ΜΕΘΟΔΟΙ ΜΕΓΙΣΤΗΣ ΚΑΘΟΔΟΥ ΜΕ ΠΡΟΒΟΛΗ

Γεώργιος Γκιουλέκας – ΑΕΜ 10636

ΤΕΧΝΙΚΕΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ Άσκηση 3η

Περιεχόμενα

Εισαγωγή 2
Η αντικειμενική συνάρτηση 2
Το πρόβλημα χωρίς περιορισμούς 3
Η μαθηματική ανάλυση της σύγκλισης 3
Αποτελέσματα της μεθόδου για διάφορα γ 3
Το πρόβλημα με περιορισμούς 8
Μέθοδος μέγιστης καθόδου με προβολή 8
Για s=5, γ=0.5, αρχικό σημείο (5,-5) 9
Για s=15, γ=0.1, αρχικό σημείο (-5,10) 10
Για s=0.1, γ=0.2, αρχικό σημείο (8,-10) 12

Εισαγωγή

Στον φάκελο που βρίσκεται αυτή η αναφορά υπάρχει ένας φάκελος Project με τον κώδικα του project και ένας φάκελος Figures με τα διαγράμματα που υπάρχουν και στην παρούσα αναφορά, χωρισμένα σε επιμέρους φακέλους ανά θέμα. Στον φάκελο του project υπάρχει ένα αρχείο **main.m** το οποίο κάνει αρχικοποίηση της αντικειμενικής συνάρτησης και του διανύσματος κλίσης, καθώς και των παραμέτρων που χρησιμοποιούνται σε κάθε θέμα. Μέσα από αυτό το αρχείο μπορεί κανείς να τρέξει το αντίστοιχο section για κάθε θέμα ξεχωριστά. Απαραίτητη προϋπόθεση για να λειτουργήσει σωστά το πρόγραμμα είναι **να έχει εκτελεστεί το πρώτο section αρχικοποίησης στο main.m** . Έπειτα υπάρχουν 3 αρχεία .m με την υλοποίηση της μεθόδου Μέγιστης Καθόδου και Μέγιστης Καθόδου με Προβολή, την βοηθητική συνάρτηση που υλοποιεί την Προβολή, καθώς και άλλο 1 αρχείο grapher.m το οποίο δημιουργεί τα γραφήματα.

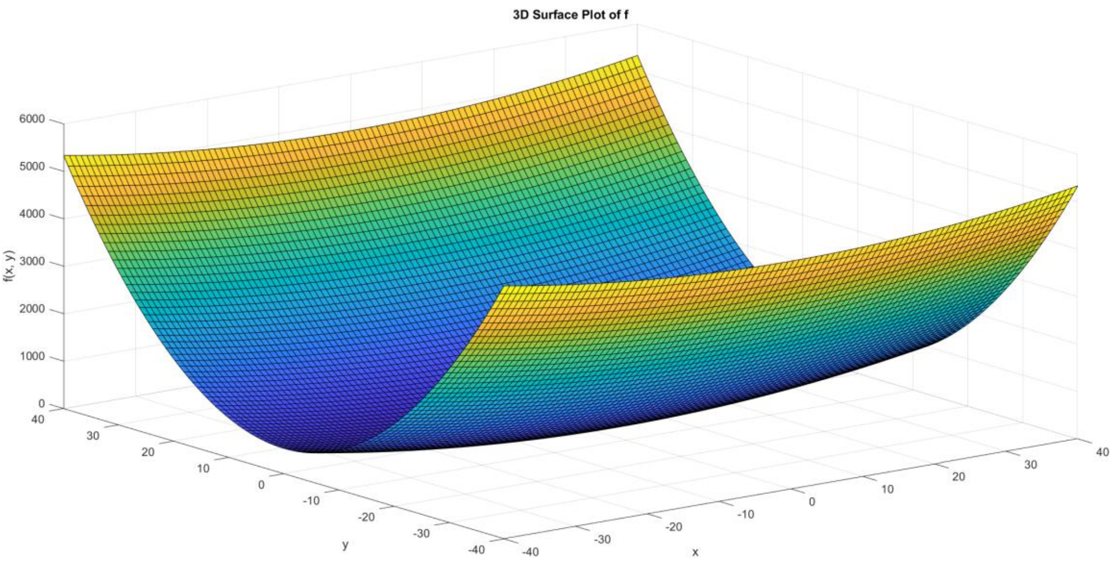
Τέλος, όλα τα .m αρχεία έχουν αναλυτικές εξηγήσεις σε σχόλια, για το τι κάνει ο κώδικας σε κάθε γραμμή, πώς λειτουργεί κάθε συνάρτηση, τι ορίσματα δέχεται, τι επιστρέφει και πολλές άλλες πληροφορίες.

