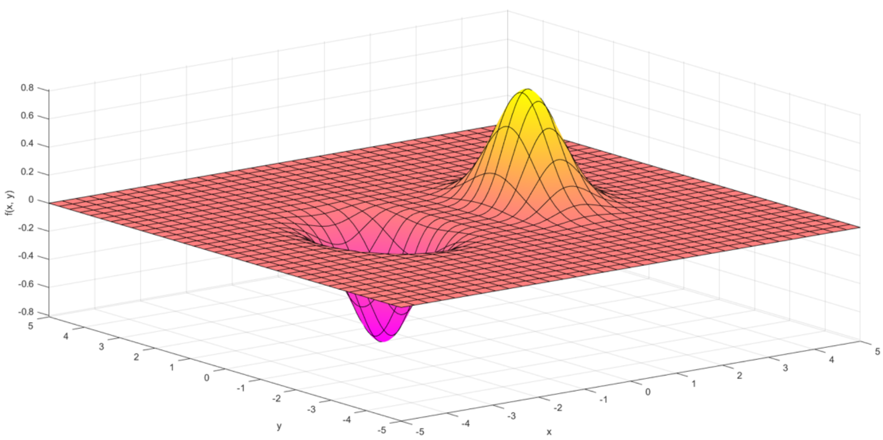
$$f(x,y) = x^5 e^{-x^2-y^2}.$$

Σχεδιάστε την f για να πάρετε μια γενική εικόνα της μορφής της.

3.1.1 Προσομοίωση Σε Matlab

Με το αρχείο *MATLAB*, "*f_function.m*", δημιουργούμε το παρακάτω γράφημα της συνάρτησης f, για να έχουμε μία καλή οπτική εικόνα προς βαθύτερη κατανόηση.

Figure 3.1: Graph of $f(x,y) = x^5 e^{-x^2-y^2}$



Από την τρισδιάστατη απεικόνιση της συνάρτησης f, παρατηρούμε ότι το ελάχιστο της συνάρτησης βρίσκεται σε αρνητικές τιμές του x και κοντά στο μηδέν στον άξονα y. Αντίστοιχα, η μέγιστη τιμή της συνάρτησης εμφανίζεται σε θετικές τιμές του x και επίσης κοντά στο μηδέν στον άξονα y. Επιπλέον, εντοπίζουμε ένα τοπικό ελάχιστο στο σημείο $(0,0)$.

Για την έναρξη των αλγορίθμων από το σημείο $(0,0)$, παρατηρούμε ότι κάθε αλγόριθμος σταματά αμέσως. Αυτό είναι λογικό αν αναλογιστούμε την προϋπόθεση τερματισμού, δηλαδή όταν το μέτρο της κλίσης $|\nabla f|$ είναι μικρότερο από μια θετική σταθερά ϵ. Συγκεκριμένα, η κλίση της συνάρτησης f είναι:

$$\nabla f = \left[5x^4e^{-x^2-y^2} - 2x^6e^{-x^2-y^2}, -2x^5ye^{-x^2-y^2}\right]$$

και για το σημείο $(0,0)$, αυτή δίνει το μηδενικό διάνυσμα. Συνεπώς, το σημείο αυτό λειτουργεί ως σημείο "εγλωβισμού", καθώς πληροί την προϋπόθεση τερματισμού από την αρχή.

3.2 Θέμα 2

Εκφώνηση

Ελαχιστοποιήστε την $f(x,y) = x^5e^{-x^2-y^2}$ με τη μέθοδο **Μέγιστης Καθόδου**, χρησιμοποιώντας ως αρχικά σημεία (x_0, y_0) τα

i) $(0,0)$,

ii) $(-1,1)$, και

iii) $(1,-1)$.

Το βήμα γ_k θα επιλεγεί:

α) σταθερό (της επιλογής σας),

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

γ) βάσει του κανόνα Armijo.

Σχολιάστε τις διαφορές στα αποτελέσματα, σε περίπτωση που προκαλούνται, λόγω της επιλογής του σημείου έναρξης (x_0, y_0) του αλγορίθμου, καθώς επίσης και λόγω της επιλογής του βήματος γ_k. Οδηγούμαστε πάντα σε σωστό αποτέλεσμα; Αν όχι, τι πιστεύετε ότι φταίει;

3.2.1 Μαθηματική Ανάλυση

Για το ερώτημα (α) με σταθερό βήμα, θα εφαρμόσουμε τη μέθοδο Μέγιστης Καθόδου ξεχωριστά για κάθε ένα από τα τρία αρχικά σημεία: $(0,0)$, $(-1,1)$, και $(1,-1)$. Στην ανάλυση αυτή, το βήμα γ θα παραμείνει σταθερό κατά τη διάρκεια όλων των επαναλήψεων.

Η γενική μορφή της διαδικασίας με σταθερό βήμα είναι η εξής:

1. Υπολογισμός της κλίσης $\nabla f(x,y)$:

$$\nabla f(x,y) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]$$

όπου $f(x,y) = x^5e^{-x^2-y^2}$. Οι μερικές παράγωγοι είναι:

$$\frac{\partial f}{\partial x} = \left(5x^4 - 2x^6\right)e^{-x^2-y^2}$$

$$\frac{\partial f}{\partial y} = -2yx^5e^{-x^2-y^2}$$

2. Επιλογή κατεύθυνσης καθόδου: Η κατεύθυνση καθόδου d_k ορίζεται ως:

$$d_k = -\nabla f(x_k, y_k)$$

3. Υπολογισμός του επόμενου σημείου: Το επόμενο σημείο (x_{k+1}, y_{k+1}) ορίζεται από τη σχέση:

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \gamma d_k$$

όπου γ είναι σταθερό.

4. Συνθήκη σύγκλισης: Επαναλαμβάνουμε τα βήματα μέχρι η κλίση να γίνει μικρότερη από ένα προκαθορισμένο όριο, δηλαδή $\|\nabla f(x_k, y_k)\| < \epsilon$.

Ανάλυση για Κάθε Αρχικό Σημείο

i. Για αρχικό σημείο $(x_0, y_0) = (0, 0)$:

- Υπολογισμός της κλίσης: Στο σημείο $(0, 0)$, έχουμε:

$$\nabla f(0, 0) = [0, 0]$$

Η κλίση είναι μηδενική, επομένως ο αλγόριθμος τερματίζει αμέσως χωρίς να πραγματοποιήσει επιπλέον βήματα. Αυτό σημαίνει ότι το σημείο $(0, 0)$ είναι πιθανό να είναι σημείο τοπικού ελαχίστου ή σέλας.

ii. Για αρχικό σημείο $(x_0, y_0) = (-1, 1)$:

Έστω ότι επιλέγουμε:

- Σταθερό βήμα $\gamma = 0.1$
- Όριο σύγκλισης $\epsilon = 10^{-3}$

Με αυτή την επιλογή, θα επαναλαμβάνουμε τη διαδικασία μέχρι η κλίση $\|\nabla f(x_k, y_k)\|$ να είναι μικρότερη από ϵ.

Βήμα 0: Αρχικές συνθήκες

- Αρχικό σημείο: $(x_0, y_0) = (-1, 1)$
- Υπολογισμός της κλίσης στο (x_0, y_0):

$$\nabla f(-1, 1) = \left[3e^{-2}, -2e^{-2}\right] \approx [0.406, -0.271]$$

- Κατεύθυνση καθόδου:

$$d_0 = -\nabla f(-1, 1) = [-0.406, 0.271]$$

- Ενημέρωση του σημείου:

$$(x_1, y_1) = (x_0, y_0) + \gamma d_0 = (-1, 1) + 0.1 \times [-0.406, 0.271]$$

$$(x_1, y_1) = (-1 - 0.0406, 1 + 0.0271) = (-1.0406, 1.0271)$$

Βήμα 1: Υπολογισμός κλίσης στο (x_1, y_1)

Υπολογίζουμε την κλίση στο νέο σημείο:

$$\nabla f(-1.0406, 1.0271)$$

(Εδώ απαιτείται αριθμητικός υπολογισμός για τη συνέχεια.)

Η διαδικασία συνεχίζεται με τον ίδιο τρόπο, με κάθε νέο σημείο να υπολογίζεται και η κλίση να ελέγχεται αν είναι μικρότερη από το όριο ϵ.

iii. Για αρχικό σημείο $(x_0, y_0) = (1, -1)$:

Έστω ότι επιλέγουμε:

- Σταθερό βήμα $\gamma = 0.1$
- Όριο σύγκλισης $\epsilon = 10^{-3}$

Με αυτή την επιλογή, θα επαναλαμβάνουμε τη διαδικασία μέχρι η κλίση $\|\nabla f(x_k, y_k)\|$ να είναι μικρότερη από ϵ.

Βήμα 0: Αρχικές συνθήκες

- Αρχικό σημείο: $(x_0, y_0) = (1, -1)$
- Υπολογισμός της κλίσης στο (x_0, y_0):

$$\nabla f(1, -1) = \left[3e^{-2}, 2e^{-2}\right] \approx [0.406, 0.271]$$

- Κατεύθυνση καθόδου:

$$d_0 = -\nabla f(1, -1) = [-0.406, -0.271]$$

- Ενημέρωση του σημείου:

$$(x_1, y_1) = (x_0, y_0) + \gamma d_0 = (1, -1) + 0.1 \times [-0.406, -0.271]$$

$$(x_1, y_1) = (1 - 0.0406, -1 - 0.0271) = (0.9594, -1.0271)$$

Βήμα 1: Υπολογισμός κλίσης στο (x_1, y_1)

Υπολογίζουμε την κλίση στο νέο σημείο:

$$\nabla f(0.9594, -1.0271)$$

(Εδώ απαιτείται αριθμητικός υπολογισμός για τη συνέχεια.)

Η διαδικασία συνεχίζεται με τον ίδιο τρόπο, με κάθε νέο σημείο να υπολογίζεται και η κλίση να ελέγχεται αν είναι μικρότερη από το όριο ϵ.

Για το ερώτημα (β), όπου το βήμα γ επιλέγεται έτσι ώστε να ελαχιστοποιεί τη συνάρτηση $f(x + \gamma d)$, θα εφαρμόσουμε τη μέθοδο Μέγιστης Καθόδου ξεχωριστά για κάθε ένα από τα τρία αρχικά σημεία: $(0, 0)$, $(-1, 1)$, και $(1, -1)$.

Η γενική μορφή της διαδικασίας είναι η εξής:

- Υπολογισμός της Κλίσης $\nabla f(x_k, y_k)$ στο σημείο (x_k, y_k).

- Υπολογισμός της Κατεύθυνσης Καθόδου $d_k = -\nabla f(x_k, y_k)$.
- Εύρεση του Βέλτιστου Βήματος γ: Για το βήμα αυτό, λύνουμε το πρόβλημα ελαχιστοποίησης μιας μεταβλητής:

$$\min_{\gamma} f(x_k + \gamma d_{k,x}, y_k + \gamma d_{k,y})$$

Αυτό σημαίνει ότι πρέπει να βρούμε την τιμή του γ που ελαχιστοποιεί τη συνάρτηση κατά μήκος της κατεύθυνσης d_k.

Ανάλυση για Κάθε Αρχικό Σημείο

i. Για αρχικό σημείο $(x_0, y_0) = (0, 0)$:

- Υπολογισμός της Κλίσης: Στο σημείο $(0, 0)$, έχουμε:

$$\nabla f(0, 0) = [0, 0]$$

- Κατεύθυνση Καθόδου: Η κατεύθυνση καθόδου είναι $d_0 = -\nabla f(0, 0) = [0, 0]$.
- Βέλτιστο Βήμα γ: Δεδομένου ότι η κλίση είναι μηδενική, δεν χρειάζεται να υπολογίσουμε κάποιο βήμα. Το σημείο $(0, 0)$ είναι ήδη ένα σημείο τοπικού ελαχίστου ή σέλας.

i. Για αρχικό σημείο $(x_0, y_0) = (-1, 1)$:

1. Υπολογισμός της Κλίσης:

$$\nabla f(-1, 1) = \left[3e^{-2}, -2e^{-2}\right] \approx [0.406, -0.271]$$

2. Κατεύθυνση Καθόδου:

$$d_0 = -\nabla f(-1, 1) = [-0.406, 0.271]$$

3. Εύρεση του Βέλτιστου Βήματος γ: Θέλουμε να ελαχιστοποιήσουμε τη συνάρτηση $f(x, y) = x^5 e^{-x^2-y^2}$ κατά μήκος της κατεύθυνσης d_0.

   Αυτό σημαίνει ότι πρέπει να βρούμε το γ που ελαχιστοποιεί:

$$f(x_0 + \gamma d_{0,x}, y_0 + \gamma d_{0,y}) = f(-1 + \gamma \cdot (-0.406), 1 + \gamma \cdot 0.271)$$

4. Υπολογισμός της Παραγώγου ως προς γ: Για να βρούμε το ελάχιστο, υπολογίζουμε την παράγωγο της f ως προς γ και τη θέτουμε ίση με το μηδέν:

$$\frac{d}{d\gamma} f(-1 + \gamma \cdot (-0.406), 1 + \gamma \cdot 0.271) = 0$$

   (Η συγκεκριμένη παράγωγος απαιτεί αριθμητικό υπολογισμό σε ένα κατάλληλο περιβάλλον, όπως το $MATLAB$, για τη συνέχεια.)

i. Για αρχικό σημείο $(x_0, y_0) = (1, -1)$:

1. Υπολογισμός της Κλίσης:

$$\nabla f(1, -1) = \left[3e^{-2}, 2e^{-2}\right] \approx [0.406, 0.271]$$

2. Κατεύθυνση Καθόδου:

$$d_0 = -\nabla f(1, -1) = [-0.406, -0.271]$$

3. Εύρεση του Βέλτιστου Βήματος γ: Θέλουμε να ελαχιστοποιήσουμε τη συνάρτηση $f(x, y) = x^5 e^{-x^2-y^2}$ κατά μήκος της κατεύθυνσης d_0.

   Αυτό σημαίνει ότι πρέπει να βρούμε το γ που ελαχιστοποιεί:

$$f(x_0 + \gamma d_{0,x}, y_0 + \gamma d_{0,y}) = f(1 + \gamma \cdot (-0.406), -1 + \gamma \cdot (-0.271))$$

4. Υπολογισμός της Παραγώγου ως προς γ: Για να βρούμε το ελάχιστο, υπολογίζουμε την παράγωγο της f ως προς γ και τη θέτουμε ίση με το μηδέν:

$$\frac{d}{d\gamma} f(1 + \gamma \cdot (-0.406), -1 + \gamma \cdot (-0.271)) = 0$$

   (Αυτός ο υπολογισμός επίσης απαιτεί αριθμητική επίλυση, για αυτό θα παρουσιαστούν προσομοιώσεις στο *MATALAB*, στη συνέχεια.)

