Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρονικής

Τεχνικές Βελτιστοποίησης - 3η Εργασία

Θεόδωρος Παπαφωτίου
ΑΕΜ: 9708
papafotit@ece.auth.gr

20 Δεκεμβρίου 2021

Περιεχόμενα

Κεφάλαιο 1

Περιγραφή 3ης Εργασίας

Στόχος της παρούσας εργασίας είναι η σύγκριση των μεθόδων Μέγιστης Καθόδου και Μέγιστης Καθόδου με Προβολή για την **εύρεση ελαχίστου** μια δοσμένης συνάρτησης πολλών μεταβλητών. Για την πρώτη μέθοδο δεν ισχύουν περιορισμοί στην συνάρτηση, σε αντίθεση με τη δεύτερη μέθοδο, όπου η συνάρτηση βρίσκεται υπό την **παρουσία περιορισμών**.

1.1 Μέθοδοι

Οι μέθοδοι που καλούμαστε να υλοποιήσουμε και να μελετήσουμε είναι οι παρακάτω:

- Μέθοδος Μέγιστης Καθόδου (Steepest Descent) (προηγούμενη εργασία)
- Μέθοδος Μέγιστης Καθόδου με Προβολή (Steepest Descent with Projection)

Η μέθοδος Μέγιστης Καθόδου χρησιμοποιείται για να συγκριθούν τα αποτελέσματα που προκύπτουν από τη δεύτερη μέθοδο. Η μελέτη της μεθόδου Μέγιστης Καθόδου με Προβολή γίνεται σε συνάρτηση με 2 μεταβλητές $f(x, y)$ με περιορισμούς. Η συνάρτηση f χρειάζεται αν είναι τουλάχιστον 1 φορά παραγωγίσιμη για τις 2 μεθόδους.

1.2 Υλοποίηση

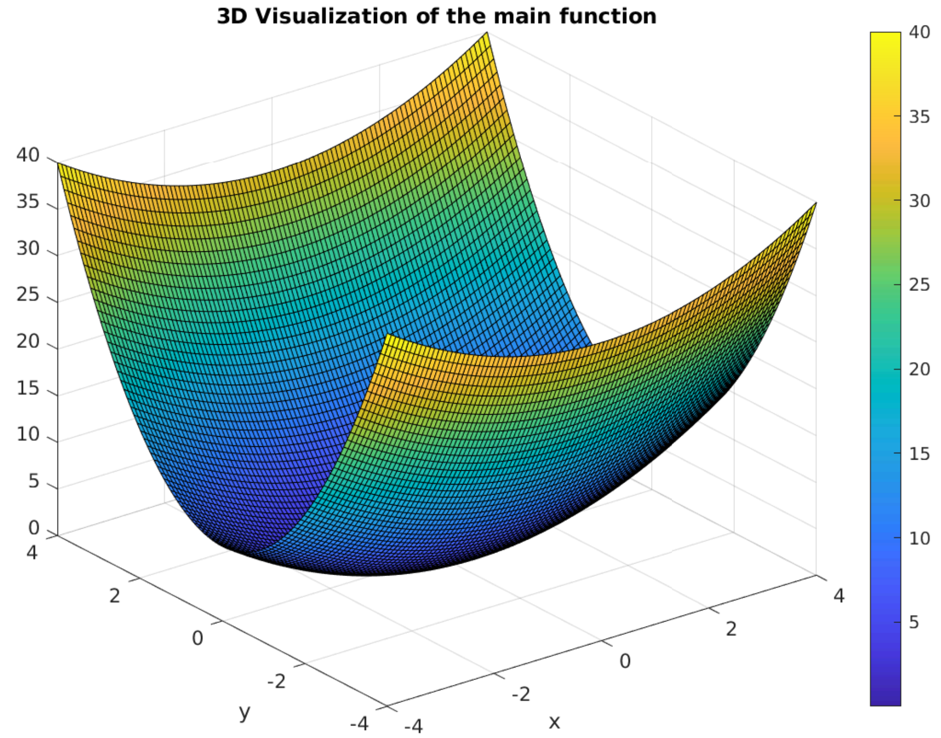
Η υλοποίηση των παραπάνω μεθόδων πραγματοποιήθηκε με τη χρήση του λογισμικού **MATLAB** (v.2021a). Για κάθε μέθοδο έχει υλοποιηθεί η αντίστοιχη συνάρτηση (function), καθώς και ένα αρχείο (Project3_i, $i = 1, 2, 3, 4$) που καλεί τη συνάρτηση αυτή, σύμφωνα με τα ερωτήματα της εργασίας. Το αρχείο Project3_0 αξιοποιείται για να οπτικοποιήσουμε τη συνάρτηση σε δισδιάστατη και τρισδιάστατη μορφή. Όλα τα αρχεία επισυνάπτονται με την παρούσα τεχνική αναφορά στο αποθετήριο της πλατφόρμας eLearning.

1.3 Μελέτη

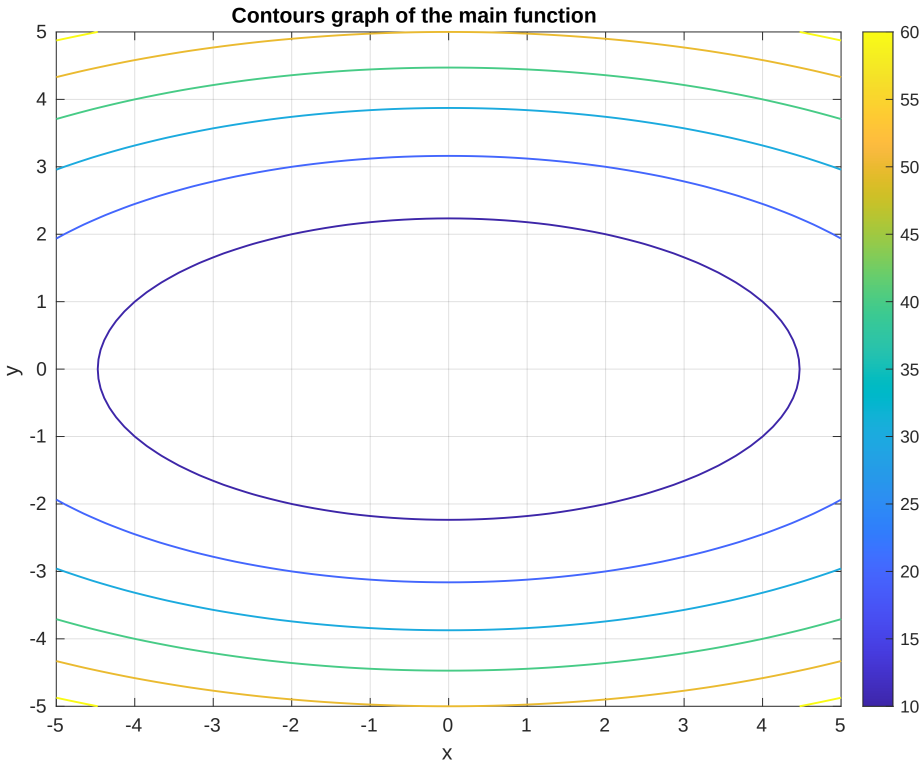
Η συνάρτηση προς ελαχιστοποίηση είναι η παρακάτω:

$$f(x, y) = \frac{1}{2} \cdot x^2 + 2 \cdot y^2, \quad x, y \in \Re \tag{1.1}$$

Παρακάτω μπορούμε να παρατηρήσουμε τη μορφή της συνάρτησης τόσο σε 3D μορφή, 1.1, όσο και μέσω κάποιων ισοβαρών καμπυλών της (2D μορφή), 1.2.



Σχήμα 1.1: 3D Visualisation of the function



Σχήμα 1.2: Contours graph of the function

Η συνάρτηση παρουσιάζει **ένα ελάχιστο** (ολικό), όπως μπορούμε να παρατηρήσουμε στα παραπάνω διαγράμματα. Συγκεκριμένα:

$$f_{\min} = f(0,0) = 0$$

Στην παρούσα εργασία καλούμαστε να αναζητήσουμε το ελάχιστο της παραπάνω συνάρτησης, χρησιμοποιώντας τις παραπάνω μεθόδους. Επίσης, μελετούμε τη συμπεριφορά των μεθόδων για διαφορετικές τιμές των παραμέτρων που χρησιμοποιούνται από την καθεμία. Η συγκεκριμένη συνάρτηση είναι **τουλάχιστον 1 φορά παραγωγίσιμη**, οπότε μπορούμε να εφαρμόσουμε και τις 2 μεθόδους.