Η αντικειμενική συνάρτηση

Η αντικειμενική συνάρτηση που καλούμαστε να ελαχιστοποιήσουμε είναι η

$$f(x,y) = \frac{1}{3}x^2 + 3y^2$$

Κάνουμε στο MATLAB το διάγραμμα της επιφάνειάς της για να έχουμε μια διαίσθηση της εικόνας της προτού προχωρήσουμε παρακάτω:



Το πρόβλημα χωρίς περιορισμούς

Θα χρησιμοποιήσουμε την Μέθοδο Μέγιστης Καθόδου από την προηγούμενη εργασία για να ελαχιστοποιήσουμε την συνάρτηση $f(x,y) = \frac{1}{3}x^2 + 3y^2,\ (x,y) \in \mathbb{R}^2$ χωρίς περιορισμούς, με ακρίβεια $\varepsilon = 0.001$ και θα εξετάσουμε τι συμβαίνει για διαφορετικά σταθερά βήματα γ_k. Θα κάνουμε επίσης την μαθηματική ανάλυση για την συνθήκη σύγκλισης της μεθόδου με σταθερό βήμα, ώστε να βρούμε για ποιες τιμές του γ_k η μέθοδος συγκλίνει και να επιβεβαιώσουμε έτσι τα αποτελέσματα που βγάλαμε με το MATLAB.

Η μαθηματική ανάλυση της σύγκλισης

Στο εξής θα ισχύει $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$ για συντομία συμβολισμών.

Ισχύει ότι $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x}$ όπου $Q = \begin{bmatrix} \frac{2}{3} & 0 \\ 0 & 6 \end{bmatrix}$, δηλαδή ο Q είναι ο Εσσιανός πίνακας της f. Οπότε προφανώς $\nabla f(\boldsymbol{x}) = Q\boldsymbol{x}$.

Θα δείξουμε πως το διάνυσμα κλίσης της f είναι συνεχές κατά Lipschitz.

$$\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| = \|Q\boldsymbol{x} - Q\boldsymbol{y}\| \leq \|Q\|\,\|\boldsymbol{x} - \boldsymbol{y}\| = L\|\boldsymbol{x} - \boldsymbol{y}\|$$

Όπου L είναι η σταθερά Lipschitz και είναι $L = \lambda_{max}(Q)$ αφού ο Q είναι συμμετρικός και θετικά ορισμένος πίνακας. Δηλαδή η f έχει συνεχές κατά Lipschitz διάνυσμα κλίσης με σταθερά $L = \lambda_{\max}(Q) = 6$.