Για το ερώτημα (γ) βήμα που επιλέγεται βάσει του κανόνα *Armijo* θα εφαρμόσουμε τη μέθοδο Μέγιστης Καθόδου ξεχωριστά για κάθε ένα από τα τρία αρχικά σημεία: $(0, 0)$, $(-1, 1)$, και $(1, -1)$.

Η γενική μορφή της διαδικασίας είναι η εξής:

Κανόνας του *Armijo*

Ο κανόνας του *Armijo* χρησιμοποιείται για να εξασφαλιστεί ότι το βήμα γ που επιλέγουμε οδηγεί σε επαρκή μείωση της τιμής της συνάρτησης. Η επιλογή του γ βασίζεται στην ανισότητα:

$$f(x_k + \gamma d_k) \leq f(x_k) + \sigma\gamma \nabla f(x_k)^T d_k$$

όπου:

- $f(x_k + \gamma d_k)$ είναι η τιμή της συνάρτησης στο νέο σημείο,
- $f(x_k)$ είναι η τιμή της συνάρτησης στο τρέχον σημείο,
- $\sigma \in (0, 1)$ είναι μια μικρή θετική σταθερά, συνήθως $\sigma = 0.1$ ή $\sigma = 0.01$,
- $\nabla f(x_k)^T d_k$ είναι το εσωτερικό γινόμενο της κλίσης στο τρέχον σημείο με την κατεύθυνση καθόδου.

Ανάλυση για Κάθε Αρχικό Σημείο

Ας προχωρήσουμε στην πλήρη ανάλυση της μεθόδου για κάθε σημείο $(0, 0)$, $(-1, 1)$, $(1, -1)$, εφαρμόζοντας τον κανόνα του *Armijo* για τον υπολογισμό του βήματος γ.

i. Για αρχικό σημείο $(x_0, y_0) = (0, 0)$:

- Υπολογισμός της κλίσης: Στο σημείο $(0, 0)$, η κλίση είναι:

$$\nabla f(0, 0) = [0, 0]$$

- Κατεύθυνση καθόδου: Η κατεύθυνση καθόδου είναι $d_0 = -\nabla f(0, 0) = [0, 0]$.
- Συμπέρασμα: Η κλίση είναι μηδενική, επομένως ο αλγόριθμος τερματίζει αμέσως χωρίς να πραγματοποιήσει επιπλέον βήματα. Το σημείο $(0, 0)$ είναι πιθανό να είναι σημείο τοπικού ελαχίστου ή σέλας.

i. Για αρχικό σημείο $(x_0, y_0) = (-1, 1)$:

1. Υπολογισμός της κλίσης: Στο σημείο $(-1, 1)$, η κλίση είναι:

$$\nabla f(-1, 1) = \left[3e^{-2}, -2e^{-2}\right] \approx [0.406, -0.271]$$

2. Κατεύθυνση καθόδου: Η κατεύθυνση καθόδου είναι:

$$d_0 = -\nabla f(-1, 1) = [-0.406, 0.271]$$

3. Επιλογή βήματος βάσει του κανόνα *Armijo*:
   - Αρχικοποιούμε το γ με μια αρχική τιμή, π.χ., $\gamma = 1$.
   - Ελέγχουμε την ανισότητα του Armijo:

$$f(x_0 + \gamma d_0, y_0 + \gamma d_0) \leq f(x_0, y_0) + \sigma\gamma\nabla f(x_0, y_0)^T d_0$$

   - Εάν η ανισότητα δεν ισχύει, μειώνουμε το γ (π.χ., πολλαπλασιάζοντας το με έναν συντελεστή $\beta \in (0, 1)$) μέχρι να ικανοποιηθεί η συνθήκη.

i. Για αρχικό σημείο $(x_0, y_0) = (1, -1)$:

1. Υπολογισμός της κλίσης: Στο σημείο $(1, -1)$, η κλίση είναι:

$$\nabla f(1, -1) = \left[3e^{-2}, 2e^{-2}\right] \approx [0.406, 0.271]$$

2. Κατεύθυνση καθόδου: Η κατεύθυνση καθόδου είναι:

$$d_0 = -\nabla f(1, -1) = [-0.406, -0.271]$$

3. Επιλογή βήματος βάσει του κανόνα *Armijo*:
   - Αρχικοποιούμε το γ με μια αρχική τιμή, π.χ., $\gamma = 1$.
   - Ελέγχουμε την ανισότητα του Armijo:

$$f(x_0 + \gamma d_0, y_0 + \gamma d_0) \leq f(x_0, y_0) + \sigma\gamma\nabla f(x_0, y_0)^T d_0$$

   - Εάν η ανισότητα δεν ισχύει, μειώνουμε το γ (π.χ., πολλαπλασιάζοντας το με έναν συντελεστή $\beta \in (0, 1)$) μέχρι να ικανοποιηθεί η συνθήκη.

3.2.2 Προσομοίωση Σε Matlab

Παρουσιάζονται τα αποτελέσματα της προσομοίωσης του Θέματος 2, στο περιβάλλον *Matlab*, για κάθε ένα από τα τρία σημεία και κάθε τρόπο επιλογής βήματος.

Για την μελέτη αυτή, έχει υλοποιηθεί ο αλγόριθμος της Μεθόδου Μέγιστης Καθόδου (*SteepestDescent*), όπως παρουσιάζεται θεωρητικά στο Κεφ. 2, ο κώδικας του οποίου περιέχεται στο αρχείο steepest_descent.m.

Στα γραφήματα παρουσιάζονται τόσο τα αποτελέσματα της μεθόδου, όσο και οι τιμές της αντικειμενικής συνάρτησης για έναν αριθμό επαναλήψεων.

Προφανώς, όπως σημειώθηκε και αποδείχθηκε στο κεφάλαιο 2, δε θα παρουσιαστεί κάποιο γράφημα για τις τιμές της αντικειμενικής συνάρτησης για αρχικό σημείο το $(0, 0)$, αφού κάθε αλγόριθμος τερματίζει ακαριαία.

Αρχικά, εξετάζεται το πλήθος των επαναλήψεων που χρειάζονται για να συγκλίνει η αντικειμενική συνάρτηση ξεκινώντας από κάθε ένα αρχικό σημείο, χρησιμοποιώντας τη μέθοδο μέγιστης καθόδου, επιλέγοντας σταθερό βήμα ίσο με 0.005.

Figure 3.2: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$. , step = 0.005, start $(0, 0)$

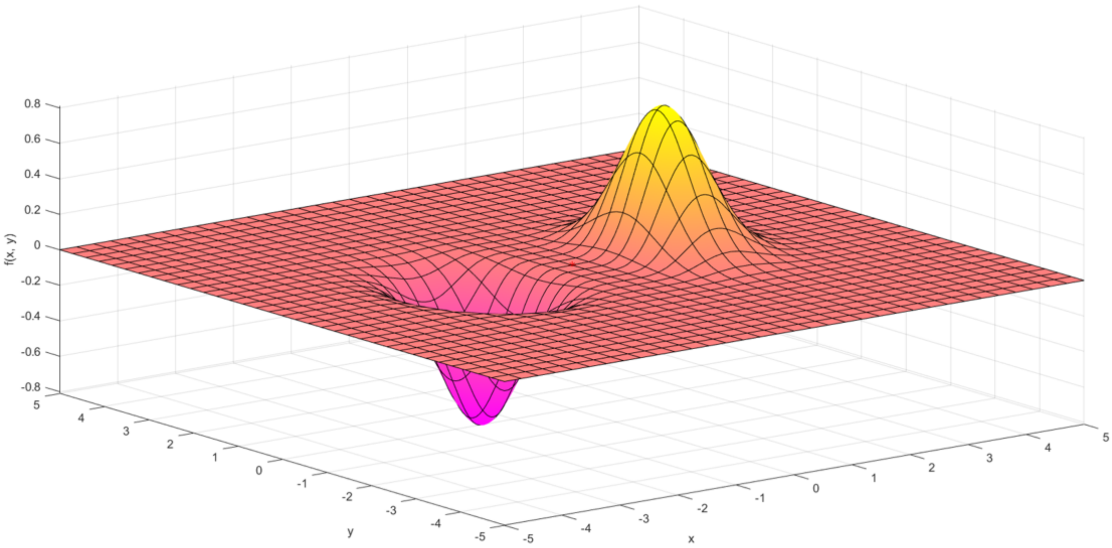


Figure 3.3: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$. , step = 0.005, start $(-1, 1)$

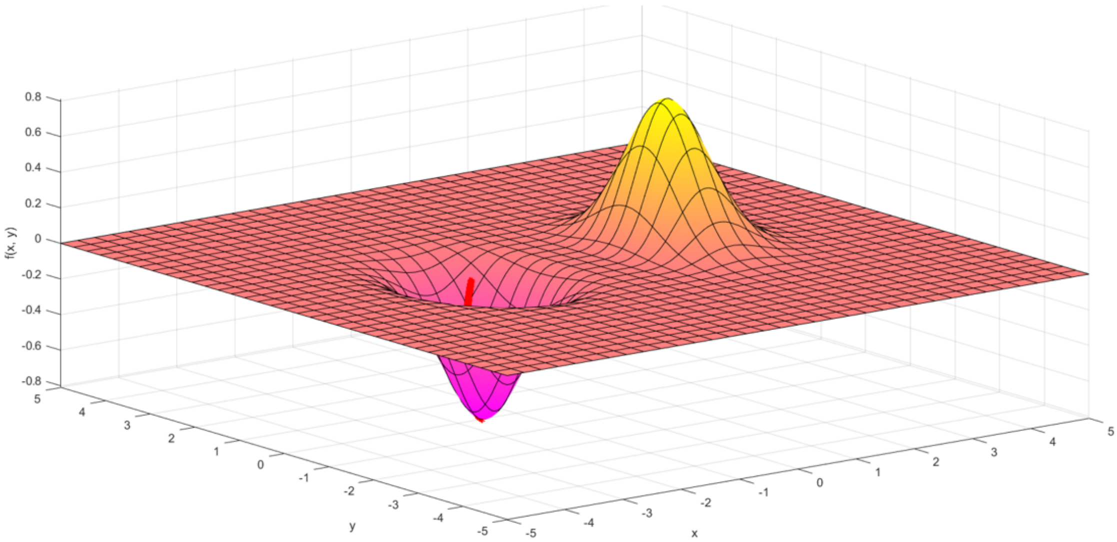


Figure 3.4: of function $f(x, y) = x^5 e^{-x^2-y^2}$. , step = 0.005, start $(-1, 1)$

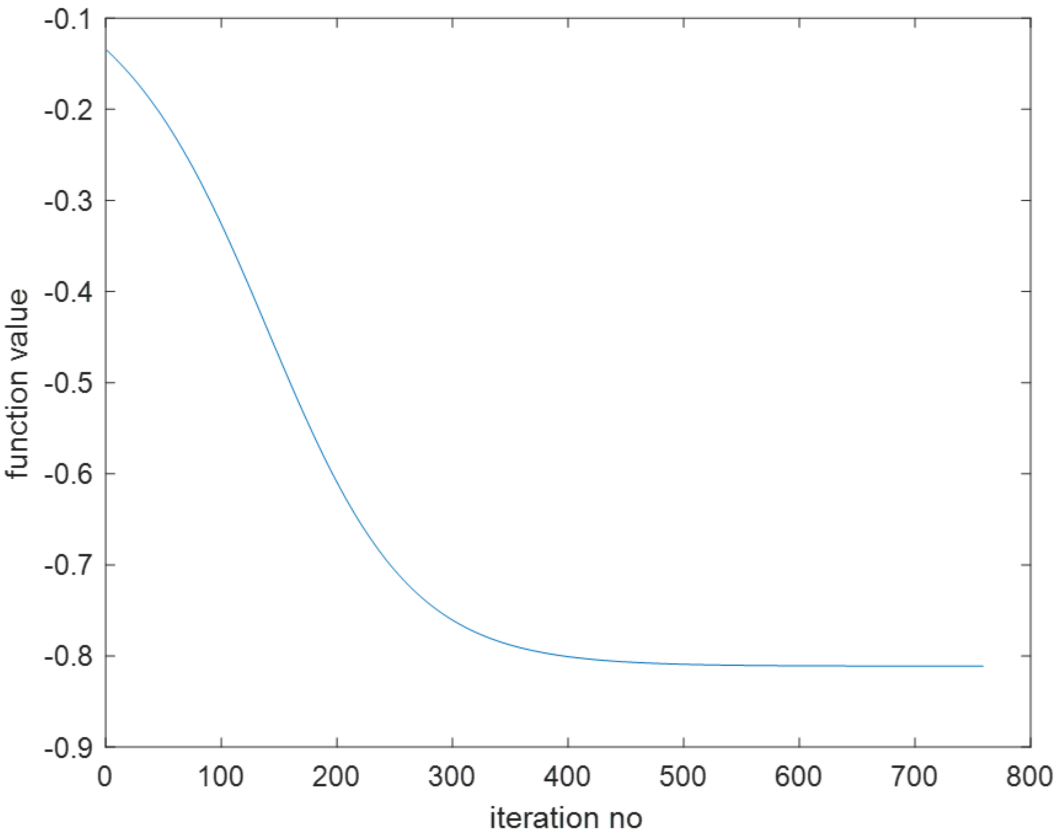


Figure 3.5: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$. , step $= 0.005$, start $(1, -1)$