Κεφάλαιο 2

Μελέτη Μεθόδων

Οι μέθοδοι που μελετούμε στην παρούσα εργασία βασίζονται στην ιδέα της επαναληπτικής καθόδου. Ξεκινάμε από κάποιο σημείο $x_0 \in \Re^n$ και παράγουμε διαδοχικά τα διανύσματα $x_1, x_2, x_3, \ldots$ έτσι ώστε:

$$f(x_{k+1}) < f(x_k), k = 0, 1, 2, \ldots$$

Ο αλγόριθμος μας οδηγεί σε βελτιωμένες τιμές της f προς ελαχιστοποίησή της. Η διαδικασία επιλογής των $x \in \Re^n$ συνοψίζεται από την παρακάτω αναδρομική σχέση:

$$x_{k+1} = xk + \gamma_k d_k, k = 0, 1, 2, \ldots$$

όπου k ο αριθμός της επανάληψης, $\gamma_k > 0$ το βήμα και d_k το διάνυσμα κατεύθυνσης, ο υπολογισμός του οποίου μεταβάλλεται μεταξύ των μεθόδων, ικανοποιώντας πάντα τη συνθήκη:

$$\nabla f^T(x_k) d_k < 0, k = 0, 1, 2, \ldots$$

Αξίζει να σημειωθεί εδώ ότι κατά τη μελέτη της μεθόδου Μέγιστης Καθόδου μπορούμε να μεταβάλουμε τη επιλογή του βήματος, με τις παρακάτω μεθόδους:

- $\gamma_k = \text{const}$
- $\gamma_k = \min(f(x_k + \gamma_k d_k))$, ως προς $\gamma_k > 0$
- Κανόνας Armijo

Ο κανόνας Armijo αποτελεί μια μέθοδο διαδοχικής μείωσης του γ_k, όπου το βήμα επιλέγεται ως:

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

όπου m_k είναι ο μικρότερος μη-αρνητικός ακέραιος που ικανοποιεί την παρακάτω σχέση:

$$f(x_k) - f(x_{k+1}) \geq -a\beta^{m_k} s {d_k}^T \nabla f(x_k)$$

Ο συντελεστής s δηλώνει το αρχικό βήμα, ενώ τα a, β τα προκαθορίζουμε. Για την παρούσα εργασία χρησιμοποιήθηκαν:

$$s = \gamma_0 = const \qquad a = 10^{-3} \qquad \beta = 0.1$$

Να σημειώσουμε εδώ ότι στη δεύτερη μέθοδο υπολογισμού του γ αξιοποιείται η μέθοδος **Golden Section** που υλοποιήθηκε στην προηγούμενη εργασία.

Στην εργασία αυτή ωστόσο θα αξιοποιηθεί **μόνο η πρώτη μέθοδος επιλογής βήματος**, γ_k = const.

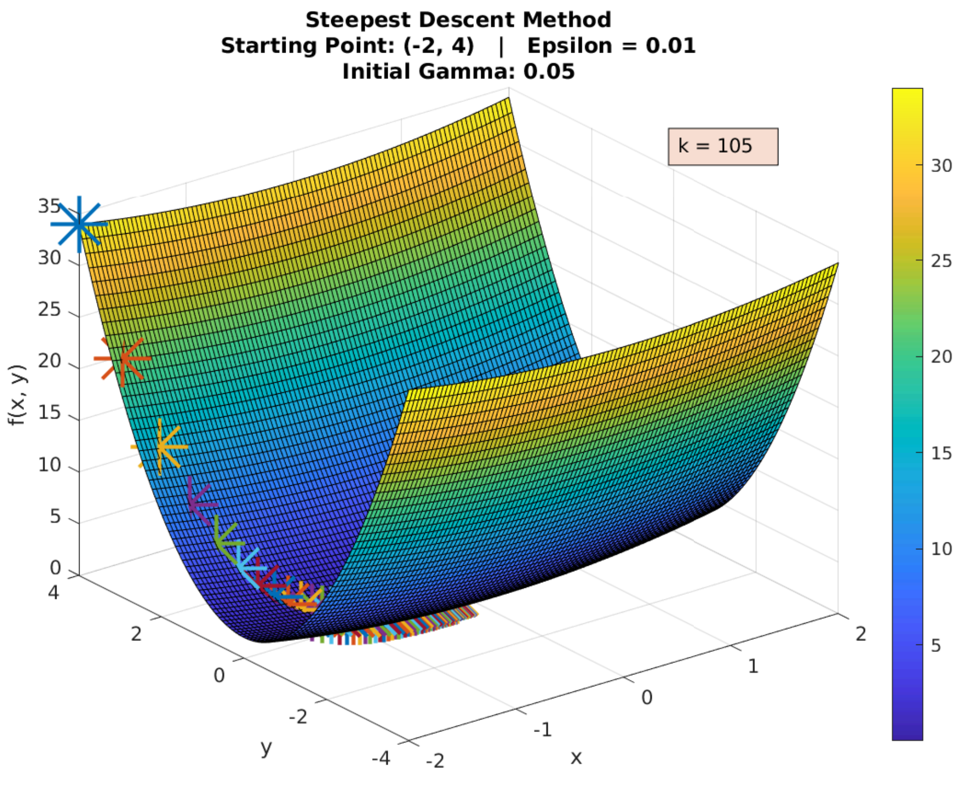
Τέλος, σε όλους τους αλγορίθμους χρησιμοποιήθηκε **μέγιστο όριο 1000 βημάτων** για τον υπολογισμό του ελαχίστου. Σε περίπτωση που ξεπερνάται αυτό το όριο, ο αλγόριθμος τερματίζει και μας επιστρέφει την τελευταία υπολογισμένη τιμή ελαχίστου.

2.1 Μέθοδος Μέγιστης Καθόδου

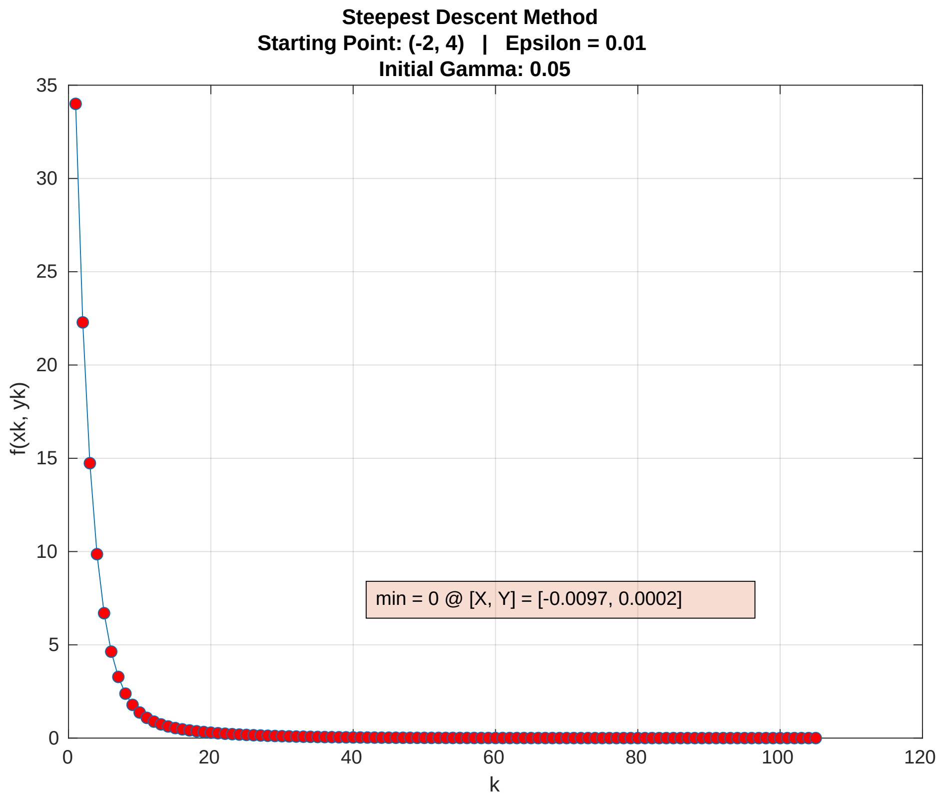
Εφόσον στην προηγούμενη εργασία είχαμε αναλύσει τη Μέθοδο Μέγιστης Καθόδου, στην παρούσα εργασία παραθέσουμε απλώς τα ζητούμενα διαγράμματα για τις περιπτώσεις διαφορετικών γ που ζητούνται.