Η επαναληπτική σχέση που χρησιμοποιεί η μέθοδος είναι η $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \gamma\nabla f(\boldsymbol{x}_k)$

Από το λήμμα καθόδου (βιβλίο σελ. 18 Λήμμα 2.3.3) προκύπτει ότι

$$f(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{x}_k) + \nabla f^T(\boldsymbol{x}_k)(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) + \frac{L}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2$$

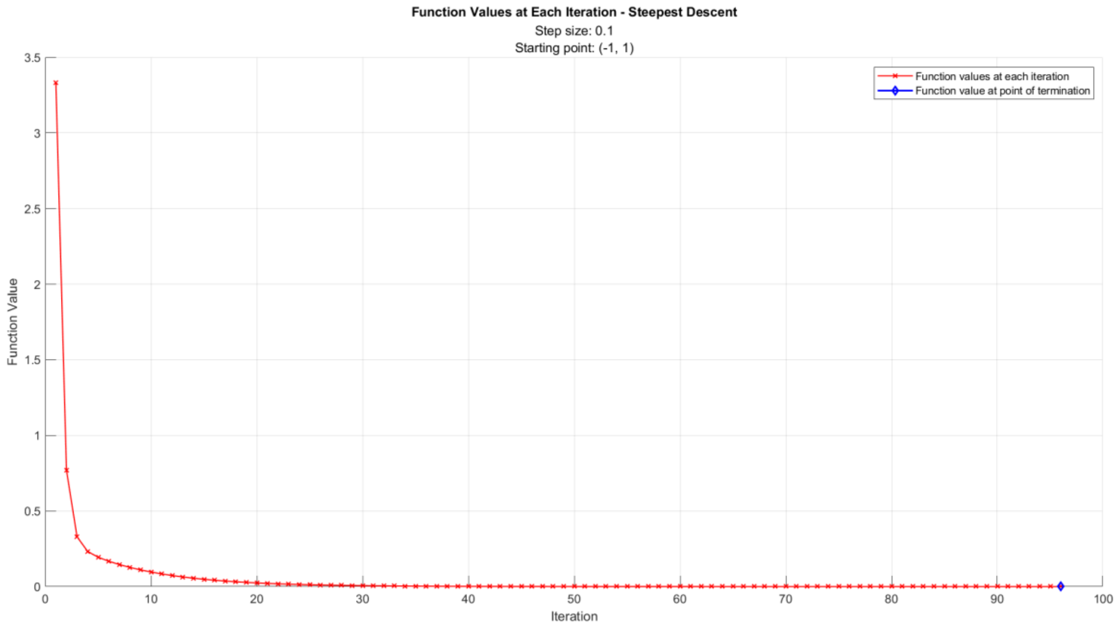
Και αν θέσουμε $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \gamma\nabla f(\boldsymbol{x}_k)$ στην παραπάνω σχέση ο τετραγωνικός όρος που περιλαμβάνει το γ καθορίζει την επίδραση του γ στην τιμή της συνάρτησης.

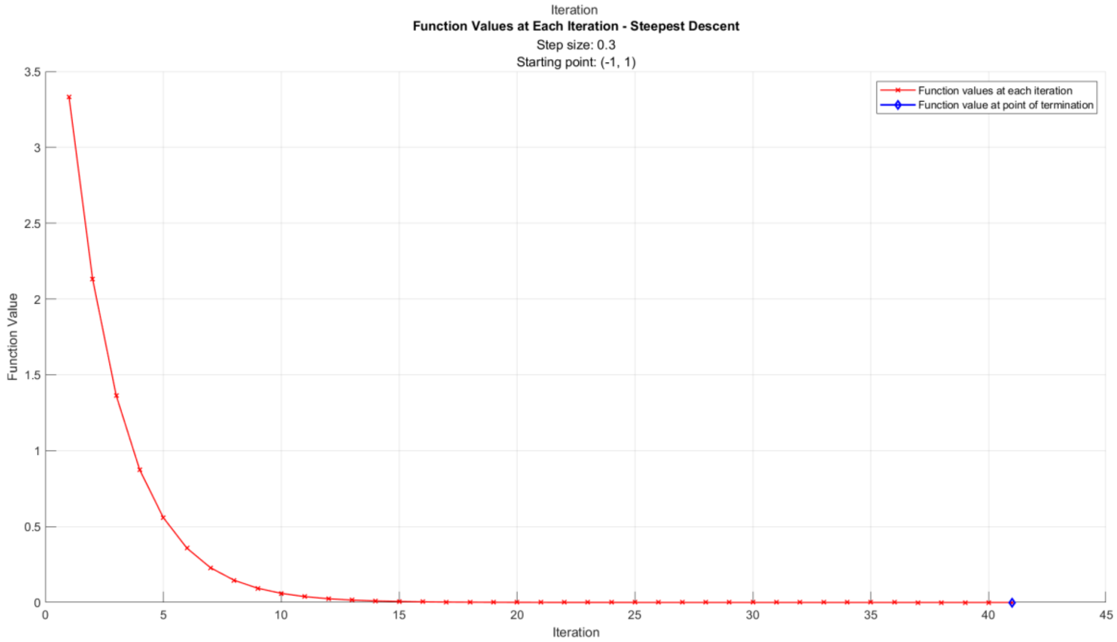
$$f(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{x}_k) - \gamma\|\nabla f(\boldsymbol{x}_k)\|^2 + \frac{L}{2}\|\gamma\nabla f(\boldsymbol{x}_k)\|^2$$