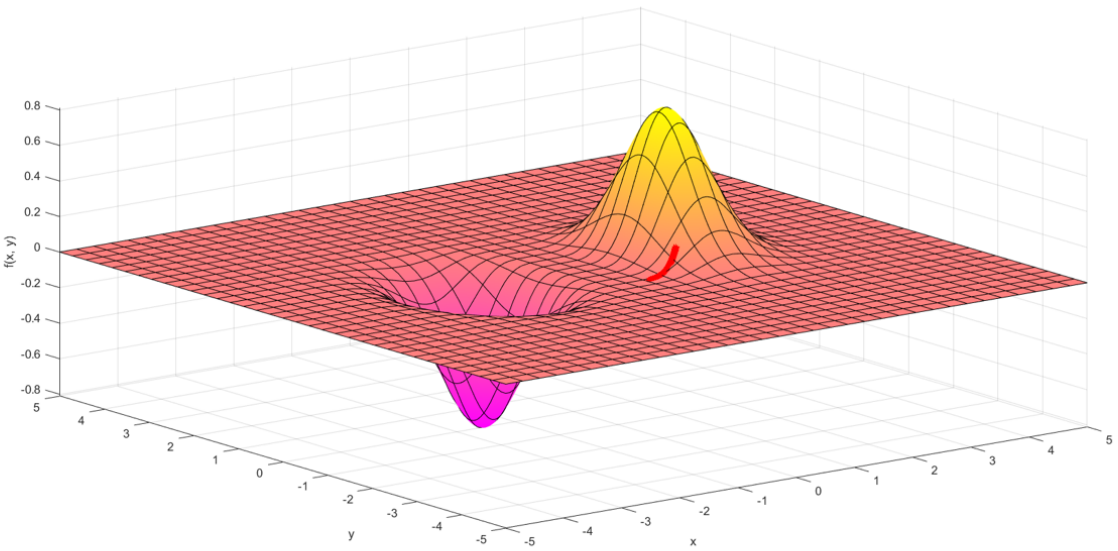
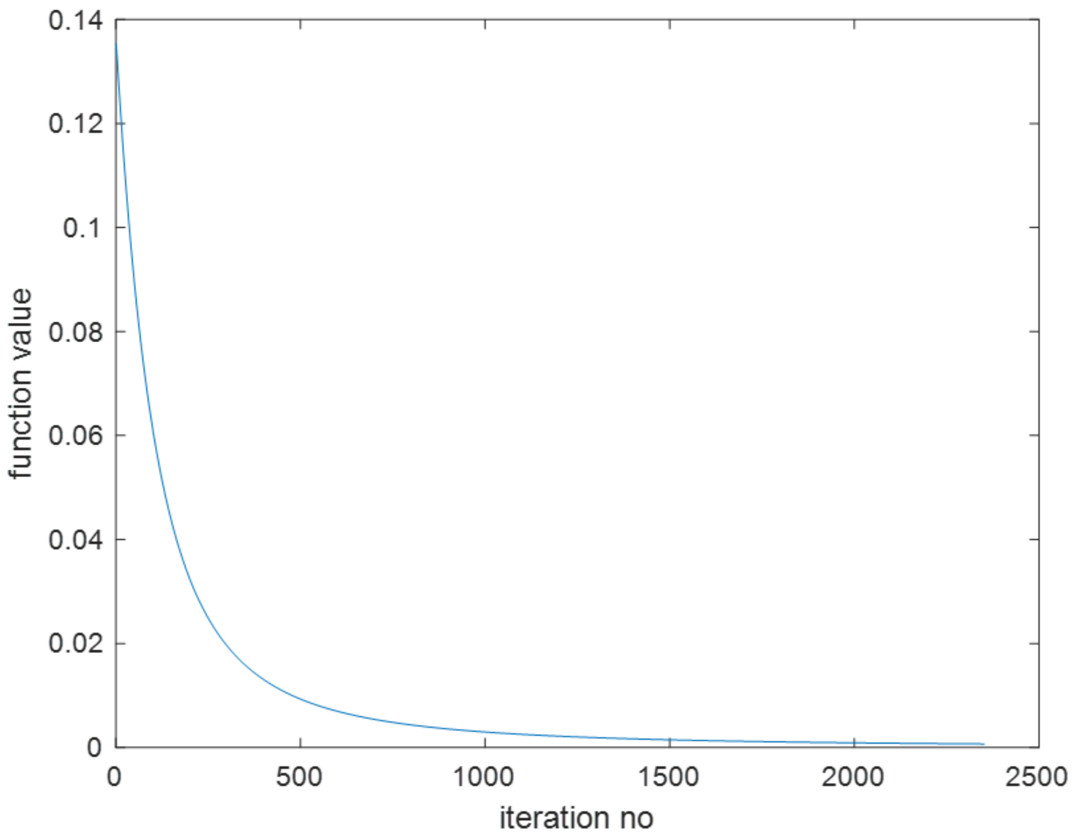


Figure 3.6: of function $f(x, y) = x^5 e^{-x^2-y^2}$. , step $= 0.005$, start $(1, -1)$



Έπειτα, εξετάζεται το πλήθος των επαναλήψεων που χρειάζονται για να συγκλίνει η αντικειμενική συνάρτηση ξεκινώντας από κάθε ένα αρχικό σημείο, χρησιμοποιώντας τη μέθοδο μέγιστης καθόδου, επιλέγοντας βήμα τέτοιο ώστε να ελαχιστοποιεί την συνάρτηση $f(x_\kappa + \gamma_\kappa d_\kappa)$.

Figure 3.7: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$., Minimizing step, start $(0, 0)$

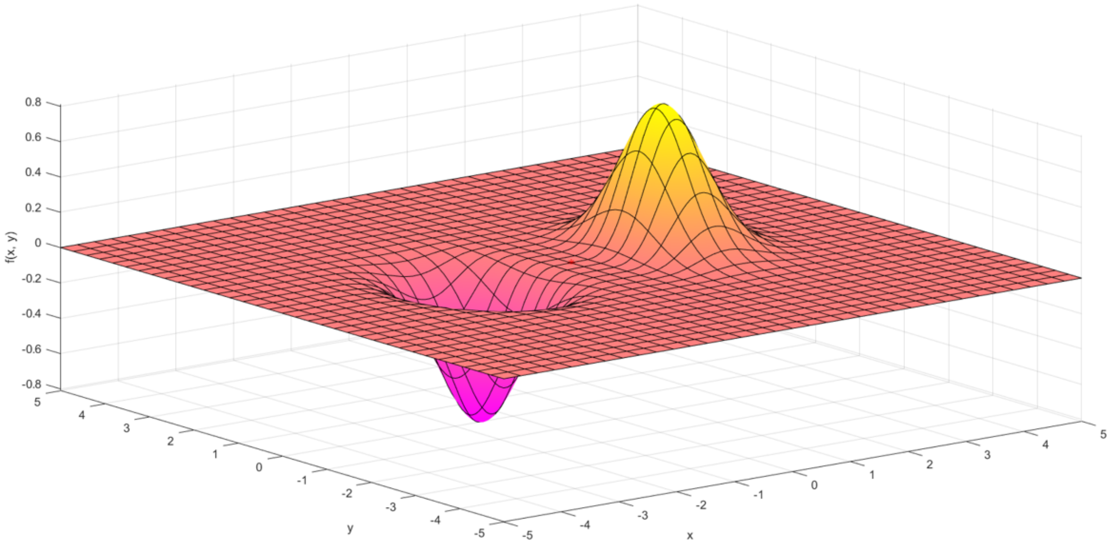


Figure 3.8: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$., Minimizing step, start $(-1, 1)$

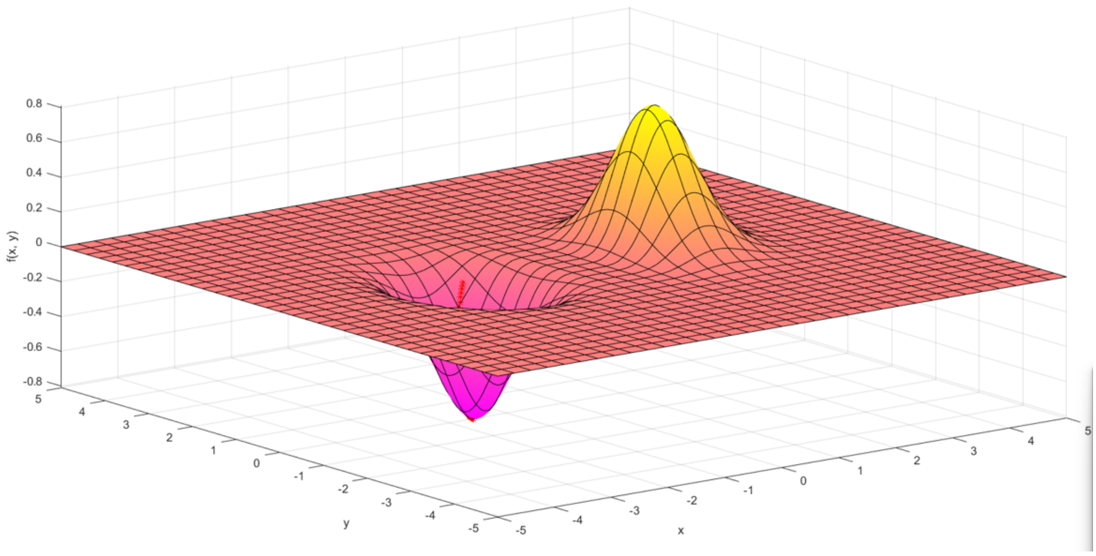


Figure 3.9: of function $f(x, y) = x^5 e^{-x^2-y^2}$., Minimizing step, start $(-1, 1)$

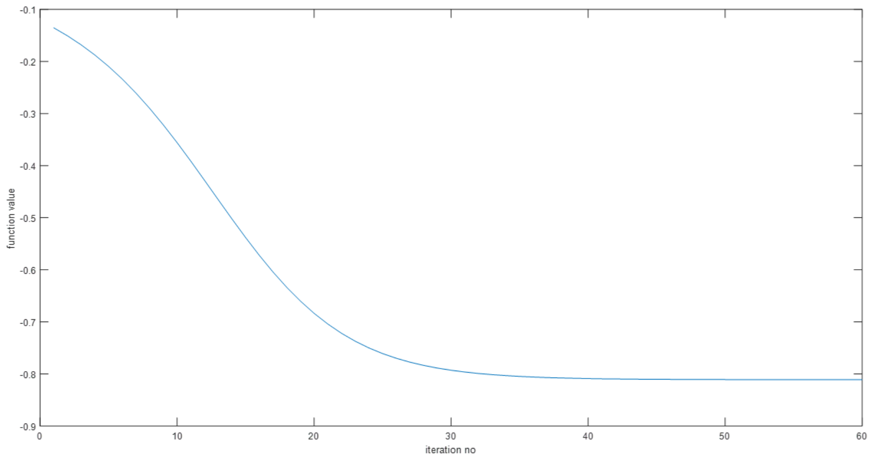


Figure 3.10: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$., Minimizing step, start $(1, -1)$

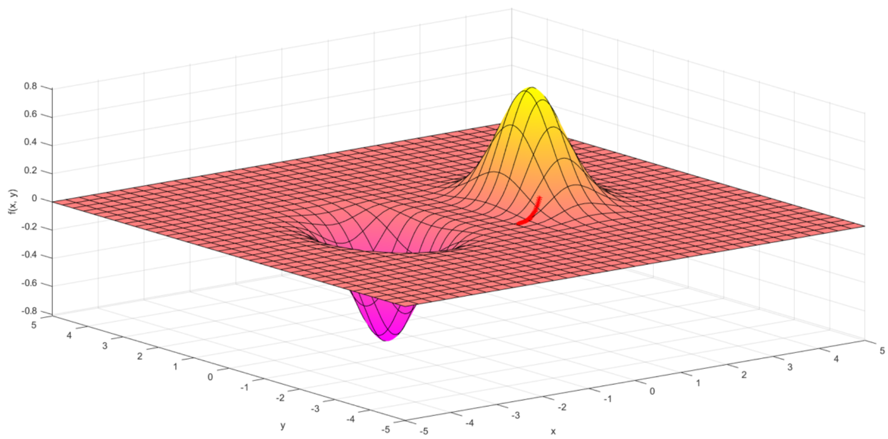
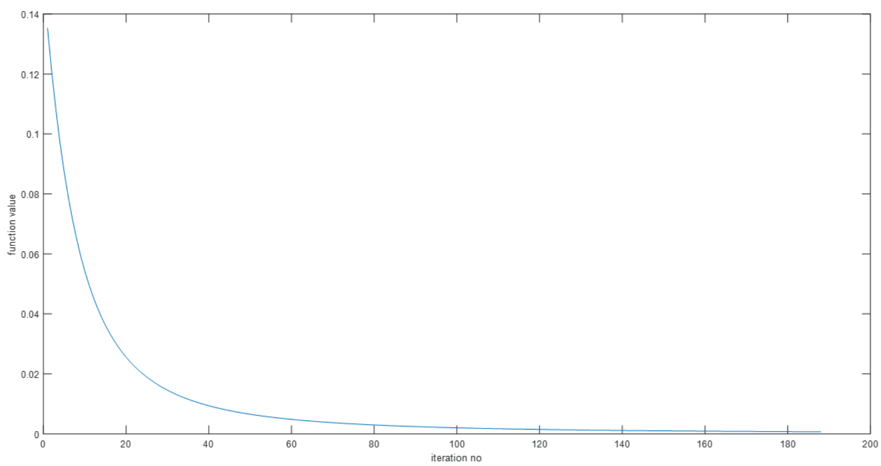


Figure 3.11: of function $f(x, y) = x^5 e^{-x^2-y^2}$., Minimizing step, start $(1, -1)$



Τέλος, εξετάζεται το πλήθος των επαναλήψεων που χρειάζονται για να συγκλίνει η αντικειμενική συνάρτηση ξεκινώντας από κάθε ένα αρχικό σημείο, χρησιμοποιώντας τη μέθοδο μέγιστης καθόδου, επιλέγοντας βήμα βάσει του κανόνα Armijo.