2.1.1 Γραφήματα

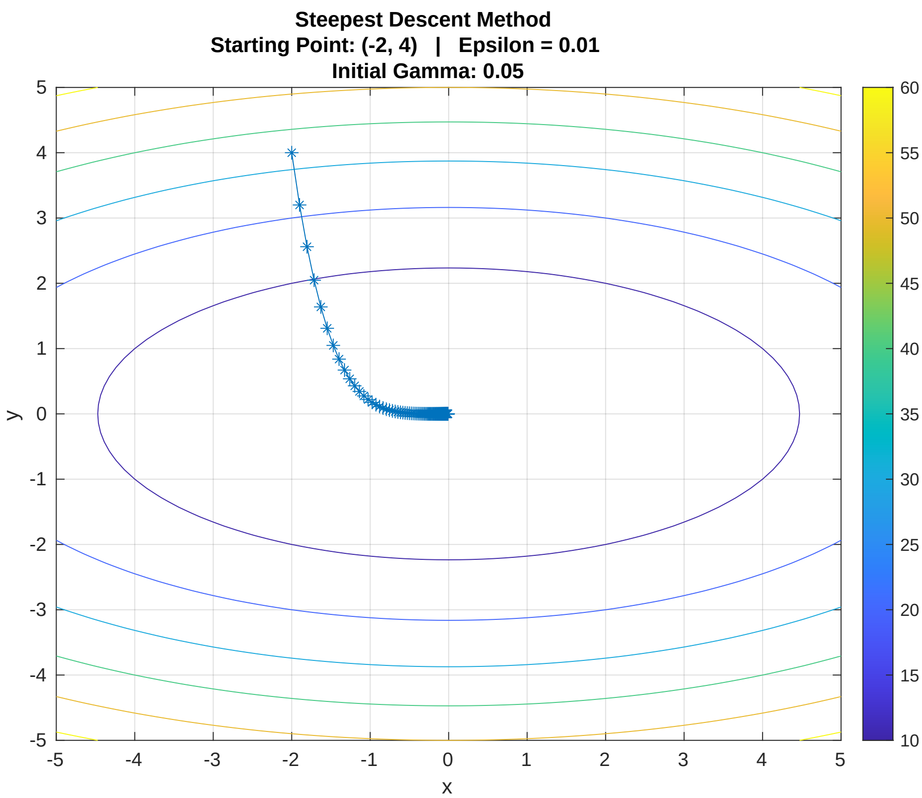
Σημείο εκκίνησης $[x_0, y_0] = [-2, 4]$**,** $\gamma_k = 0.05$**,** $\epsilon = 0.01$



Σχήμα 2.1: 3D visualisation

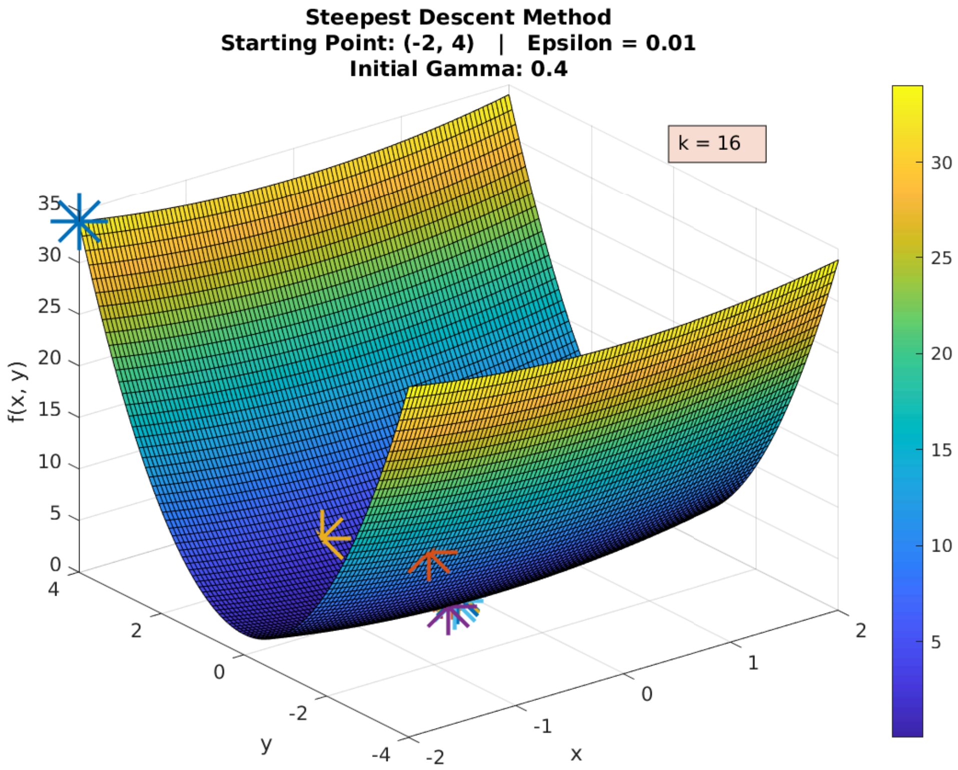


Σχήμα 2.2: 2D visualisation

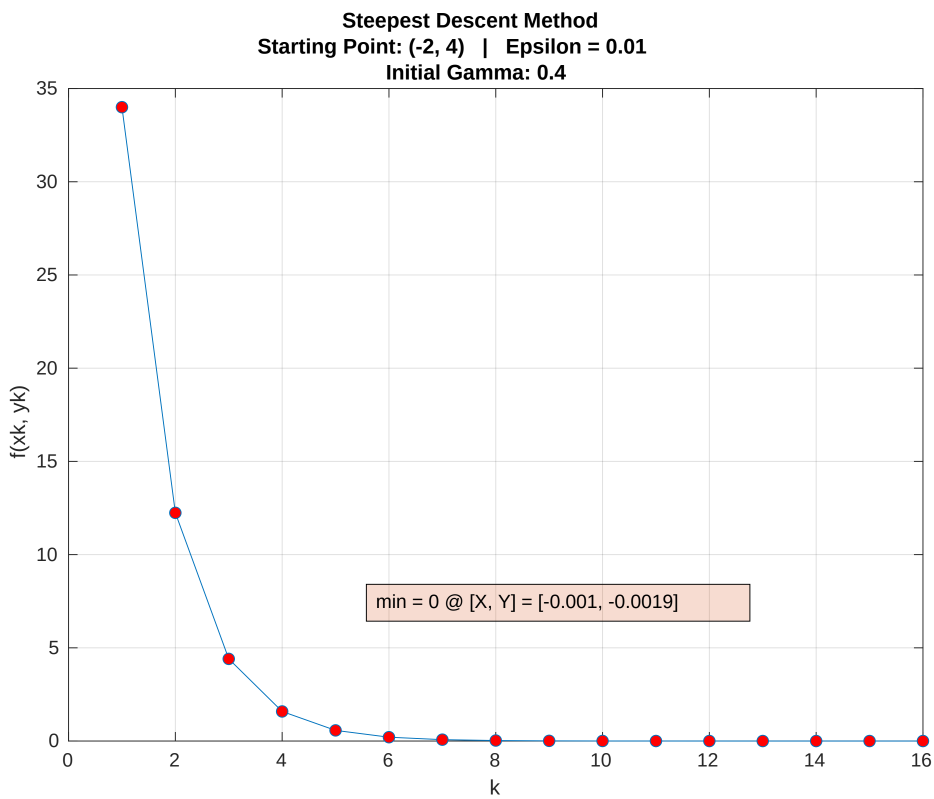


Σχήμα 2.3: Contours visualisation

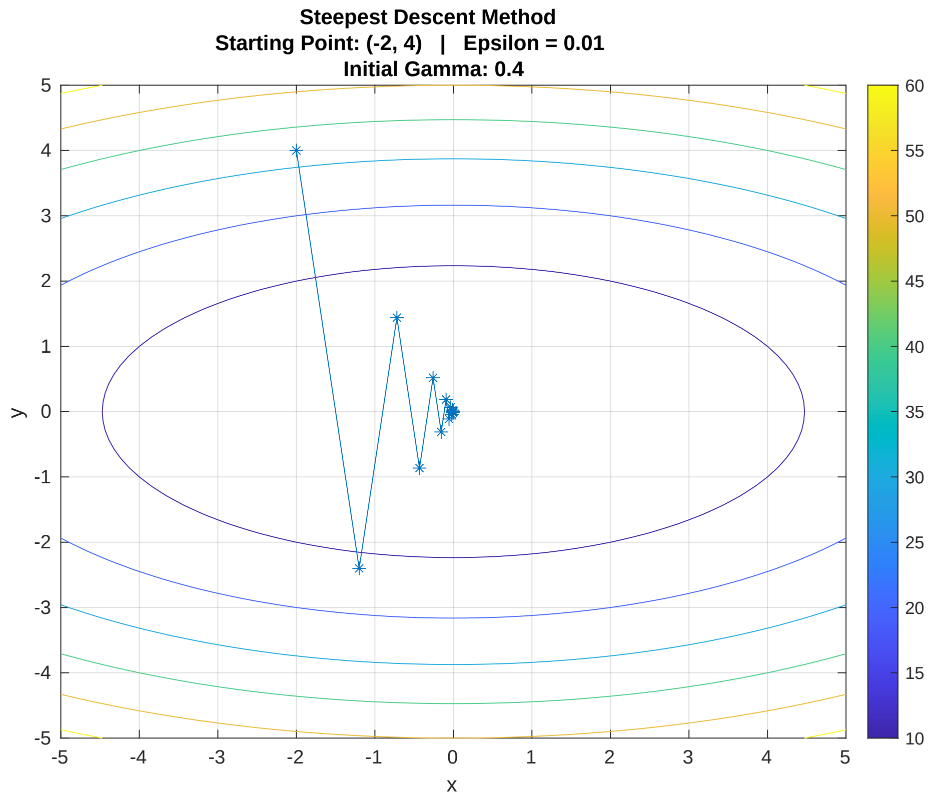
Σημείο εκκίνησης $[x_0, y_0] = [-2, 4]$, $\gamma_k = 0.4$, $\epsilon = 0.01$