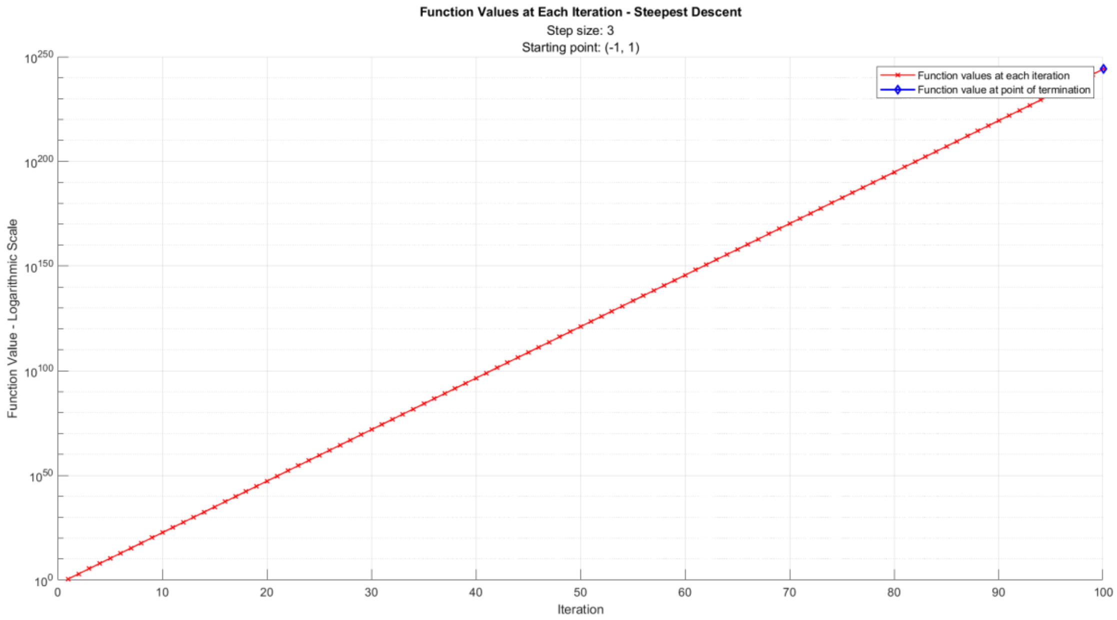
Άρα πρέπει $-\gamma + \frac{L}{2}\gamma^2 < 0 \ \Rightarrow\ \gamma \in (0, \frac{1}{3})$ (αφού L=6) για να συγκλίνει η μέθοδος μέγιστης καθόδου για την συγκεκριμένη συνάρτηση.