Figure 3.12: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$., Armijo's Rule step, start $(0, 0)$

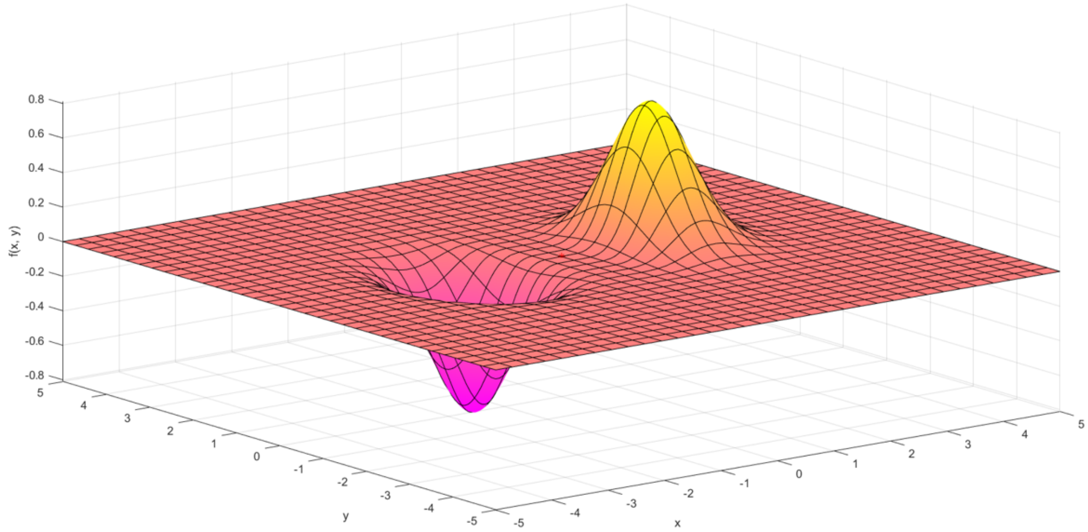


Figure 3.13: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$., Armijo's Rule step, start $(-1, 1)$

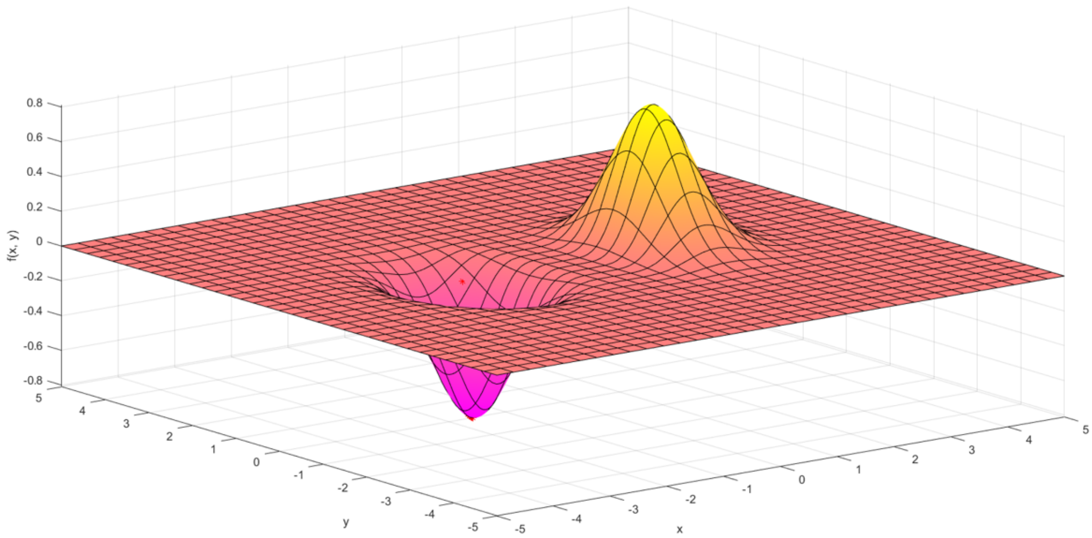


Figure 3.14: of function $f(x, y) = x^5 e^{-x^2-y^2}$., Armijo′s Rule step, start $(-1, 1)$

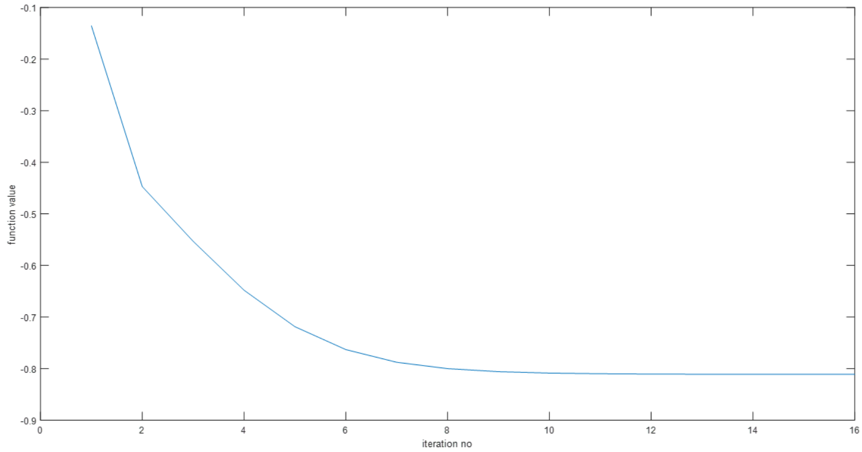


Figure 3.15: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$., Armijo′s Rule step, start $(1, -1)$

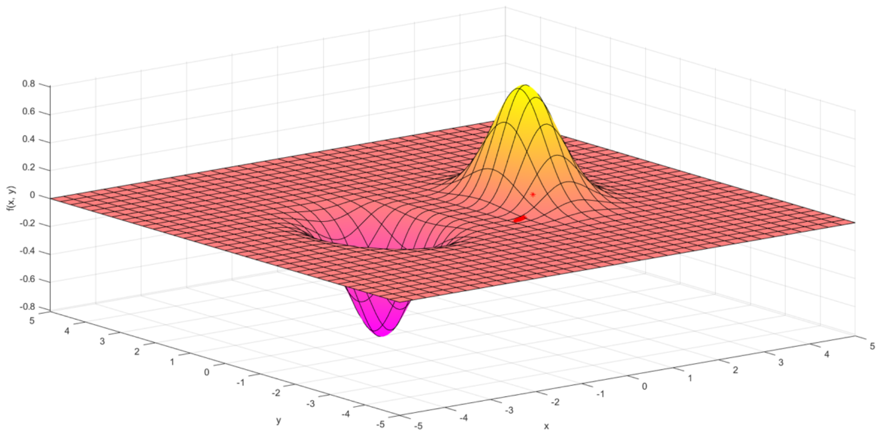
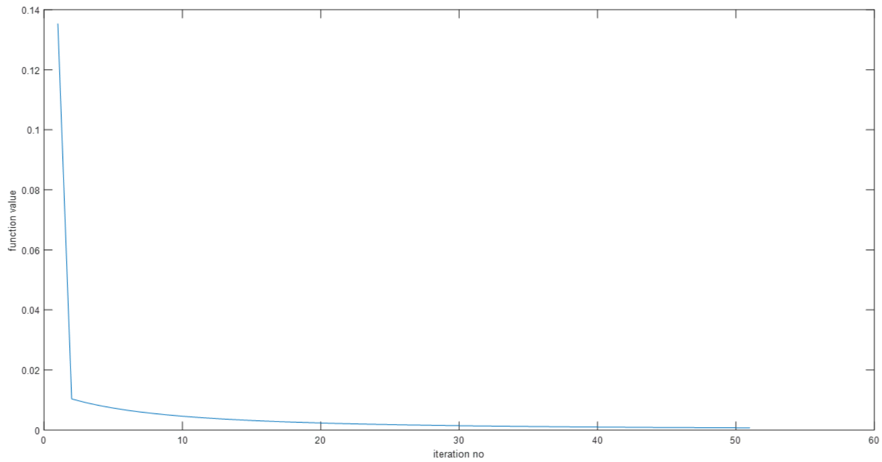


Figure 3.16: of function $f(x, y) = x^5 e^{-x^2-y^2}$., Armijo′s Rule step, start $(1, -1)$



Παρατηρήσεις:

Παρατηρούμε ότι η μέθοδος που υπολογίζει το βήμα με στόχο τη μείωση της τιμής της συνάρτησης απαιτεί λιγότερες επαναλήψεις για να συγκλίνει, συγκριτικά με τη μέθοδο που χρησιμοποιεί σταθερό βήμα. Επιπλέον, η μέθοδος που επιλέγει το βήμα βάσει του κανόνα του Armijo απαιτεί ακόμα λιγότερες επαναλήψεις, αποδεικνύοντας την αποδοτικότητά της.

Επίσης, παρατηρούμε ότι όταν οι μέθοδοι ξεκινούν από το σημείο $(1, -1)$, τείνουν να εγκλωβίζονται σε σημεία της μορφής $(0, y)$. Αυτό είναι αναμενόμενο, καθώς στα σημεία αυτά η κλίση της συνάρτησης ικανοποιεί τη συνθήκη $|\nabla f| = 0$.

Πιο αναλυτικά και συγκεντρωτικά, από την παρατήρηση αυτών των γραφημάτων μπορούμε να εξάγουμε τα εξής συμπεράσματα:

- **Επίδραση Σημείου Εκκίνησης:**
  - Τα σημεία εκκίνησης επηρεάζουν σημαντικά την πορεία της σύγκλισης. Για παράδειγμα, όταν το σημείο εκκίνησης βρίσκεται μακριά από το τοπικό ελάχιστο, οι επαναλήψεις που απαιτούνται για τη σύγκλιση αυξάνονται.
  - Σε ορισμένα σημεία, όπως το $(1, -1)$, παρατηρείται ¨εγκλωβισμός' σε συγκεκριμένες περιοχές, κάτι που οφείλεται στη μορφή της συνάρτησης και στη μη επαρκή επιλογή κατεύθυνσης.
- **Επίδραση Μεγέθους Βήματος:**
  - Ένα πολύ μικρό βήμα οδηγεί σε αργή σύγκλιση, καθώς η αλλαγή της τιμής της συνάρτησης ανά επανάληψη είναι μικρή.
  - Αντίθετα, ένα υπερβολικά μεγάλο βήμα μπορεί να οδηγήσει σε αποσταθεροποίηση ή μη σύγκλιση, καθώς η μέθοδος ενδέχεται να υπερβεί το ελάχιστο.
- **Χρήση Κανόνα Armijo:**
  - Η προσαρμογή του βήματος βάσει του κανόνα Armijo προσφέρει μια ισορροπημένη προσέγγιση, εξασφαλίζοντας τόσο τη σταθερότητα όσο και τη γρηγορότερη σύγκλιση, όπως φαίνεται από τα σχετικά γραφήματα.
  - Η μέθοδος αυτή φαίνεται να είναι η πιο αποδοτική επιλογή από πλευράς αριθμού επαναλήψεων.
- **Συνολική Απόδοση:**
  - Γενικά, η μέθοδος της Μέγιστης Καθόδου λειτουργεί αποτελεσματικά, εφόσον οι παράμετροι έχουν επιλεγεί κατάλληλα. Ωστόσο, η επιλογή της αρχικής τιμής και του βήματος είναι κρίσιμες για την απόδοση της μεθόδου.
- **Τριδιάστατη Απεικόνιση:**
  - Τα τριδιάστατα γραφήματα δείχνουν τη διαδρομή που ακολουθεί η μέθοδος προς το ελάχιστο. Μπορεί να παρατηρηθεί πώς η κατεύθυνση της καθόδου προσαρμόζεται στη γεωμετρία της συνάρτησης.

Αυτά τα αποτελέσματα υποδεικνύουν ότι η καλή επιλογή παραμέτρων, όπως το σημείο εκκίνησης και ο κανόνας βήματος, είναι ουσιαστική για τη βελτιστοποίηση της απόδοσης της μεθόδου.

3.3 Θέμα 3

Εκφώνηση

Ελαχιστοποιήστε την $f(x, y) = x^5 e^{-x^2-y^2}$ με τη μέθοδο **Newton**, χρησιμοποιώντας ως αρχικά σημεία (x_0, y_0) τα

i) $(0, 0)$,

ii) $(-1, 1)$, και

iii) $(1, -1)$.

Το βήμα γ_k θα επιλεγεί:

α) σταθερό (της επιλογής σας),

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

γ) βάσει του κανόνα Armijo.

Σχολιάστε τις διαφορές στα αποτελέσματα, σε περίπτωση που προκαλούνται, λόγω της επιλογής του σημείου έναρξης (x_0, y_0) του αλγορίθμου, καθώς επίσης και λόγω της επιλογής του βήματος γ_k. Οδηγούμαστε πάντα σε σωστό αποτέλεσμα; Αν όχι, τι πιστεύετε ότι φταίει;

3.3.1 Μαθηματική Ανάλυση

Αντίθετα με τη μέθοδο μέγιστης καθόδου, η μέθοδος Newton αναζητεί σημείο ελαχίστου στην κατεύθυνση

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

χρησιμοποιεί δηλαδή και τον Hessian πίνακα, ο οποίος θα πρέπει όμως να είναι θετικά ορισμένος.

Υπολογισμός του Εσσιανού Πίνακα

Ο εσσιανός πίνακας $H(x, y)$ είναι ο πίνακας με τις δεύτερες μερικές παραγώγους της συνάρτησης:

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

Για τη συνάρτηση $f(x, y) = x^5 e^{-x^2-y^2}$, οι μερικές παράγωγοι είναι:

1. **Πρώτη μερική παράγωγος ως προς** x:

$$\frac{\partial f}{\partial x} = \left(5x^4 - 2x^6\right) e^{-x^2-y^2}.$$

2. **Πρώτη μερική παράγωγος ως προς** y:

$$\frac{\partial f}{\partial y} = -2xy^5 e^{-x^2-y^2}.$$

3. Δεύτερες μερικές παράγωγοι:

- $\frac{\partial^2 f}{\partial x^2}$:

$$\frac{\partial^2 f}{\partial x^2} = \left(20x^3 - 6x^5 - 2x^5 + 4x^7\right) e^{-x^2-y^2}.$$

- $\frac{\partial^2 f}{\partial y^2}$:

$$\frac{\partial^2 f}{\partial y^2} = \left(-2x^5 + 4x^5y^2\right) e^{-x^2-y^2}.$$

- $\frac{\partial^2 f}{\partial x \partial y}$ και $\frac{\partial^2 f}{\partial y \partial x}$:

$$\frac{\partial^2 f}{\partial x \partial y} = -2x^4 y e^{-x^2-y^2}.$$

Υπολογισμός του Εσσιανού Πίνακα για τα Τρία Σημεία

1. **Σημείο** $(0, 0)$:

$$H = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

2. **Σημείο** $(-1, 1)$:

$$H = \begin{bmatrix} -2e^{-2} & -6e^{-2} \\ -6e^{-2} & -2e^{-2} \end{bmatrix}.$$

3. **Σημείο** $(1, -1)$:

$$H = \begin{bmatrix} 2e^{-2} & 6e^{-2} \\ 6e^{-2} & 2e^{-2} \end{bmatrix}.$$

Εξέταση κάθε σημείου:

1. Σημείο $(0, 0)$

Ο πίνακας H είναι:

$$H = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

- Το πρώτο στοιχείο είναι $H_{11} = 0$, το οποίο δεν είναι θετικό.
- Ο προσδιοριστής του πίνακα είναι:

$$\det(H) = 0.$$

Άρα, ο πίνακας δεν είναι θετικά ορισμένος.

2. Σημείο $(-1, 1)$

Ο πίνακας H είναι:

$$H = \begin{bmatrix} -2e^{-2} & -6e^{-2} \\ -6e^{-2} & -2e^{-2} \end{bmatrix}.$$

- Το πρώτο στοιχείο είναι $H_{11} = -2e^{-2}$, το οποίο είναι αρνητικό.

- Ο προσδιοριστής του πίνακα είναι:
$$\det(H) = (-2e^{-2})(-2e^{-2}) - (-6e^{-2})(-6e^{-2}) = 4e^{-4} - 36e^{-4} = -32e^{-4},$$
που είναι αρνητικός.

Άρα, ο πίνακας δεν είναι θετικά ορισμένος.