Σχήμα 2.4: 3D visualisation

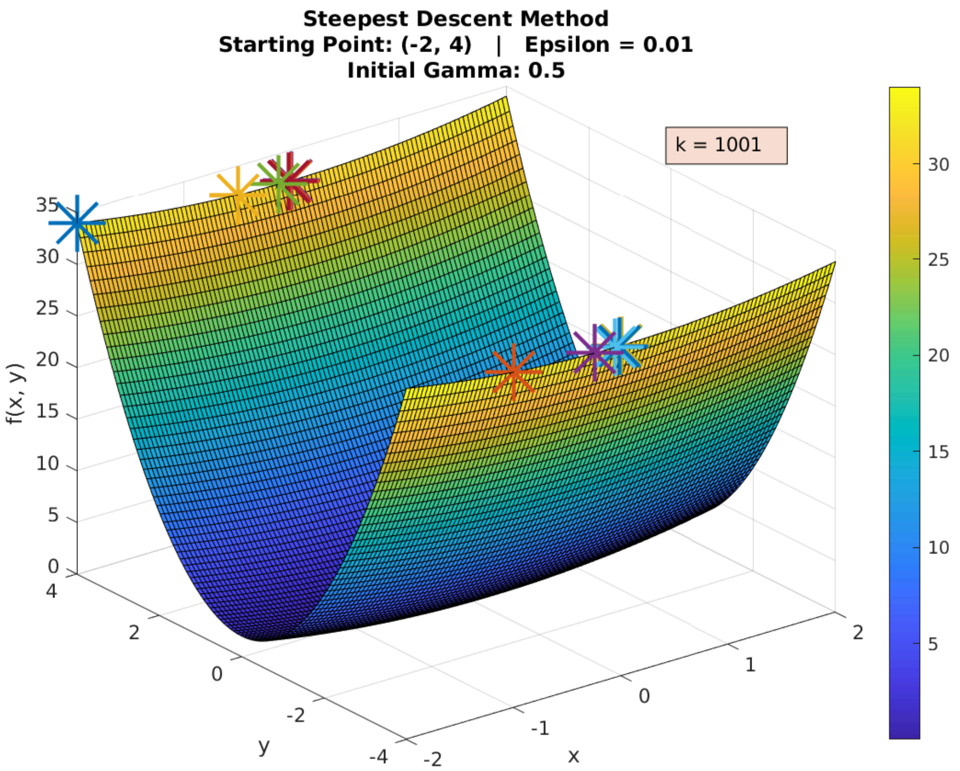


Σχήμα 2.5: 2D visualisation

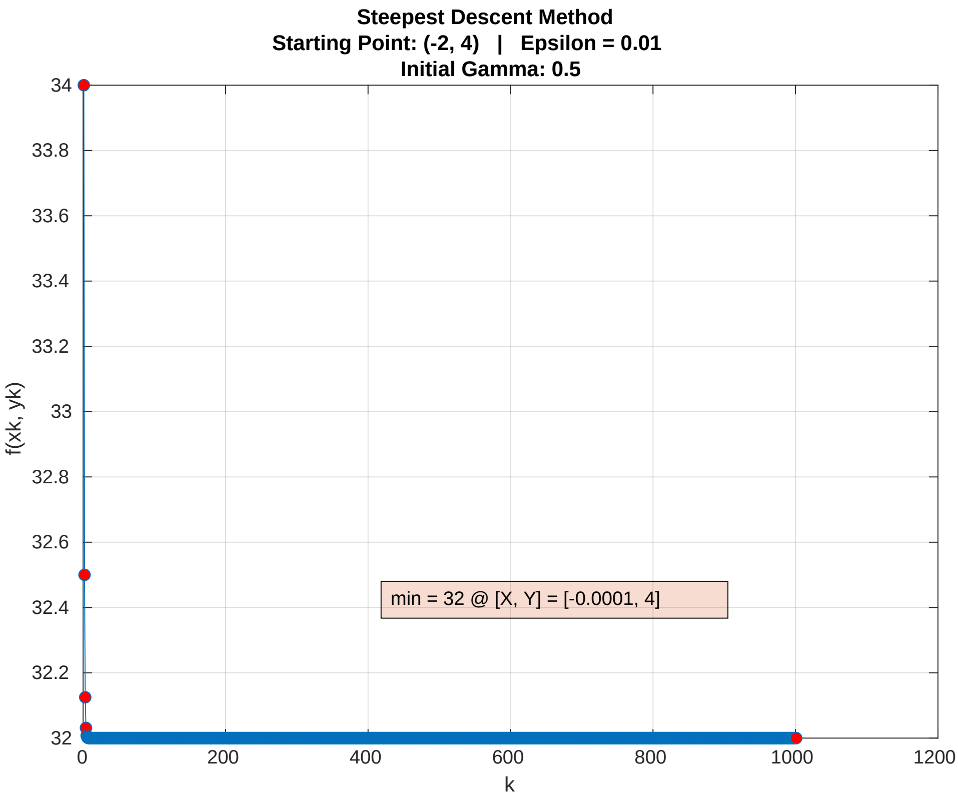


Σχήμα 2.6: Contours visualisation

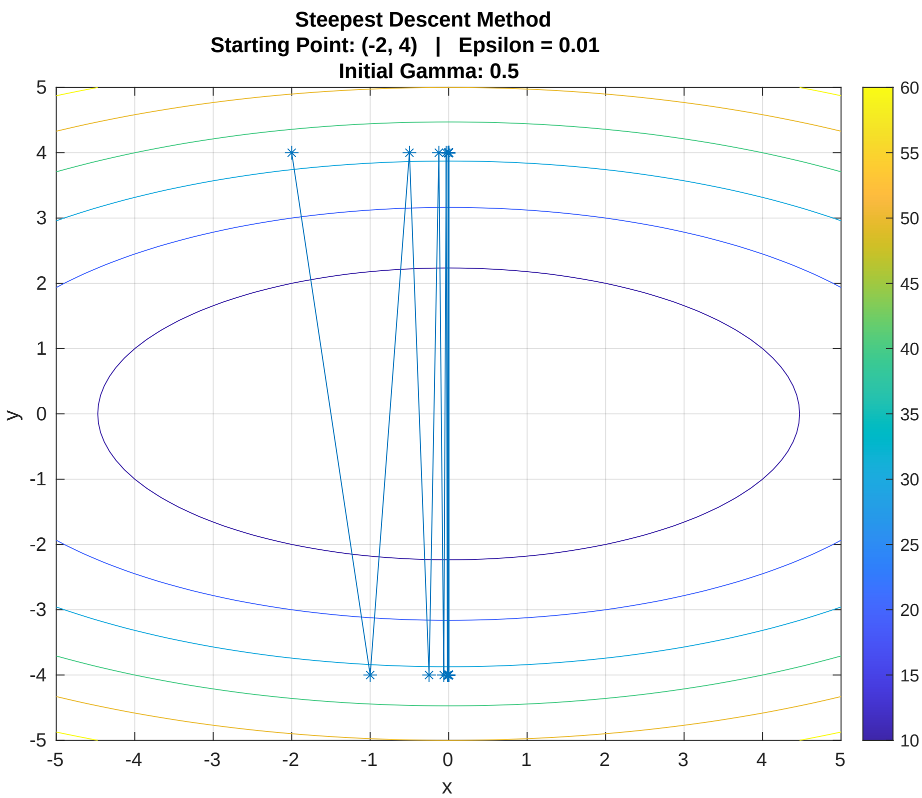
Σημείο εκκίνησης $[x_0, y_0] = [-2, 4]$, $\gamma_k = 0.5$, $\epsilon = 0.01$



Σχήμα 2.7: 3D visualisation



Σχήμα 2.8: 2D visualisation



Σχήμα 2.9: Contours visualisation

Σημείο εκκίνησης $[x_0, y_0] = [-2, 4]$, $\gamma_k = 2$, $\epsilon = 0.01$

Λόγω των αριθμών που προκύπτουν, τάξεως $> 10^{300}$*, το πρόγραμμα δεν δύναται να τυπώσει τα διαγράμματα*

Σημείο εκκίνησης $[x_0, y_0] = [-2, 4]$, $\gamma_k = 10$, $\epsilon = 0.01$

Λόγω των αριθμών που προκύπτουν, τάξεως $> 10^{300}$*, το πρόγραμμα δεν δύναται να τυπώσει τα διαγράμματα*

2.2 Μέθοδος Μέγιστης Καθόδου με Προβολή

2.2.1 Σύντομη περιγραφή της μεθόδου

Η συγκεκριμένη μέθοδος ανήκει στις μεθόδου ελαχιστοποίησης με **προβολή** (projection methods σε συγκεκριμένο μη-κενό, κλειστό και κυρτό **σύνολο X**, υποσύνολο του $\Re^n$. Η μέθοδος μέγιστης καθόδου με προβολή είναι **αλγόριθμος εφικτών σημείων** της μορφής:

$$x_{k+1} = x_k + \gamma_k(\bar{x}_k - x_k), \quad \gamma_k \in (0, 1] \tag{2.1}$$

όπου

$$\bar{x}_k = \mathrm{Pr}_X\{x_k - s_k \nabla f(x_k)\}, \quad s_k > 0 \tag{2.2}$$

Χρησιμοποιώντας την προβολή του $x_k - s_k \nabla f(x_k)$ στο X καταλήγουμε πάντα σε εφικτό σημείο x_{k+1}.