Αποτελέσματα της μεθόδου για διάφορα γ

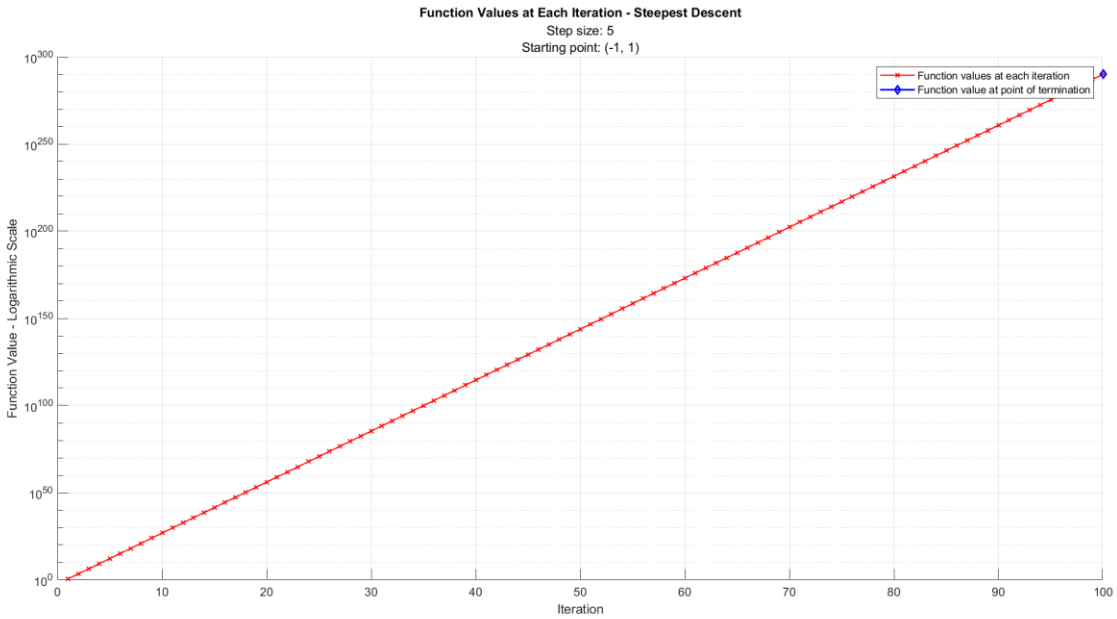
Για να επιβεβαιώσουμε την παραπάνω ανάλυση τρέχουμε στο MATLAB την μέθοδο μέγιστης καθόδου για αρχικό σημείο (-1,1), ακρίβεια ε=0.001, και βήματα γ=0.1, 0.3, 3 και 5. Αναμένουμε βάσει της μαθηματικής ανάλυσης η μέθοδος να συγκλίνει μόνο για τις 2 πρώτες επιλογές βήματος όπου ισχύει η συνθήκη $\gamma < \frac{1}{3}$.


Function Values at Each Iteration - Steepest Descent
Step size: 0.1
Starting point: (-1, 1)
Function values at each iteration
Function value at point of termination
Function Value
Iteration