3. **Σημείο** $(1, -1)$

Ο πίνακας H είναι:
$$H = \begin{bmatrix} 2e^{-2} & 6e^{-2} \\ 6e^{-2} & 2e^{-2} \end{bmatrix}.$$

- Το πρώτο στοιχείο είναι $H_{11} = 2e^{-2}$, το οποίο είναι θετικό.
- Ο προσδιοριστής του πίνακα είναι:
$$\det(H) = (2e^{-2})(2e^{-2}) - (6e^{-2})(6e^{-2}) = 4e^{-4} - 36e^{-4} = -32e^{-4},$$
που είναι αρνητικός.

Άρα, ο πίνακας δεν είναι θετικά ορισμένος.

Συμπέρασμα

Κανένας από τους πίνακες για τα σημεία $(0, 0)$, $(-1, 1)$ και $(1, -1)$ δεν είναι θετικά ορισμένος. Η μέθοδος Newton εξ ορισμού δεν μπορεί να εφαρμοστεί για την συγκεκριμένη συνάρτηση.

Παρόλα αυτά, ο κώδικας δίνεται για λόγους πληρότητας στο αρχείο newton.m. Είναι εύλογο να πούμε ότι η μέθοδος Newton δεν μας δίνει καθόλου καλά συμπεράσματα για ελάχιστο της συνάρτησης εξαιτίας της μορφής αυτής και του μη θετικά ορισμένου εσσιανού πίνακα της. Ακόμη, σε ορισμένες περιπτώσεις η μέθοδος δεν τερματίζει.

3.4 Θέμα 4

Εκφώνηση

Ελαχιστοποιήστε την $f(x, y) = x^5 e^{-x^2-y^2}$ με τη μέθοδο **Levenberg Marquardt**, χρησιμοποιώντας ως αρχικά σημεία (x_0, y_0) τα

i) $(0, 0)$,

ii) $(-1, 1)$, και

iii) $(1, -1)$.

Το βήμα γ_k θα επιλεγεί:

α) σταθερό (της επιλογής σας),

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

γ) βάσει του κανόνα Armijo.

Σχολιάστε τις διαφορές στα αποτελέσματα, σε περίπτωση που προκαλούνται, λόγω της επιλογής του σημείου έναρξης (x_0, y_0) του αλγορίθμου, καθώς επίσης και λόγω της επιλογής του βήματος γ_k. Οδηγούμαστε πάντα σε σωστό αποτέλεσμα; Αν όχι, τι πιστεύετε ότι φταίει;

3.4.1 Μαθηματική Ανάλυση

Η γενική διαδικασία της μεθόδου Levenberg Marquardt με σταθερό βήμα μπορεί να περιγραφεί ως εξής:

1. Αρχικοποίηση:
   - Επιλέγουμε το αρχικό σημείο (x_0, y_0).
   - Θέτουμε την παράμετρο $\lambda > 0$ και το σταθερό βήμα $\gamma > 0$.
   - Καθορίζουμε το όριο σύγκλισης $\epsilon > 0$.
2. Υπολογισμός της Κλίσης:
$$\nabla f(x_k, y_k) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}.$$
3. Υπολογισμός του Εσσιανού Πίνακα:
$$H(x_k, y_k) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}.$$
4. Τροποποίηση του Εσσιανού Πίνακα:
$$H_k + \lambda I = \begin{bmatrix} H_{11} + \lambda & H_{12} \\ H_{21} & H_{22} + \lambda \end{bmatrix}.$$
5. Υπολογισμός της Κατεύθυνσης Ενημέρωσης:
$$d_k = -(H_k + \lambda I)^{-1} \cdot \nabla f(x_k, y_k).$$
6. Ενημέρωση του Σημείου:
$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \gamma d_k.$$
7. Συνθήκη Σύγκλισης: Επαναλαμβάνουμε τα παραπάνω βήματα μέχρι να ισχύει:
$$\|\nabla f(x_k, y_k)\| < \epsilon.$$

Ανάλυση για τη Μέθοδο Levenberg Marquardt με Σταθερό Βήμα ($\gamma = 0.1$)

Σημείο 1: $(x_0, y_0) = (0, 0)$

1. Υπολογισμός Κλίσης:
$$\nabla f(0, 0) = [0, 0].$$
   Η κλίση είναι μηδενική.
2. Εσσιανός Πίνακας:
$$H(0, 0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$
3. Τροποποιημένος Πίνακας: Με $\lambda = 1$,
$$H_k + \lambda I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

4. Κατεύθυνση Ενημέρωσης:
$$d_k = -(H_k + \lambda I)^{-1} \cdot \nabla f_k = [0, 0].$$
5. Ενημέρωση Σημείου:
$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \gamma d_k = (0, 0).$$

Συμπέρασμα: Ο αλγόριθμος τερματίζει άμεσα, καθώς η κλίση είναι μηδενική.

Σημείο 2: $(x_0, y_0) = (-1, 1)$

1. Υπολογισμός Κλίσης:
$$\nabla f(-1, 1) = [3e^{-2}, -2e^{-2}] \approx [0.406, -0.271].$$
2. Εσσιανός Πίνακας:
$$H(-1, 1) = \begin{bmatrix} -2e^{-2} & -6e^{-2} \\ -6e^{-2} & -2e^{-2} \end{bmatrix}.$$
3. Τροποποιημένος Πίνακας: Με $\lambda = 1$,
$$H_k + \lambda I = \begin{bmatrix} -2e^{-2} + 1 & -6e^{-2} \\ -6e^{-2} & -2e^{-2} + 1 \end{bmatrix}.$$
4. Κατεύθυνση Ενημέρωσης:
$$d_k = -(H_k + \lambda I)^{-1} \cdot \nabla f_k.$$
5. Ενημέρωση Σημείου:
$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \gamma d_k.$$
6. Επανάληψη: Επαναλαμβάνουμε τη διαδικασία μέχρι η κλίση να ικανοποιεί την προϋπόθεση $\|\nabla f_k\| < \epsilon$.

Σημείο 3: $(x_0, y_0) = (1, -1)$

1. Υπολογισμός Κλίσης:
$$\nabla f(1, -1) = [-3e^{-2}, 2e^{-2}] \approx [-0.406, 0.271].$$
2. Εσσιανός Πίνακας:
$$H(1, -1) = \begin{bmatrix} 2e^{-2} & 6e^{-2} \\ 6e^{-2} & 2e^{-2} \end{bmatrix}.$$
3. Τροποποιημένος Πίνακας: Με $\lambda = 1$,
$$H_k + \lambda I = \begin{bmatrix} 2e^{-2} + 1 & 6e^{-2} \\ 6e^{-2} & 2e^{-2} + 1 \end{bmatrix}.$$
4. Κατεύθυνση Ενημέρωσης:
$$d_k = -(H_k + \lambda I)^{-1} \cdot \nabla f_k.$$
5. Ενημέρωση Σημείου:
$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \gamma d_k.$$
6. Επανάληψη: Επαναλαμβάνουμε τη διαδικασία μέχρι η κλίση να ικανοποιεί την προϋπόθεση $\|\nabla f_k\| < \epsilon$.

Γενική Μορφή της Διαδικασίας με Βήμα γ που Ελαχιστοποιεί τη Συνάρτηση $f(x + \gamma d)$

Η γενική διαδικασία της μεθόδου Levenberg Marquardt, όταν το βήμα γ επιλέγεται έτσι ώστε να ελαχιστοποιεί τη συνάρτηση $f(x + \gamma d)$, έχει ως εξής:

1. Αρχικοποίηση:
   - Επιλέγουμε το αρχικό σημείο (x_0, y_0).
   - Θέτουμε την παράμετρο $\lambda > 0$.
   - Καθορίζουμε το όριο σύγκλισης $\epsilon > 0$.
2. Υπολογισμός της Κλίσης:
$$\nabla f(x_k, y_k) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}.$$
3. Υπολογισμός του Εσσιανού Πίνακα:
$$H(x_k, y_k) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}.$$
4. Τροποποίηση του Εσσιανού Πίνακα:
$$H_k + \lambda I = \begin{bmatrix} H_{11} + \lambda & H_{12} \\ H_{21} & H_{22} + \lambda \end{bmatrix}.$$
5. Υπολογισμός της Κατεύθυνσης Ενημέρωσης:
$$d_k = -(H_k + \lambda I)^{-1} \cdot \nabla f(x_k, y_k).$$
6. Υπολογισμός του Βέλτιστου Βήματος γ_k: Το γ_k επιλέγεται ώστε να ελαχιστοποιεί τη συνάρτηση:
$$f(x_k + \gamma d_k).$$
   Αυτό επιτυγχάνεται λύνοντας την εξίσωση:
$$\frac{\partial}{\partial \gamma} f(x_k + \gamma d_k) = 0.$$
   Η εξίσωση αυτή μπορεί να λυθεί αναλυτικά ή αριθμητικά, ανάλογα με τη μορφή της συνάρτησης f.
7. Ενημέρωση του Σημείου:
$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \gamma_k d_k.$$
8. Συνθήκη Σύγκλισης: Επαναλαμβάνουμε τα παραπάνω βήματα μέχρι να ισχύει:
$$\|\nabla f(x_k, y_k)\| < \epsilon.$$

Ανάλυση για τη Μέθοδο Levenberg Marquardt με Βήμα γ που Ελαχιστοποιεί τη Συνάρτηση $f(x_k + \gamma d_k)$

Σημείο 1: $(x_0, y_0) = (0, 0)$

1. Υπολογισμός Κλίσης:
$$\nabla f(0, 0) = [0, 0].$$
Η κλίση είναι μηδενική.

Συμπέρασμα: Ο αλγόριθμος τερματίζει άμεσα, καθώς η κλίση είναι μηδενική.

Σημείο 2: $(x_0, y_0) = (-1, 1)$

1. Υπολογισμός Κλίσης:
$$\nabla f(-1, 1) = [3e^{-2}, -2e^{-2}] \approx [0.406, -0.271].$$

2. Υπολογισμός Εσσιανού Πίνακα:
$$H(-1, 1) = \begin{bmatrix} -2e^{-2} & -6e^{-2} \\ -6e^{-2} & -2e^{-2} \end{bmatrix}.$$

3. Τροποποιημένος Εσσιανός Πίνακας: Για $\lambda = 1$,
$$H_k + \lambda I = \begin{bmatrix} -2e^{-2} + 1 & -6e^{-2} \\ -6e^{-2} & -2e^{-2} + 1 \end{bmatrix}.$$

4. Υπολογισμός της Κατεύθυνσης Ενημέρωσης:
$$d_k = -(H_k + \lambda I)^{-1} \cdot \nabla f_k.$$

5. Υπολογισμός του Βέλτιστου Βήματος γ_k: Το γ_k επιλέγεται έτσι ώστε να ελαχιστοποιεί τη συνάρτηση:
$$f(x_k + \gamma d_k).$$
Υπολογίζουμε:
$$\frac{\partial}{\partial \gamma} f(x_k + \gamma d_k) = 0.$$

6. Ενημέρωση του Σημείου:
$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \gamma_k d_k.$$

7. Επανάληψη: Επαναλαμβάνουμε τη διαδικασία μέχρι να ισχύει:
$$\|\nabla f(x_k, y_k)\| < \epsilon.$$

Σημείο 3: $(x_0, y_0) = (1, -1)$

1. Υπολογισμός Κλίσης:

$$\nabla f(1, -1) = [-3e^{-2}, 2e^{-2}] \approx [-0.406, 0.271].$$

2. Υπολογισμός Εσσιανού Πίνακα:

$$H(1, -1) = \begin{bmatrix} 2e^{-2} & 6e^{-2} \\ 6e^{-2} & 2e^{-2} \end{bmatrix}.$$

3. Τροποποιημένος Εσσιανός Πίνακας: Για $\lambda = 1$,

$$H_k + \lambda I = \begin{bmatrix} 2e^{-2} + 1 & 6e^{-2} \\ 6e^{-2} & 2e^{-2} + 1 \end{bmatrix}.$$

4. Υπολογισμός της Κατεύθυνσης Ενημέρωσης:

$$d_k = -(H_k + \lambda I)^{-1} \cdot \nabla f_k.$$

5. Υπολογισμός του Βέλτιστου Βήματος γ_k: Το γ_k επιλέγεται έτσι ώστε να ελαχιστοποιεί τη συνάρτηση:

$$f(x_k + \gamma d_k).$$

Υπολογίζουμε:

$$\frac{\partial}{\partial \gamma} f(x_k + \gamma d_k) = 0.$$

6. Ενημέρωση του Σημείου:

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \gamma_k d_k.$$

7. Επανάληψη: Επαναλαμβάνουμε τη διαδικασία μέχρι να ισχύει:

$$\|\nabla f(x_k, y_k)\| < \epsilon.$$

Γενική Μορφή της Διαδικασίας με Βήμα που Επιλέγεται Βάσει του Κανόνα Armijo

Η μέθοδος Levenberg-Marquardt με επιλογή βήματος σύμφωνα με τον κανόνα Armijo εξασφαλίζει επαρκή μείωση της συνάρτησης μέσω μιας ανισότητας.Η γενική διαδικασία περιγράφεται ως εξής:

1. Αρχικοποίηση:
   - Επιλέγουμε το αρχικό σημείο (x_0, y_0).
   - Θέτουμε την παράμετρο $\lambda > 0$.
   - Καθορίζουμε το όριο σύγκλισης $\epsilon > 0$.
   - Επιλέγουμε σταθερά $\sigma \in (0, 1)$, συνήθως $\sigma = 0.1$ ή $\sigma = 0.01$.

2. Υπολογισμός της Κλίσης:

$$\nabla f(x_k, y_k) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}.$$

3. Υπολογισμός του Εσσιανού Πίνακα:

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

4. Τροποποίηση του Εσσιανού Πίνακα:

$$H_k + \lambda I = \begin{bmatrix} H_{11} + \lambda & H_{12} \\ H_{21} & H_{22} + \lambda \end{bmatrix}.$$

5. Υπολογισμός της Κατεύθυνσης Ενημέρωσης:

$$d_k = -(H_k + \lambda I)^{-1} \cdot \nabla f(x_k, y_k).$$

6. Επιλογή του Βέλτιστου Βήματος γ_k: Το βήμα γ_k επιλέγεται έτσι ώστε να ικανοποιεί την ανισότητα του κανόνα Armijo:

$$f(x_k + \gamma_k d_k) \leq f(x_k) + \sigma \gamma_k \nabla f(x_k)^T d_k.$$

   Το γ_k υπολογίζεται χρησιμοποιώντας μια γραμμική αναζήτηση, ξεκινώντας από μια αρχική τιμή $\gamma_k = \gamma_0$ και μειώνοντάς την με έναν παράγοντα $\beta \in (0, 1)$, π.χ., $\beta = 0.5$, μέχρι να ικανοποιηθεί η παραπάνω ανισότητα.