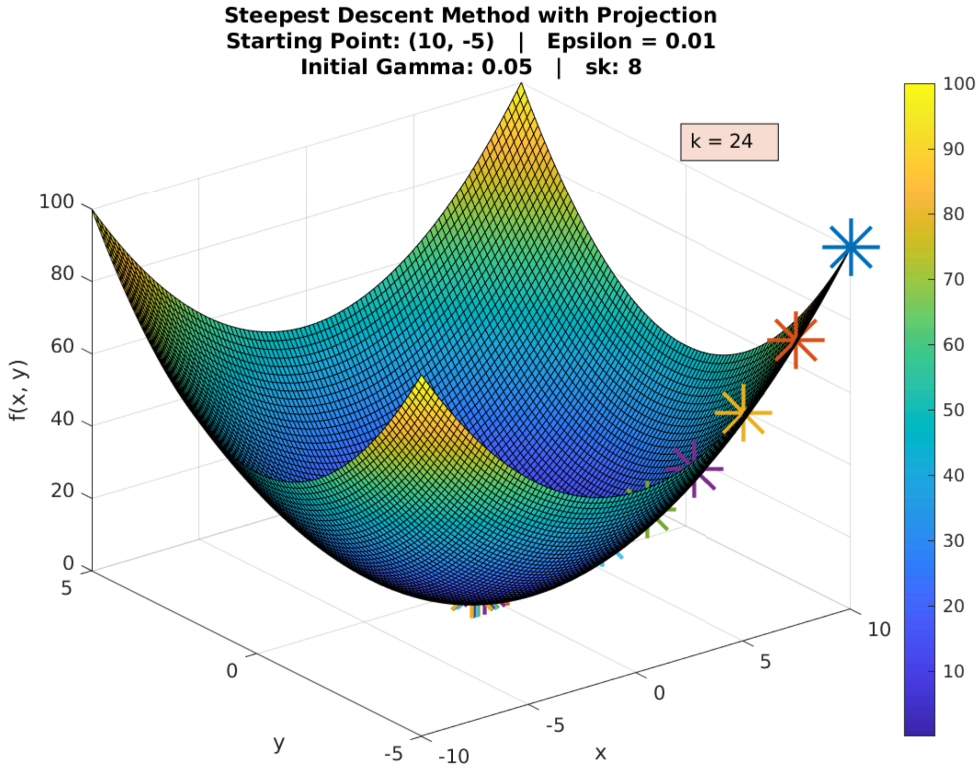
Ο αλγόριθμος ολοκληρώνεται αν και μόνο αν καταλήξει σε στάσιμο σημείο με ακρίβεια ϵ, δηλαδή όταν:

$$||x_k - \mathrm{Pr}_X(x_k - s_k \nabla f(x_k))|| \leq \epsilon \tag{2.3}$$

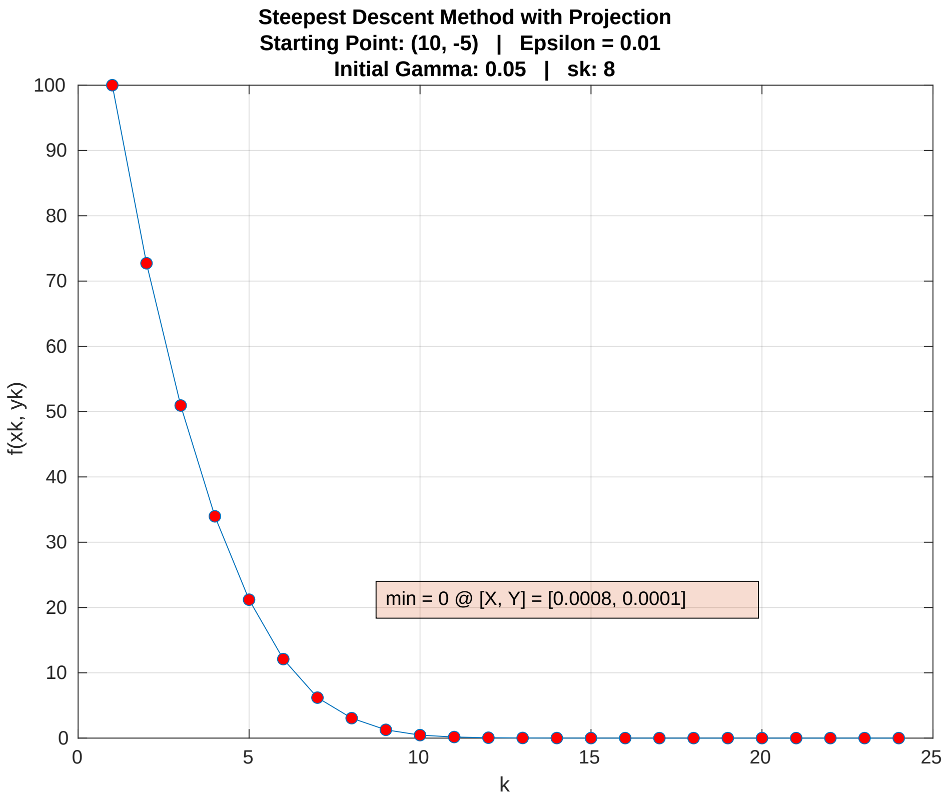
Ο αλγόριθμος βέβαια τερματίζει και όταν ξεπεράσει ορισμένο αριθμό βημάτων, ο οποίος για την παρούσα εργασία ισούται με 1000.

2.2.2 Γραφήματα

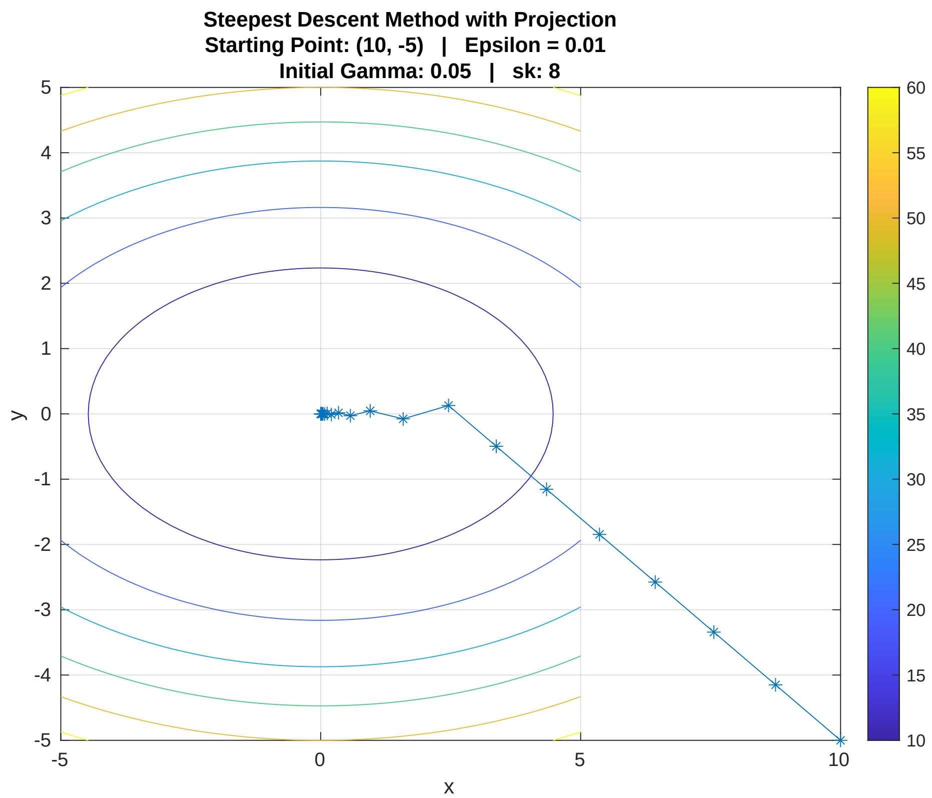
Σημείο εκκίνησης $[x_0, y_0] = [10, -5]$**,** $s_k = 8$**,** $\gamma_k = 0.05$**,** $\epsilon = 0.01$