7. Ενημέρωση του Σημείου:

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \gamma_k d_k.$$

8. Συνθήκη Σύγκλισης: Επαναλαμβάνουμε τα παραπάνω βήματα μέχρι να ισχύει:

$$\|\nabla f(x_k, y_k)\| < \epsilon.$$

Ανάλυση για τη Μέθοδο Levenberg-Marquardt με Βήμα που Επιλέγεται Βάσει του Κανόνα Armijo

Σημείο 1: $(x_0, y_0) = (0, 0)$

1. Υπολογισμός Κλίσης:

$$\nabla f(0, 0) = [0, 0].$$

   Η κλίση είναι μηδενική.

Συμπέρασμα: Ο αλγόριθμος τερματίζει άμεσα, καθώς η κλίση είναι μηδενική.

Σημείο 2: $(x_0, y_0) = (-1, 1)$

1. Υπολογισμός Κλίσης:

$$\nabla f(-1, 1) = [3e^{-2}, -2e^{-2}] \approx [0.406, -0.271].$$

2. Υπολογισμός Εσσιανού Πίνακα:

$$H(-1, 1) = \begin{bmatrix} -2e^{-2} & -6e^{-2} \\ -6e^{-2} & -2e^{-2} \end{bmatrix}.$$

3. Τροποποιημένος Εσσιανός Πίνακας: Για $\lambda = 1$,

$$H_k + \lambda I = \begin{bmatrix} -2e^{-2} + 1 & -6e^{-2} \\ -6e^{-2} & -2e^{-2} + 1 \end{bmatrix}.$$

4. Υπολογισμός της Κατεύθυνσης Ενημέρωσης:

$$d_k = -(H_k + \lambda I)^{-1} \cdot \nabla f_k.$$

5. Υπολογισμός Βήματος γ_k: Το βήμα γ_k επιλέγεται έτσι ώστε να ικανοποιεί την ανισότητα του κανόνα Armijo:

$$f(x_k + \gamma_k d_k) \leq f(x_k) + \sigma \gamma_k \nabla f_k^T d_k.$$

Ξεκινάμε με $\gamma_k = \gamma_0$ (π.χ., $\gamma_0 = 1$) και μειώνουμε το γ_k κατά έναν παράγοντα $\beta \in (0, 1)$ (π.χ., $\beta = 0.5$), μέχρι να ικανοποιηθεί η ανισότητα.

6. Ενημέρωση του Σημείου:

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \gamma_k d_k.$$

7. Επανάληψη: Επαναλαμβάνουμε τη διαδικασία μέχρι να ισχύει:

$$\|\nabla f(x_k, y_k)\| < \epsilon.$$

Σημείο 3: $(x_0, y_0) = (1, -1)$

1. Υπολογισμός Κλίσης:

$$\nabla f(1, -1) = [-3e^{-2}, 2e^{-2}] \approx [-0.406, 0.271].$$

2. Υπολογισμός Εσσιανού Πίνακα:

$$H(1, -1) = \begin{bmatrix} 2e^{-2} & 6e^{-2} \\ 6e^{-2} & 2e^{-2} \end{bmatrix}.$$

3. Τροποποιημένος Εσσιανός Πίνακας: Για $\lambda = 1$,

$$H_k + \lambda I = \begin{bmatrix} 2e^{-2} + 1 & 6e^{-2} \\ 6e^{-2} & 2e^{-2} + 1 \end{bmatrix}.$$

4. Υπολογισμός της Κατεύθυνσης Ενημέρωσης:

$$d_k = -(H_k + \lambda I)^{-1} \cdot \nabla f_k.$$

5. Υπολογισμός Βήματος γ_k: Το βήμα γ_k επιλέγεται έτσι ώστε να ικανοποιεί την ανισότητα του κανόνα Armijo:

$$f(x_k + \gamma_k d_k) \leq f(x_k) + \sigma \gamma_k \nabla f_k^T d_k.$$

Ξεκινάμε με $\gamma_k = \gamma_0$ (π.χ., $\gamma_0 = 1$) και μειώνουμε το γ_k κατά έναν παράγοντα $\beta \in (0, 1)$ (π.χ., $\beta = 0.5$), μέχρι να ικανοποιηθεί η ανισότητα.

6. Ενημέρωση του Σημείου:

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \gamma_k d_k.$$

7. Επανάληψη: Επαναλαμβάνουμε τη διαδικασία μέχρι να ισχύει:

$$\|\nabla f(x_k, y_k)\| < \epsilon.$$

3.4.2 Προσομοίωση Σε Matlab

Παρουσιάζονται τα αποτελέσματα της προσομοίωσης του Θέματος 4, στο περιβάλλον *Matlab*, για κάθε ένα από τα τρία σημεία και κάθε τρόπο επιλογής βήματος.

Για την μελέτη αυτή, έχει υλοποιηθεί ο αλγόριθμος της Μεθόδου Levenberg Marquardt, όπως παρουσιάζεται θεωρητικά στο Κεφ. 2, ο κώδικας του οποίου περιέχεται στο αρχείο levenberg_marquardt.m.

Στα γραφήματα παρουσιάζονται τόσο τα αποτελέσματα της μεθόδου, όσο και οι τιμές της αντικειμενικής συνάρτησης για έναν αριθμό επαναλήψεων.

Για αρχικό σημείο εκκίνησης του αλγορίθμου το $(0,0)$ παρουσιάζονται τα ίδια προβλήματα με προηγουμένως όπου υπάρχει δηλαδή εγκλωβισμός σε τοπικό ελάχιστο. Στο διάγραμμα παρακάτω βλέπουμε ότι δεν υπάρχει καμία επανάληψη και ο αλγόριθμος μένει στάσιμος.

Αρχικά, εξετάζεται το πλήθος των επαναλήψεων που χρειάζονται για να συγκλίνει η αντικειμενική συνάρτηση ξεκινώντας από κάθε ένα αρχικό σημείο, χρησιμοποιώντας τη μέθοδο μέγιστης καθόδου, επιλέγοντας σταθερό βήμα ίσο με 0.005.

Figure 3.17: Graph of function $f(x,y) = x^5 e^{-x^2-y^2}$. , step = 0.005, start $(0,0)$

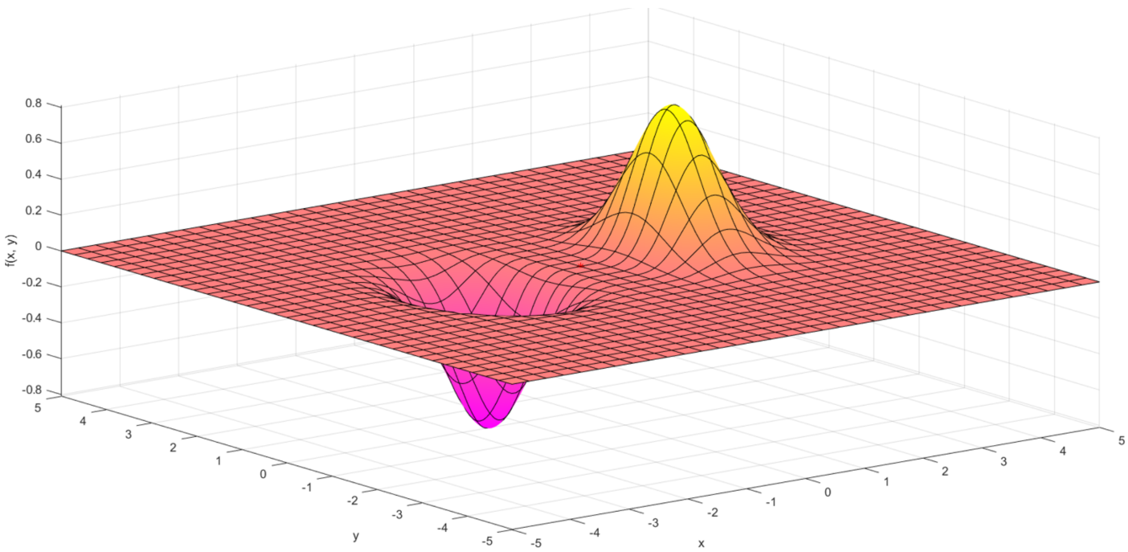


Figure 3.18: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$. , step = 0.005, start $(-1, 1)$

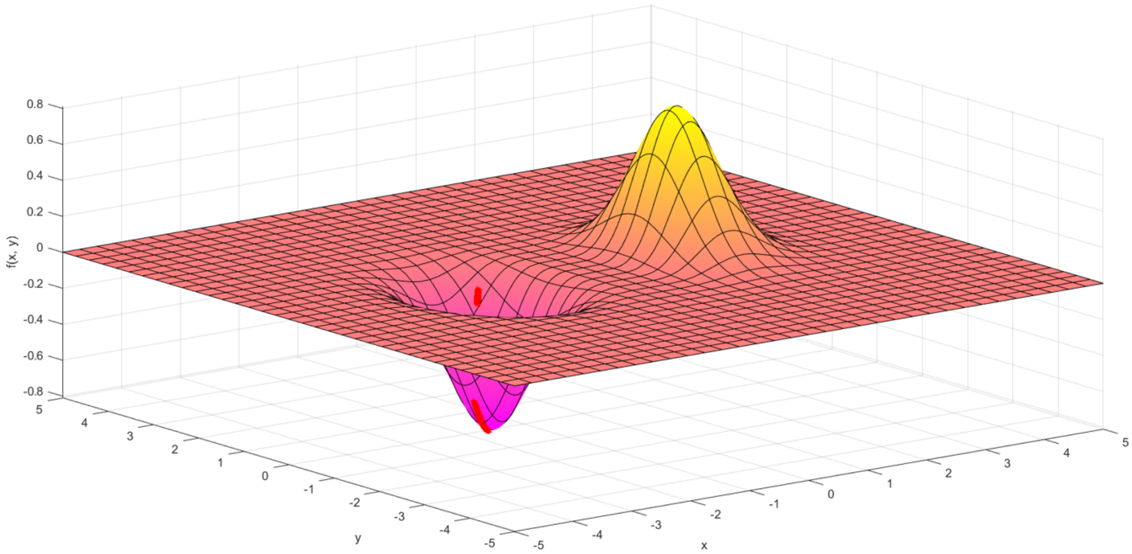


Figure 3.19: of function $f(x, y) = x^5 e^{-x^2-y^2}$. , step = 0.005, start $(-1, 1)$

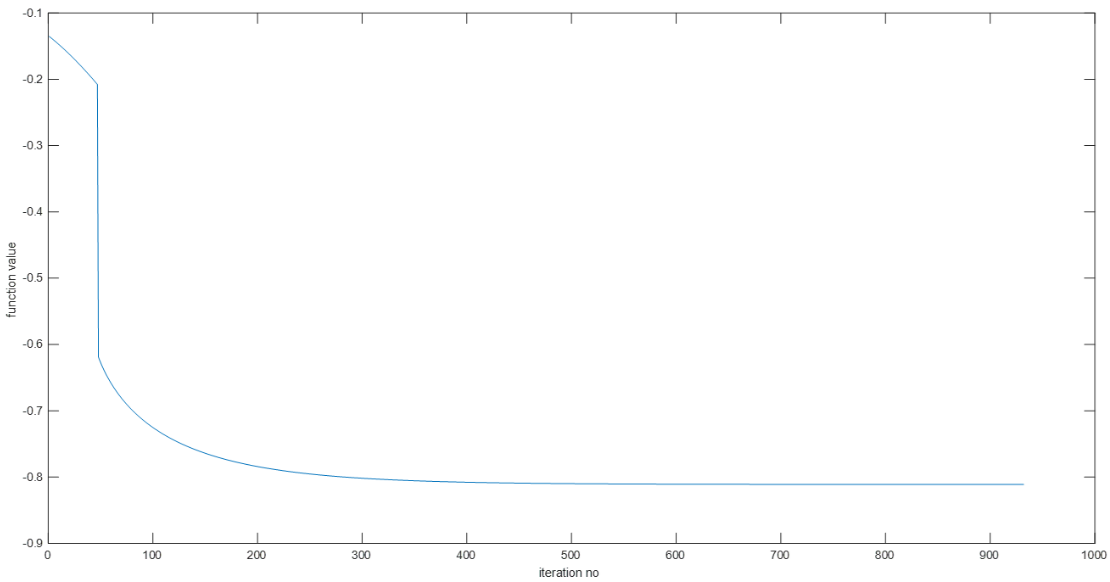


Figure 3.20: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$. , step $= 0.005$, start $(1, -1)$

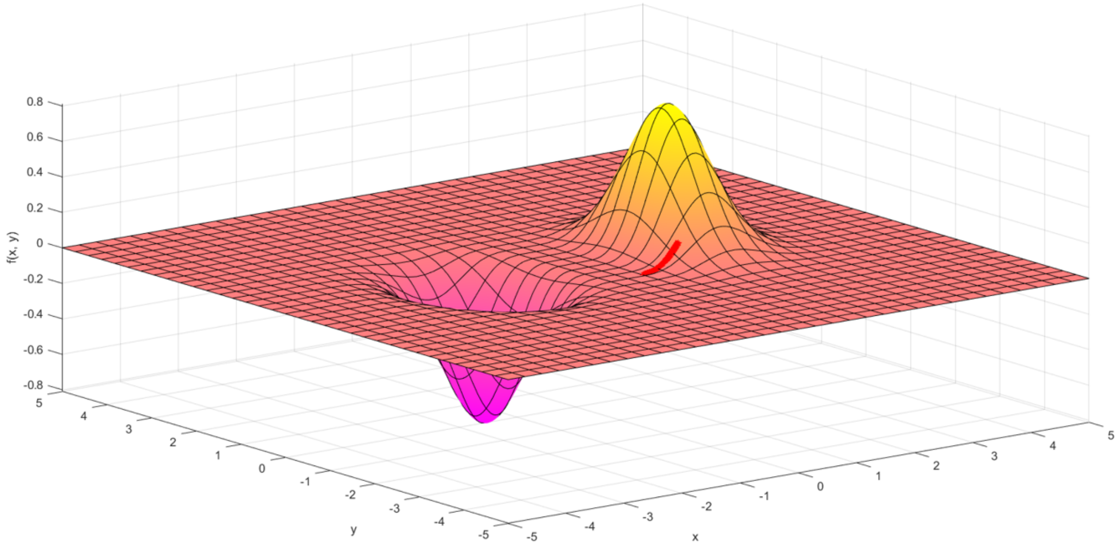
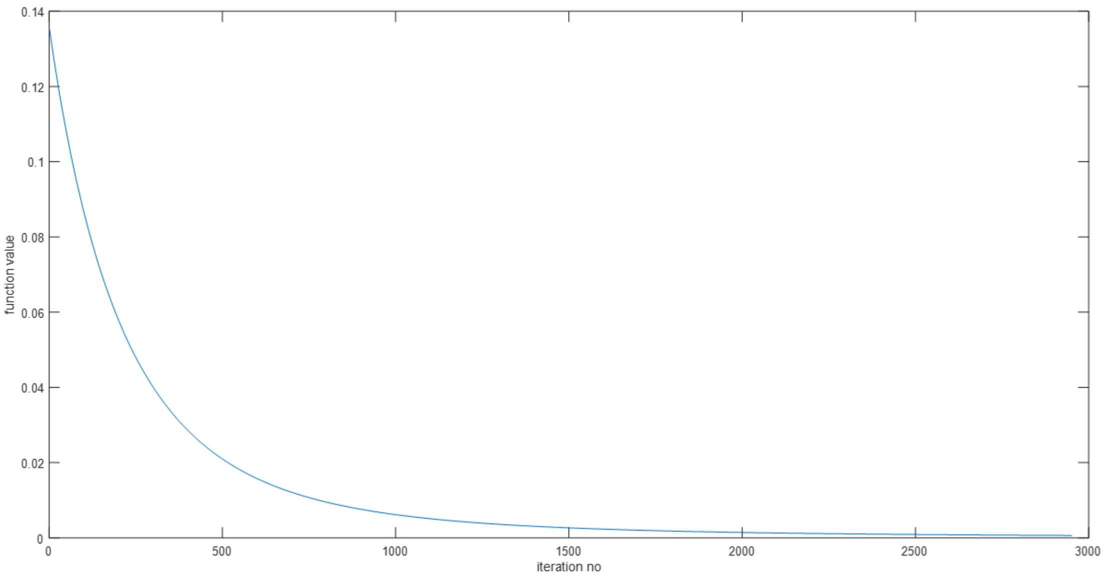


Figure 3.21: of function $f(x, y) = x^5 e^{-x^2-y^2}$. , step $= 0.005$, start $(1, -1)$



Έπειτα, εξετάζεται το πλήθος των επαναλήψεων που χρειάζονται για να συγκλίνει η αντικειμενική συνάρτηση ξεκινώντας από κάθε ένα αρχικό σημείο, χρησιμοποιώντας τη μέθοδο Levenberg Marquardt, επιλέγοντας βήμα τέτοιο ώστε να ελαχιστοποιεί την συνάρτηση $f(x_\kappa + \gamma_\kappa d_\kappa)$.

Figure 3.22: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$., Minimizing step, start $(0, 0)$

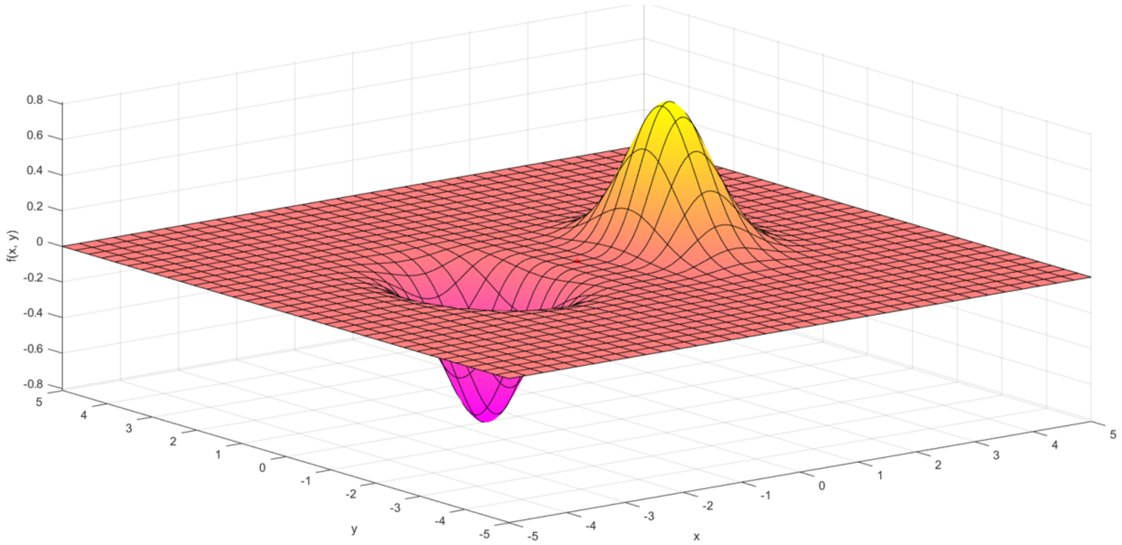


Figure 3.23: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$., Minimizing step, start $(-1, 1)$

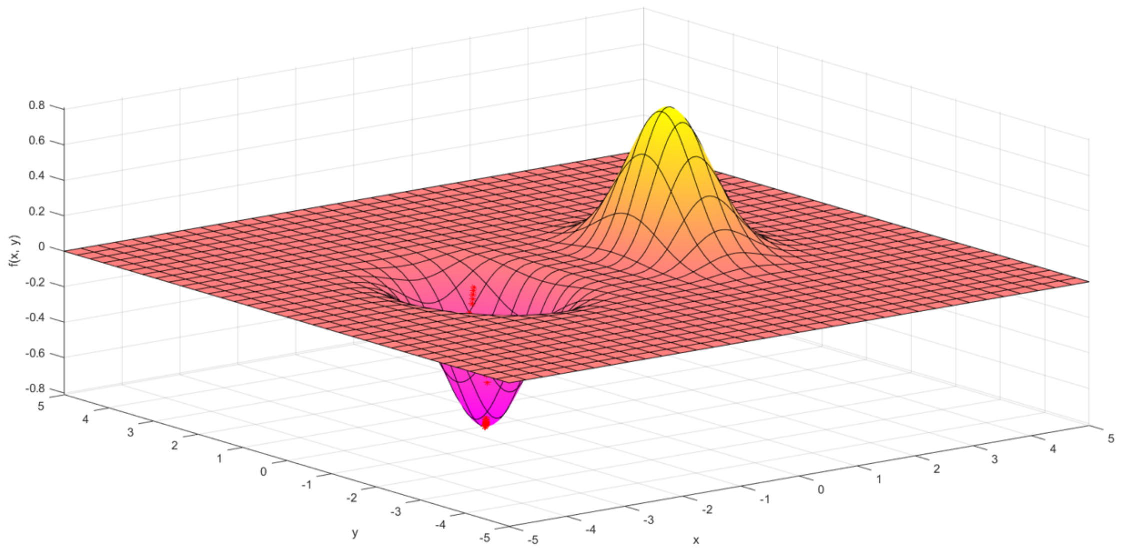


Figure 3.24: of function $f(x,y) = x^5 e^{-x^2-y^2}$., Minimizing step, start $(-1,1)$

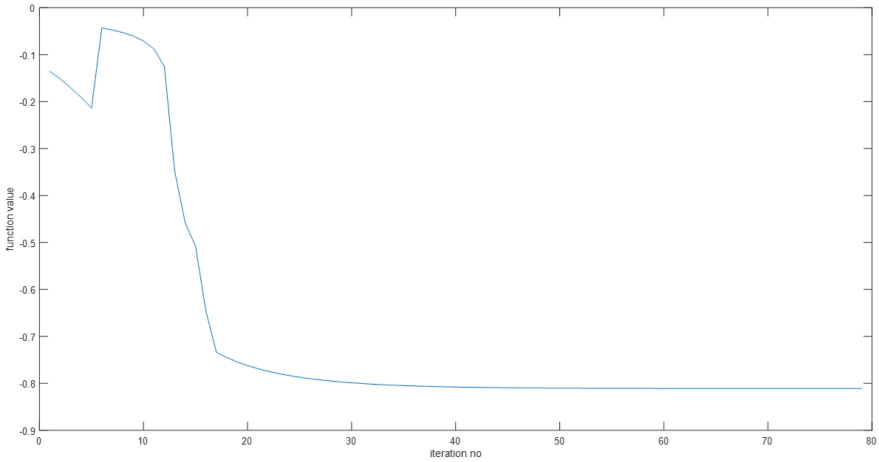


Figure 3.25: Graph of function $f(x,y) = x^5 e^{-x^2-y^2}$., Minimizing step, start $(1,-1)$

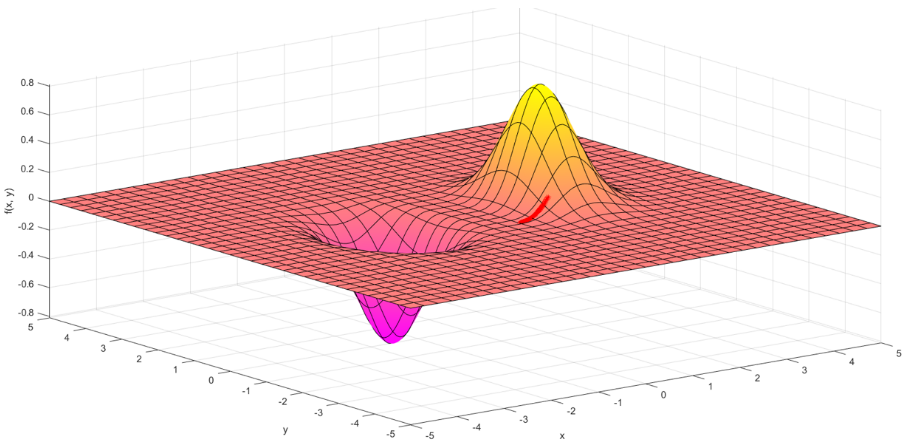
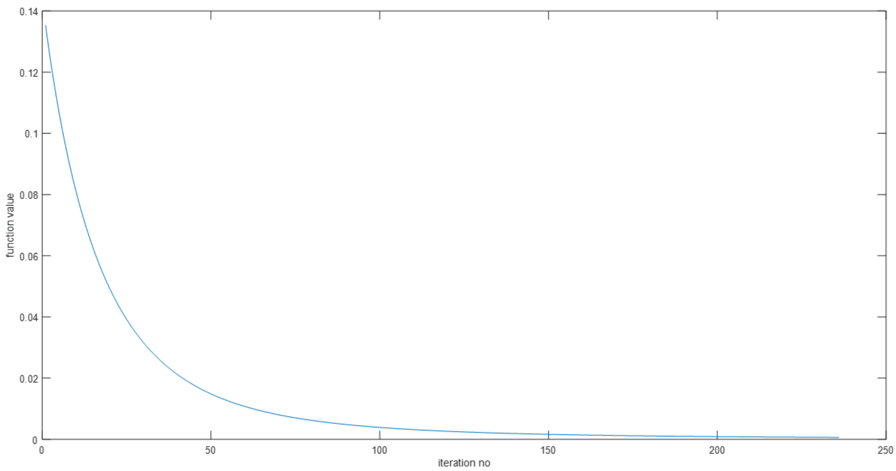


Figure 3.26: of function $f(x,y) = x^5 e^{-x^2-y^2}$., Minimizing step, start $(1,-1)$



Τέλος, εξετάζεται το πλήθος των επαναλήψεων που χρειάζονται για να συγκλίνει η αντικειμενική συνάρτηση ξεκινώντας από κάθε ένα αρχικό σημείο, χρησιμοποιώντας τη μέθοδο Levenberg Marquardt, επιλέγοντας βήμα βάσει του κανόνα Armijo.

Figure 3.27: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$., Armijo's Rule step, start $(0, 0)$

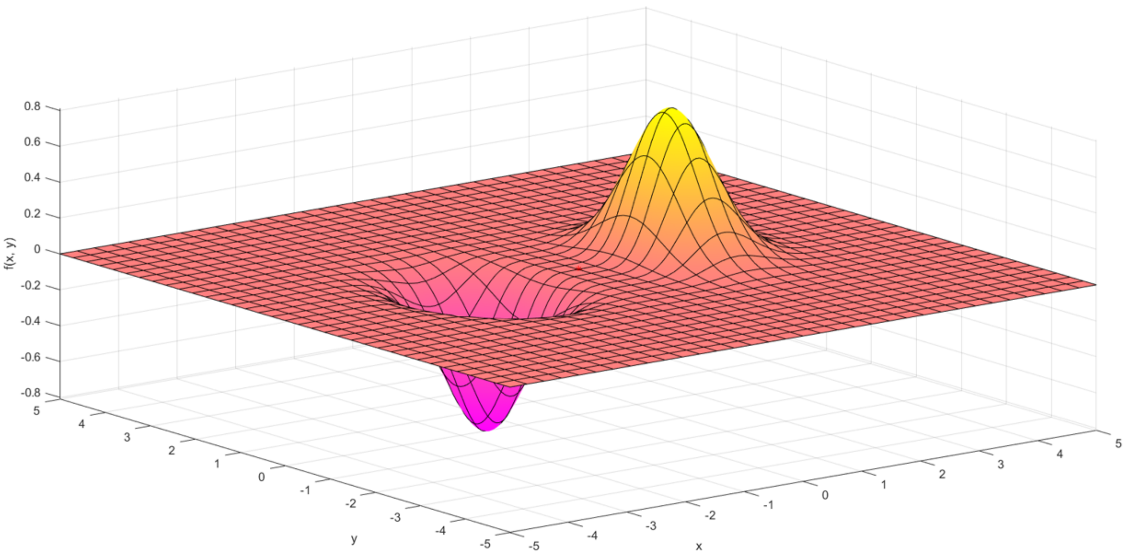


Figure 3.28: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$., Armijo's Rule step, start $(-1, 1)$

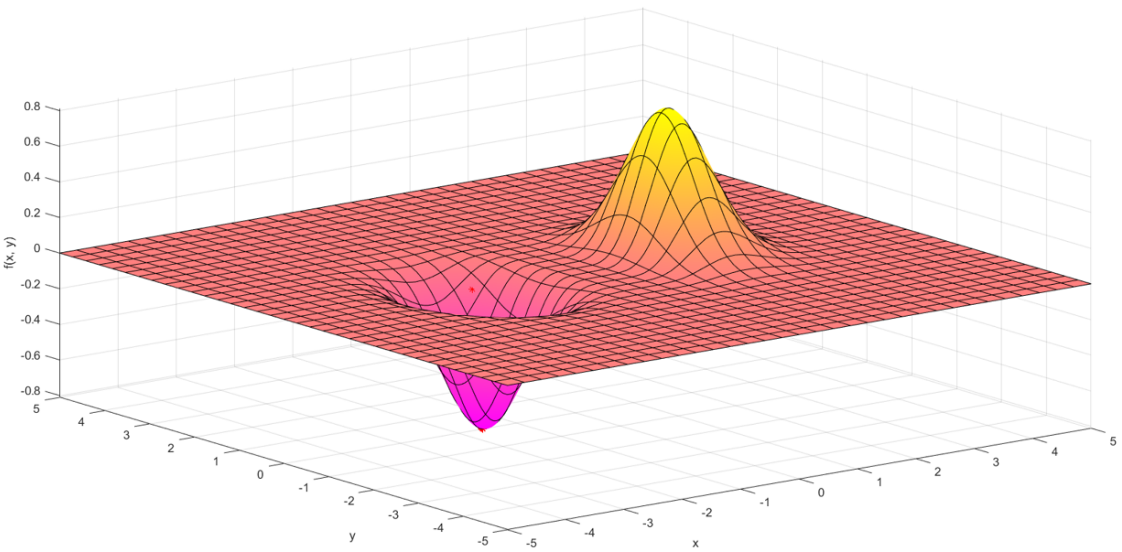


Figure 3.29: of function $f(x, y) = x^5 e^{-x^2-y^2}$., Armijo′s Rule step, start $(-1, 1)$