Σχήμα 2.10: 3D visualisation

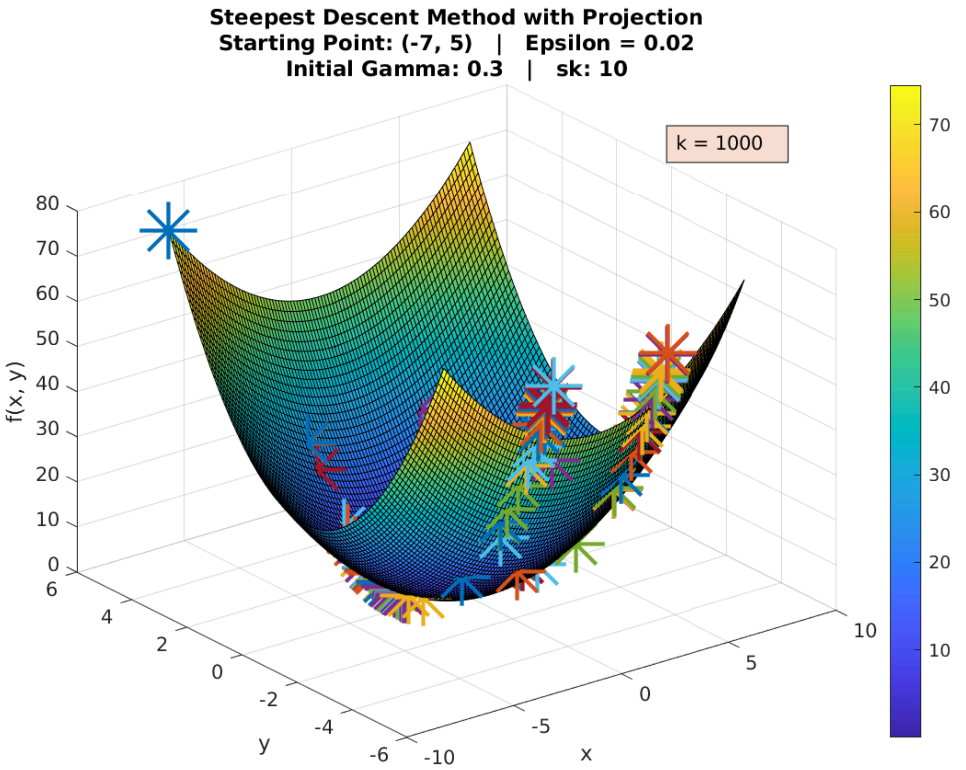


Σχήμα 2.11: 2D visualisation

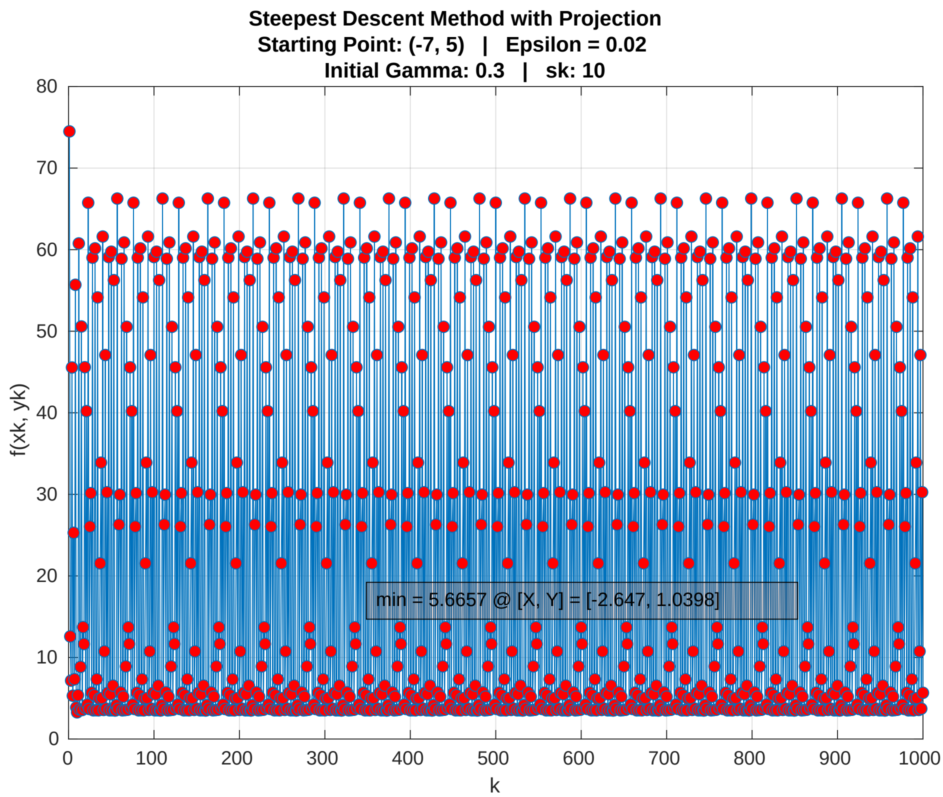


Σχήμα 2.12: Contours visualisation

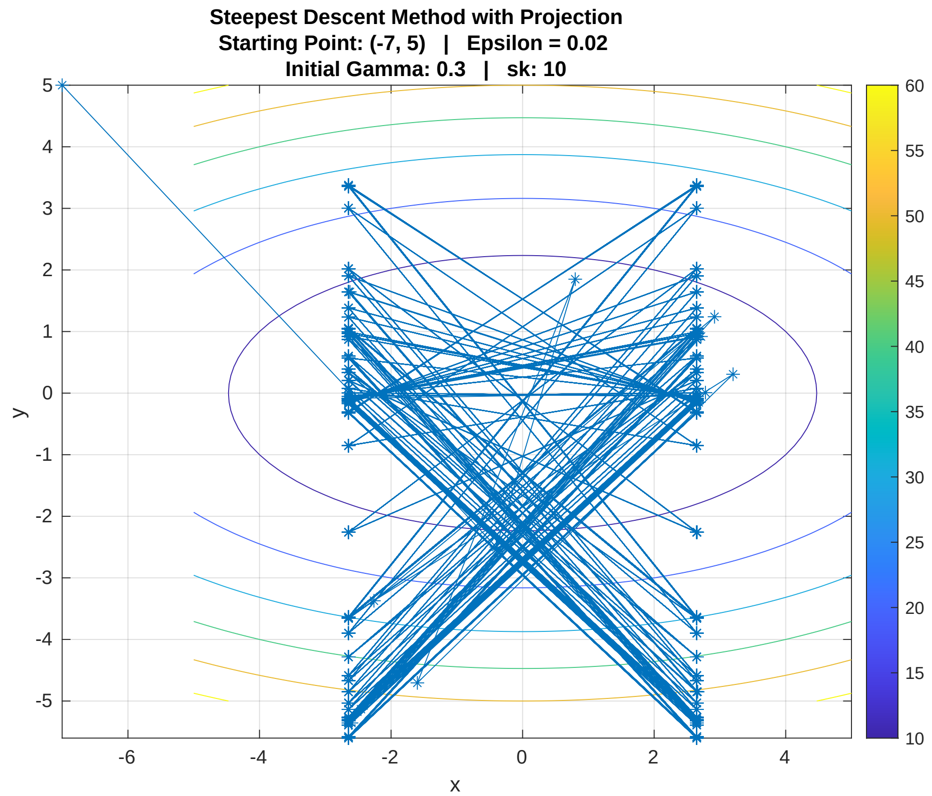
Σημείο εκκίνησης $[x_0, y_0] = [-7, 5]$, $s_k = 10$, $\gamma_k = 0.3$, $\epsilon = 0.02$



Σχήμα 2.13: 3D visualisation

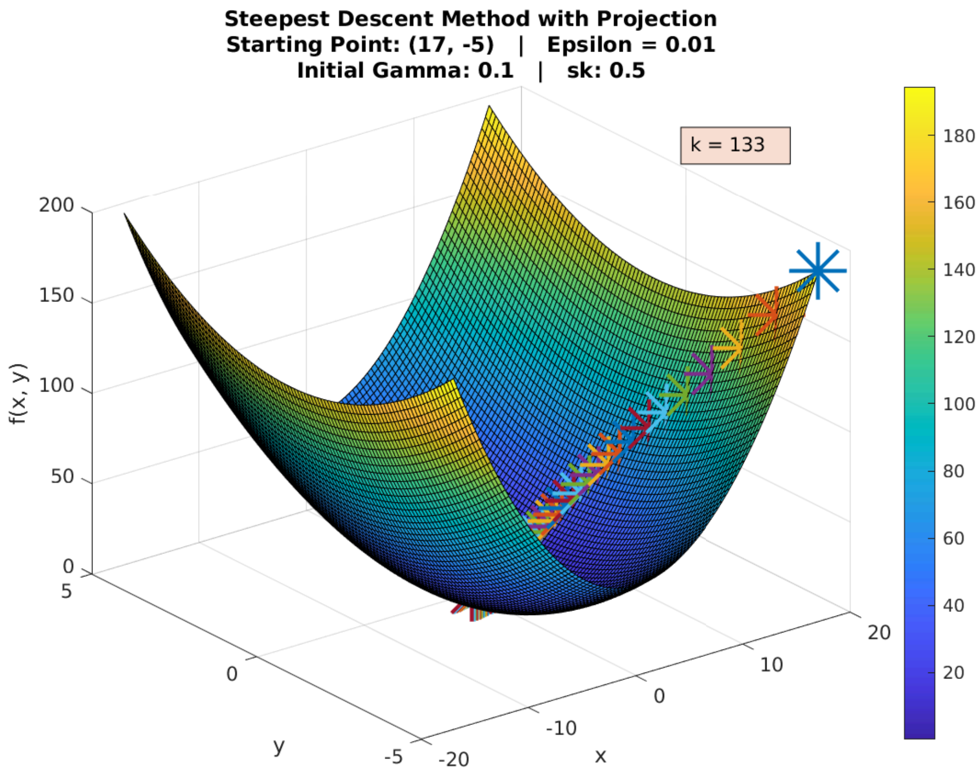


Σχήμα 2.14: 2D visualisation

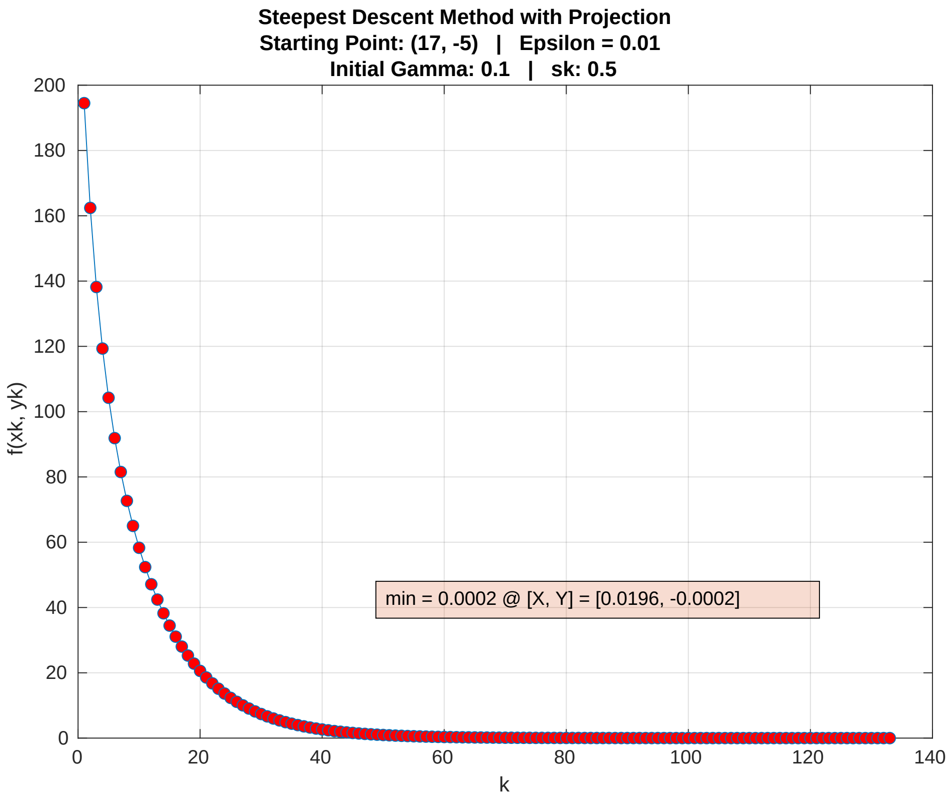


Σχήμα 2.15: Contours visualisation

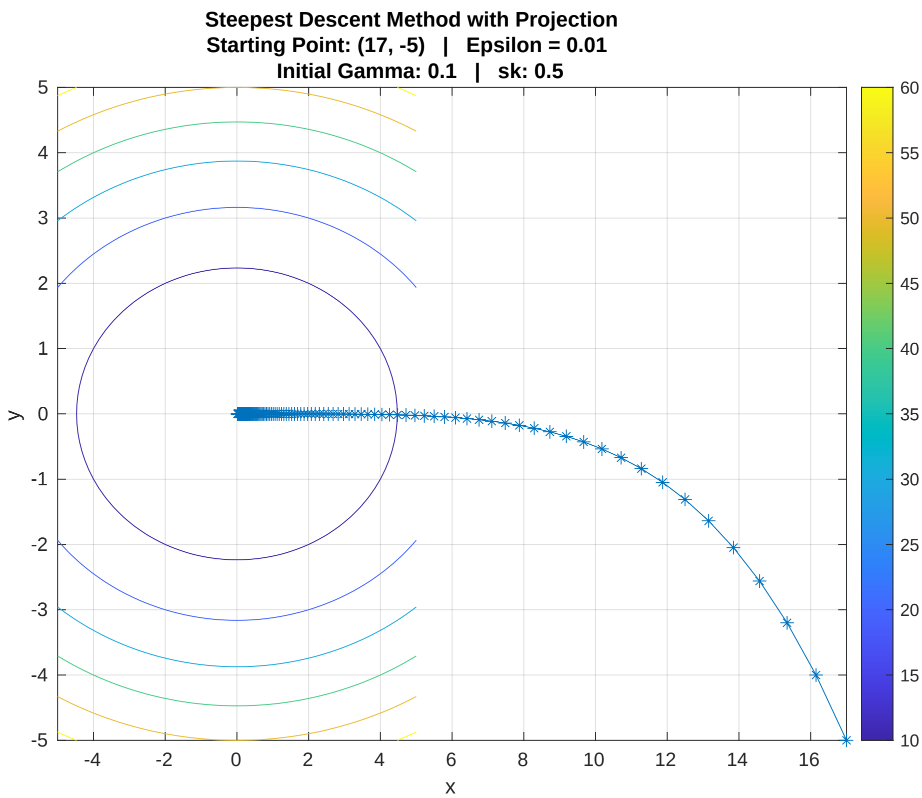
Σημείο εκκίνησης $[x_0, y_0] = [17, -5]$, $s_k = 0.5$, $\gamma_k = 0.1$, $\epsilon = 0.01$



Σχήμα 2.16: 3D visualisation



Σχήμα 2.17: 2D visualisation



Σχήμα 2.18: Contours visualisation

Κεφάλαιο 3

Παρατηρήσεις - Απαντήσεις Ερωτημάτων

3.1 Μέθοδος Μέγιστης Καθόδου

Για τη συγκεκριμένη μέθοδο ισχύει ότι

$$x_{k+1} = x_k - \gamma_k \cdot \nabla f(x_k) \tag{3.1}$$

Γνωρίζοντας τον τύπο της f, υπολογίζουμε ότι

$$\nabla f(x_k) = [x_1, 4x_2], \quad x_1, x_2 \in \Re \tag{3.2}$$

οπότε προκύπτει ότι

$$x_{k+1} = xk - \gamma_k \cdot [x_1, 4x_2]^T \tag{3.3}$$

Οπότε, για παράδειγμα στην περίπτωση όπου $\gamma_k = 0.05$ έχουμε ότι

$$x_{k+1} = (x_k, y_k) - 0.05 \cdot (x_k, y_k) = (0.95x_k, 0.8y_k) \tag{3.4}$$

οπότε για $k \to \infty$ ισχύει

$$\lim_{x \to \infty} 0.95^k \cdot |x_k| = 0 \tag{3.5}$$

δηλαδή ο αλγόριθμος βρίσκει το ελάχιστο.

Γενικότερα:

$$x_{k+1} = (x_k, y_k) - \gamma_k(x_k, 4y_k) = [(1 - \gamma_k)x_k, (1 - 4\gamma_k)y_k] \tag{3.6}$$

αντίστοιχα:

$$x_{k+2} = [(1 - \gamma_k)x_{k+1}, (1 - 4\gamma_k)y_{k+1}] = [(1 - \gamma_k)^2 x_k, (1 - 4\gamma_k)^2 y_k] \tag{3.7}$$

οπότε γενικά:

$$x_{k+N} = [(1 - \gamma_k)^N x_k, (1 - 4\gamma_k)^N y_k] \tag{3.8}$$

Άρα για $N \to \infty$ προκύπτει ότι:

$$x_k : \lim_{x \to \infty}(1 - \gamma_k)^N x_k = \begin{cases} 0 & 0 < \gamma_k < 2 \\ \infty & \gamma_k > 2 \end{cases} \tag{3.9}$$

$$y_k : \lim_{x\to\infty}(1-4\gamma_k)^N x_k = \begin{cases} 0 & 0 < \gamma_k < 0.5 \\ \infty & \gamma_k > 0.5 \end{cases} \tag{3.10}$$

Οπότε συμπεραίνουμε ότι ο αλγόριθμος συγκλίνει στο ελάχιστο όταν **συναληθεύουν** και 2 ανισότητες, δηλαδή όταν:

$$\gamma_k \in (0, 0.5) \tag{3.11}$$

Όταν $\gamma_k = 0.5$ θα έπρεπε να παρατηρούμε **ταλάντωση** μεταξύ κάποιων πεπερασμένων σημείων της συνάρτησης, ενώ για $\gamma_k > 0.5$ ο αλγόριθμος **αποκλίνει** και κινείται προς το ∞.