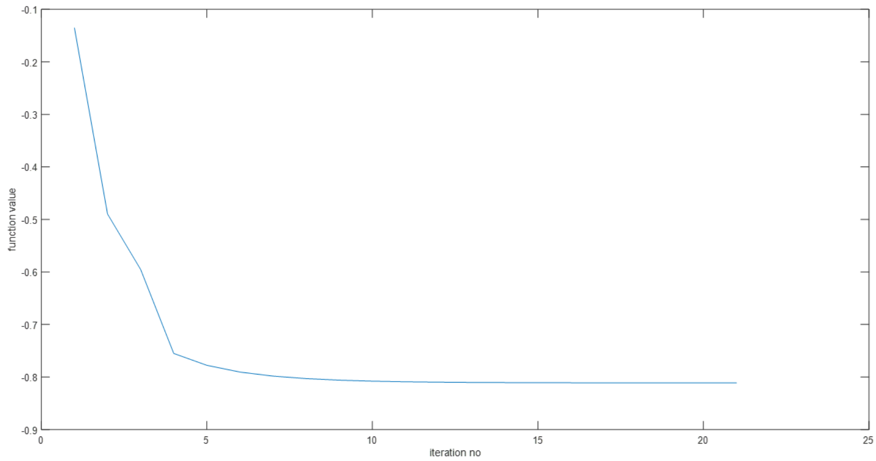


Figure 3.30: Graph of function $f(x, y) = x^5 e^{-x^2-y^2}$., Armijo′s Rule step, start $(1, -1)$

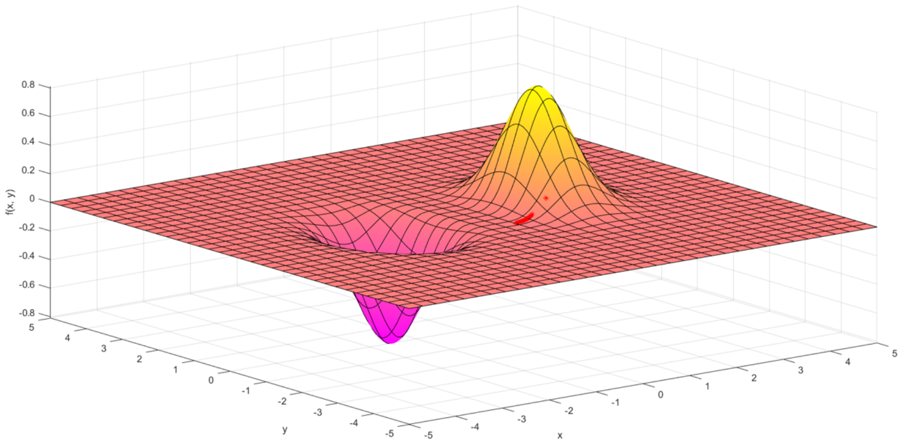
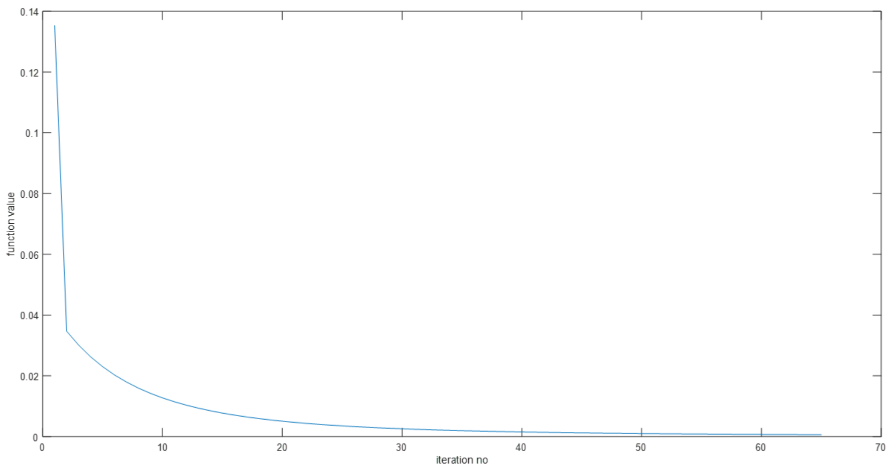


Figure 3.31: of function $f(x, y) = x^5 e^{-x^2-y^2}$., Armijo′s Rule step, start $(1, -1)$



Παρατηρήσεις:

Από τα γραφήματα που παρουσιάζουν τα αποτελέσματα της εφαρμογής της μεθόδου Levenberg-Marquardt, παρατηρούνται τα εξής:

- Συγκέντρωση και Σύγκλιση:
  - Η μέθοδος Levenberg-Marquardt οδηγεί σε σταθερή και γρήγορη σύγκλιση σε σχέση με άλλες μεθόδους (όπως η μέθοδος της Μέγιστης Καθόδου), ειδικά όταν το βήμα επιλέγεται προσεκτικά.
  - Παρατηρείται ότι το πλήθος των επαναλήψεων εξαρτάται σημαντικά από το αρχικό σημείο εκκίνησης και τη γεωμετρία της συνάρτησης.
- Επίδραση του Αρχικού Σημείου:
  - Για τα αρχικά σημεία κοντά στο ελάχιστο, ο αλγόριθμος συγκλίνει ταχύτερα.
  - Όταν το αρχικό σημείο βρίσκεται σε μεγαλύτερη απόσταση, η διαδικασία σύγκλισης επιβραδύνεται, αλλά η μέθοδος προσαρμόζεται μέσω της διόρθωσης που ενσωματώνει ο όρος απόσβεσης (λ).
- Επίδραση Επιλογής Βήματος:
  - Με σταθερό βήμα, παρατηρείται σταθερή μείωση της συνάρτησης, αλλά το πλήθος επαναλήψεων αυξάνεται για σημεία που βρίσκονται πιο μακριά από το ελάχιστο.
  - Με ελαχιστοποίηση του βήματος (γ), επιτυγχάνεται καλύτερη ισορροπία μεταξύ ταχύτητας σύγκλισης και σταθερότητας.
  - Ο κανόνας του Armijo εξασφαλίζει γρήγορη και σταθερή σύγκλιση με μικρότερο πλήθος επαναλήψεων, ιδιαίτερα για πιο απότομα ελάχιστα.
- Γραφήματα:
  - Τα τρισδιάστατα γραφήματα δείχνουν ότι η διαδρομή του αλγορίθμου ακολουθεί την καμπυλότητα της συνάρτησης, κάτι που δικαιολογεί την αποδοτικότητα της μεθόδου για λειτουργίες με πολυπλοκότερη μορφολογία.
  - Στα διαγράμματα τιμών συναρτήσεων ανά επανάληψη, παρατηρείται εκθετική μείωση της τιμής της συνάρτησης, ειδικά όταν χρησιμοποιείται ο κανόνας του Armijo.

Συμπεράσματα

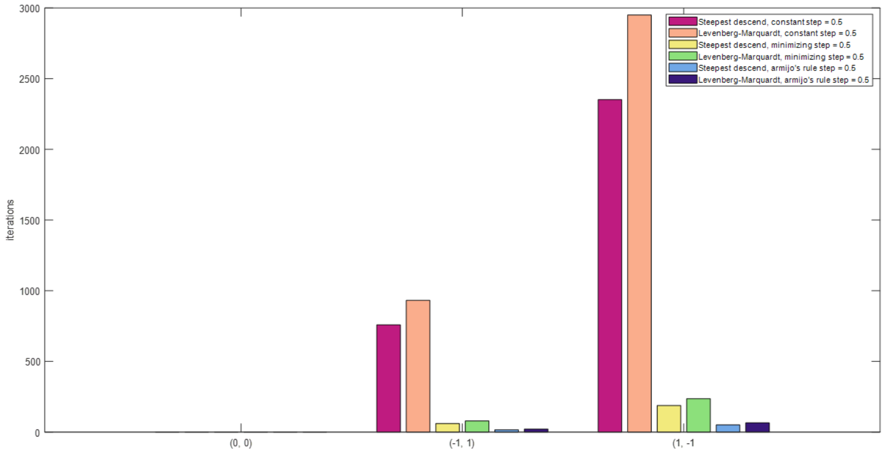
- Η μέθοδος Levenberg-Marquardt συνδυάζει στοιχεία από τις μεθόδους μέγιστης καθόδου και Νεωτον, κάτι που την καθιστά πιο ευέλικτη για διαφορετικά είδη συναρτήσεων.
- Ο κανόνας του Armijo φαίνεται να προσφέρει την καλύτερη απόδοση, καθώς προσαρμόζει το βήμα δυναμικά ανάλογα με την κλίση και την καμπυλότητα της συνάρτησης.
- Η επιλογή του αρχικού σημείου έχει καθοριστική σημασία για την ταχύτητα και την ακρίβεια της σύγκλισης.
- Συνολικά, η μέθοδος είναι ιδανική για τη βελτιστοποίηση μη γραμμικών συναρτήσεων που έχουν σύνθετα τοπικά χαρακτηριστικά.

Κεφάλαιο 4

Σύγκριση Μεθόδων

Για τον συγκριτικό σχολιασμό πάνω στην αποδοτικότητα των υπό μελέτη μεθόδων για την δοθείσα συνάρτηση τρέχουμε τους προηγούμενους αλγορίθμους των μεθόδων για τα ίδια δεδομένα και δημιουργούμε ένα γράφημα που περιέχει τους υπολογισμούς των επαναλήψεων που απαιτούνται για σύγκλιση χρησιμοποιώντας τις διαφορετικές μεθόδους, για τους τρεις διαφορετικούς τρόπους επιλογής βήματος, για καθένα από τα τρία σημεία εκκίνησης. Η σύγκριση των μεθόδων γίνεται στο αρχείο comparison2.m. Ακολουθεί το γράφημα με τα αποτελέσματα:

Figure 4.1: Iterations for Every Method



4.1 Συμπεράσματα:

Το σταθερό βήμα αποδεικνύεται ότι δεν είναι κατάλληλο για την εύρεση του ελαχίστου, καθώς συχνά οδηγεί σε εγκλωβισμό σε τοπικά ακρότατα. Ακόμα και όταν το βήμα είναι αρκετά μεγάλο ώστε να τα υπερβεί, παρουσιάζει δυσκολία στη σύγκλιση κοντά στο ολικό ελάχιστο, καθιστώντας τη μέθοδο λιγότερο αποδοτική. Η μέθοδος Newton, όπως εφαρμόζεται για τη συγκεκριμένη συνάρτηση και τα δοσμένα αρχικά σημεία, δεν μπορεί να χρησιμοποιηθεί αποτελεσματικά, πιθανώς λόγω ιδιομορφιών στον εσσιανό πίνακα.

Αντίθετα, οι μέθοδοι Steepest Descent και Levenberg-Marquardt, όταν συνδυάζονται είτε με την ελαχιστοποίηση του βήματος μέσω διχοτόμησης είτε με τον κανόνα Armijo, αποδεικνύονται ιδιαίτερα αποδοτικές. Και οι δύο μέθοδοι συγκλίνουν με παρόμοιο αριθμό επαναλήψεων, ενώ υποστηρίζουν μεγαλύτερα βήματα χωρίς να αποσταθεροποιούνται.

Από άποψη χρόνου, ο Steepest Descent εμφανίζεται πιο γρήγορος στην εκτέλεση, δεδομένου ότι η απλότητά του τον καθιστά λιγότερο απαιτητικό σε υπολογιστικούς πόρους. Παράλληλα, η ανάγκη για μεγάλα βήματα δεν είναι κρίσιμη για τη σύγκλιση του.

Συνολικά, η επιλογή της μεθόδου και της στρατηγικής βήματος εξαρτάται από τη φύση της συνάρτησης και τις απαιτήσεις του προβλήματος, με τη μέθοδο Levenberg-Marquardt να υπερέχει σε πιο περίπλοκες περιπτώσεις και τη Steepest Descent να αποτελεί μια ταχύτερη εναλλακτική για λιγότερο απαιτητικές εφαρμογές.

4.2 Συνολικές Παρατηρήσεις

- Εγκλωβισμός σε Τοπικά Ακρότατα:
  - Εάν ο αλγόριθμος ξεκινήσει από ένα σημείο κοντά σε τοπικό ελάχιστο ή μέγιστο, υπάρχει μεγάλη πιθανότητα να "εγκλωβιστεί" σε αυτό, ειδικά αν η επιλεγμένη κατεύθυνση ή το βήμα δεν είναι επαρκώς προσαρμοσμένα. Αυτό είναι εμφανές στις περιπτώσεις σταθερού βήματος, όπου ο αλγόριθμος αδυνατεί να ξεπεράσει τοπικά ακρότατα.
- Εξάρτηση από Τιμή Εκκίνησης (x_0, y_0):
  - Η επιλογή του αρχικού σημείου παίζει καθοριστικό ρόλο στην επιτυχία του αλγορίθμου. Τα γραφήματα και οι υπολογισμοί δείχνουν ότι ορισμένα σημεία εκκίνησης οδηγούν σε γρήγορη σύγκλιση, ενώ άλλα απαιτούν περισσότερες επαναλήψεις ή οδηγούν σε αποτυχία σύγκλισης στο ολικό ελάχιστο.
- Επίδραση Επιλογής Βήματος γ_k:
  - Ο τύπος επιλογής του βήματος (σταθερό, ελαχιστοποίηση, κανόνας *Armijo*) επηρεάζει σε μεγάλο βαθμό τη συμπεριφορά του αλγορίθμου. Ο κανόνας Αρμιθο και η ελαχιστοποίηση μέσω διχοτόμησης φαίνεται να παρέχουν πιο σταθερά αποτελέσματα, ακόμα και όταν το σημείο εκκίνησης δεν είναι ιδανικό.

Βιβλιογραφία

Γεώργιος Α. Ροβιθάκης. *Τεχνικές Βελτιστοποίησης*. Εκδόσεις ΤΖΙΟΛΑ, 2007.