Πράγματι, μπορούμε να παρατηρήσουμε από τα διαγράμματα 2.3, 2.6 και τα αντίστοιχα 3D plots 2.1, 2.4 ότι ο αλγόριθμος καταλήγει στο ελάχιστο μετά από **πεπερασμένο αριθμό επαναλήψεων**. Στην περίπτωση μάλιστα όπου $\gamma_k = 0.4$ είναι εμφανές στο διάγραμμα 2.6 και το **ζικ-ζακ** που εμφανίζεται κατά την εκτέλεση του αλγορίθμου από το πρώτο σημείο μέχρι το ελάχιστο (αυτός είναι και ο λόγος που προστέθηκε και αυτή η τιμή του γ_k πέρα από τις 4 τιμές που ζητήθηκαν από την εκφώνηση).

Στην περίπτωση όπου $\gamma_k = 0.5$ παρατηρούμε στα διαγράμματα 2.8, 2.7, 2.9 μια χαρακτηριστική ταλάντωση γύρω από την τιμή $f(x_k, y_k) = 32$, οπότε ο αλγόριθμος τερματίζει μετά το πέρας 1000 επαναλήψεων που ορίστηκαν ως ο μέγιστος αριθμός επαναλήψεων.

Τέλος, στις περιπτώσεις όπου $\gamma_k = 2$ και $\gamma_k = 10$ δεν μπορούσαν να εξαχθούν διαγράμματα από το MATLAB, καθώς ο αλγόριθμος έφτανε σε τιμές τις συνάρτησης τάξεως μεγαλύτερης από 10^{300}, έτεινε δηλαδή στο άπειρο.

Να αναφέρουμε ότι για όλα τα διαγράμματα που εξήχθησαν από τη Μέθοδο Μέγιστης Καθόδου χρησιμοποιήθηκε αρχικό σημείο εκκίνησης $[x_0, y_0] = [-2, 4]$ και $\epsilon = 0.01$.

3.2 Μέθοδος Μέγιστης Καθόδου με Προβολή

Για τη συγκεκριμένη μέθοδο θεωρήθηκαν στη συνάρτηση περιορισμοί:

$$-15 \le x_1 \le 15 \quad \text{και} \quad -20 \le x_2 \le 12 \tag{3.12}$$

Αυτοί οι περιορισμοί ορίζουν πρακτικά το μη-κενό, κλειστό, κυρτό **σύνολο X**, μέσα στο οποίο κινείται ο αλγόριθμος προς εύρεση του ελαχίστου.

Εφόσον οι περιορισμοί εκφράζονται ως φράγματα, ορίζουμε τις προβολές των x_1, x_2 στο X ως εξής:

$$[Pr_X\{x\}]_1 = \begin{cases} -15, & x_1 \le -15 \\ x_1, & -15 < x_1 < 15 \\ 15, & x_1 \ge 15 \end{cases} \qquad [Pr_X\{x\}]_2 = \begin{cases} -20, & x_1 \le -20 \\ x_2, & -20 < x_2 < 12 \\ 12, & x_1 \ge 12 \end{cases} \tag{3.13}$$

Από τα παραπάνω συμπεραίνουμε ότι αν το x_k είναι εφικτό σημείο:

$$\bar{x}_k = \mathrm{Pr}_X\{x_k - s_k \cdot \nabla f(x_k)\} = x_k - s_k \cdot \nabla f(x_k) \tag{3.14}$$

οπότε από την εξίσωση 2.1 συνάγεται ότι:

$$x_{k+1} = x_k - \gamma_k \cdot s_k \cdot \nabla f(x_k) \tag{3.15}$$

Οπότε πρακτικά ο αλγόριθμος για x_k εφικτά σημεία μετατρέπεται σε αλγόριθμο μέγιστης καθόδου με $\gamma_k' = \gamma_k \cdot s_k$ και ισχύουν οι ίδιοι περιορισμοί που αποδείχθηκαν παραπάνω για το γ, δηλαδή:

$$\gamma_k' \in (0, 0.5) \tag{3.16}$$

3.2.1 Θέμα 3.2

$$\gamma_k = 0.05, \quad s_k = 8, \quad [x_0, y_0] = [10, -5]$$

Παρατηρούμε ότι $\gamma_k' = 0.05 \cdot 8 = 0.4 < 0.5$, επομένως ο αλγόριθμος αναμένεται να τερματίσει και να **συγκλίνει προς το ελάχιστο**.

Πράγματι, στα διαγράμματα 2.10, 2.10, 2.12 παρατηρούμε ότι ο αλγόριθμος καταλήγει στο ελάχιστο και μάλιστα σε **λιγότερα βήματα** συγκριτικά με την απλή μέθοδο μέγιστης καθόδου με $\gamma_k = 0.05$, όπως φαίνεται στο 2.1 (k' = 24 έναντι k = 105).

3.2.2 Θέμα 3.3

$$\gamma_k = 0.3, \quad s_k = 10, \quad [x_0, y_0] = [-7, 5]$$

Παρατηρούμε ότι $\gamma_k' = 0.3 \cdot 10 = 3 > 0.5$, επομένως ο αλγόριθμος αναμένεται να είναι **ασταθής**.

Πράγματι, στα διαγράμματα 2.13, 2.13, 2.15 παρατηρούμε ότι ο αλγόριθμος είναι ασταθής και πραγματοποιεί συνεχόμενες **ταλαντώσεις** μέχρι τον τερματισμό του μετά το πέρας των 1000 βημάτων. Είναι χαρακτηριστικό, ωστόσο, ότι σε αντίθεση με την απλή μέθοδο μέγιστης καθόδου, ο αλγόριθμος για τιμές του $\gamma_k' > 0.5$ δεν οδηγείται στο άπειρο, απλώς ταλαντώνεται προσπαθώντας να μείνει εντός του συνόλου X. Αυτή είναι και η **διόρθωση που ¨επιβάλλει¨ η προβολή στο X**. Στις αντίστοιχες περιπτώσεις της απλής μεθόδου μέγιστης καθόδου, όπου $gamma_k = 2$ ή $gamma_k = 10$ παρατηρούμε ότι ο αλγόριθμος αποκλίνει κινούμενος προς το άπειρο.

Ένας πρακτικός τρόπος να συγκλίνει το συγκεκριμένο παράδειγμα στο ελάχιστο, είναι να **μειώσουμε** την παράμετρο s_k στην τιμή $s_k = 1$, για παράδειγμα, ώστε $\gamma_k' = 0.3 \cdot 1 = 0.3 < 0.5$.

3.2.3 Θέμα 3.4

$$\gamma_k = 0.1, \quad s_k = 0.5, \quad [x_0, y_0] = [17, -5]$$

Με αντίστοιχο τρόπο βλέπουμε ότι $\gamma_k' = 0.1 \cdot 0.5 = 0.05 << 0.5$, επομένως ο αλγόριθμος αναμένεται να τερματίσει και να συγκλίνει προς το ελάχιστο, ωστόσο επειδή το γ_k' είναι **πιο κοντά στο 0**, παρά στο 0.5, αναμένουμε η σύγκλιση να γίνει πολύ **πιο αργά** σε σχέση με την πρώτη περίπτωση όπου $\gamma_k' = 0.4$.

Πράγματι, στα διαγράμματα 2.16, 2.16, 2.18 παρατηρούμε ότι ο αλγόριθμος καταλήγει σε ελάχιστο, αλλά με πολύ αργό ρυθμό, άρα πολύ περισσότερα βήματα.

Πρέπει να τονίσουμε βέβαια ότι το σημείο εκκίνησης βρίσκεται **εκτός του συνόλου X**, επομένως με την πρώτη επανάληψη το σημείο τοποθετείται πάνω στο **όριο του συνόλου X**.

3.2.4 Γενικά συμπεράσματα

Από τη μελέτη που διεξήχθη στην παρούσα εργασία διαπιστώνουμε ότι ο αλγόριθμος μέγιστης καθόδου με προβολή μπορεί να συγκλίνει στο ελάχιστο μιας συνάρτησης (με περιορισμούς) σε λιγότερα βήματα από ότι ο απλός αλγόριθμος μέγιστης καθόδου. Ωστόσο χρειάζεται να γίνει **σωστή επιλογή** των παραμέτρων s_k και γ_k ώστε να μην παρατηρηθεί αστάθεια/ταλάντωση. Το γινόμενό τους πρέπει να ικανοποιεί **τυχόν περιορισμούς** που θέτει η συνάρτηση, ώστε να έχουμε σύγκλιση, και θα πρέπει να γνωρίζουμε το σύνολο X μέσα στο οποίο θα αναζητήσουμε το ελάχιστο της συνάρτησης.