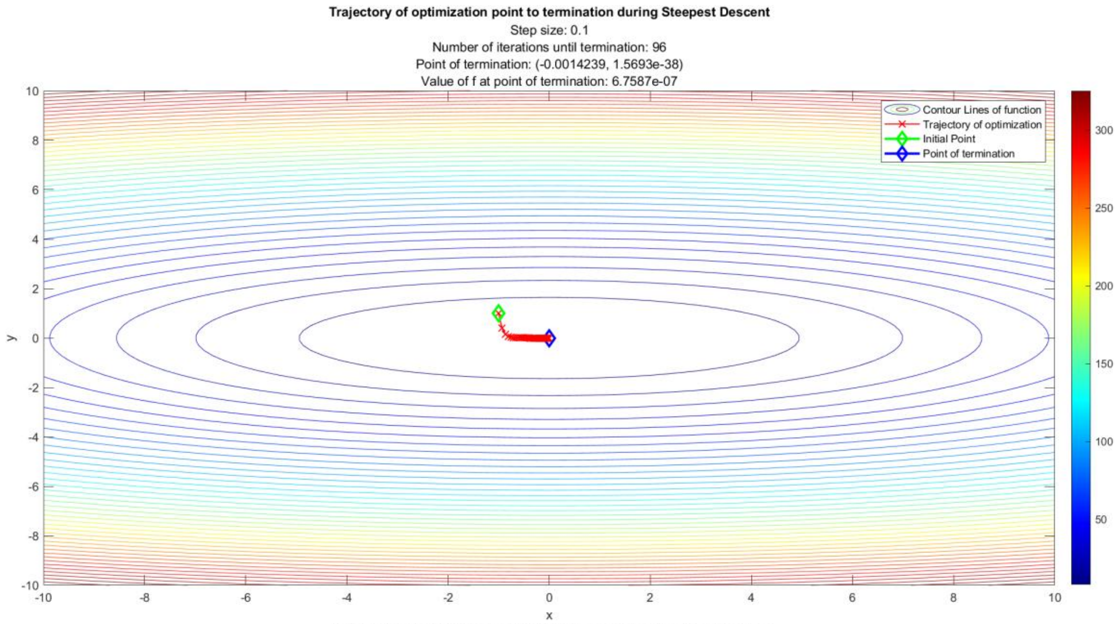

Function Values at Each Iteration - Steepest Descent
Step size: 0.3
Starting point: (-1, 1)
Function values at each iteration
Function value at point of termination
Function Value
Iteration

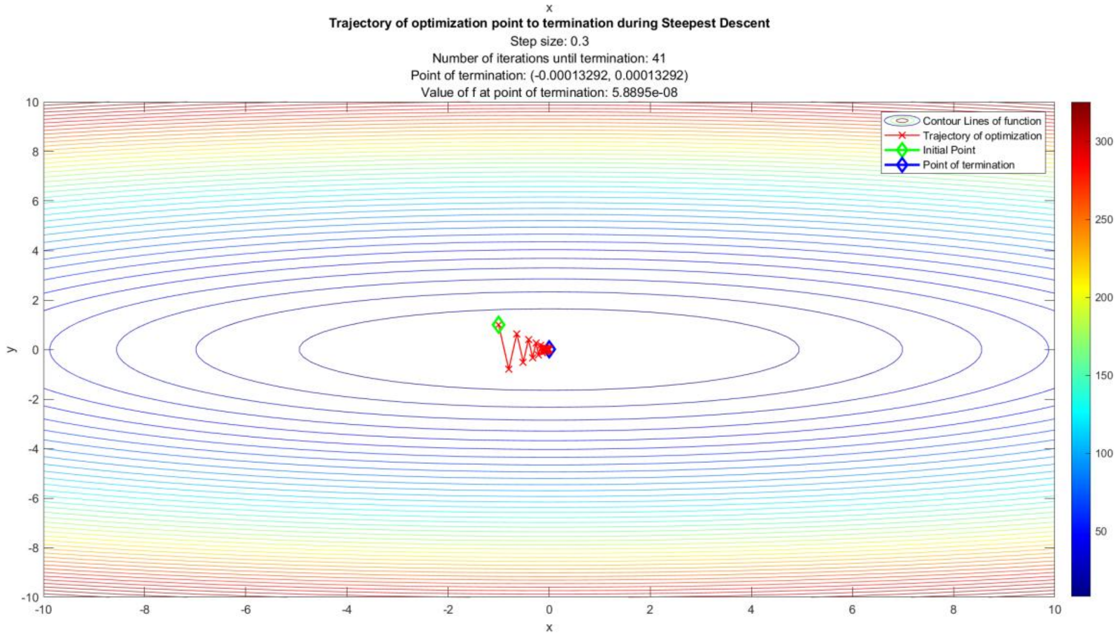


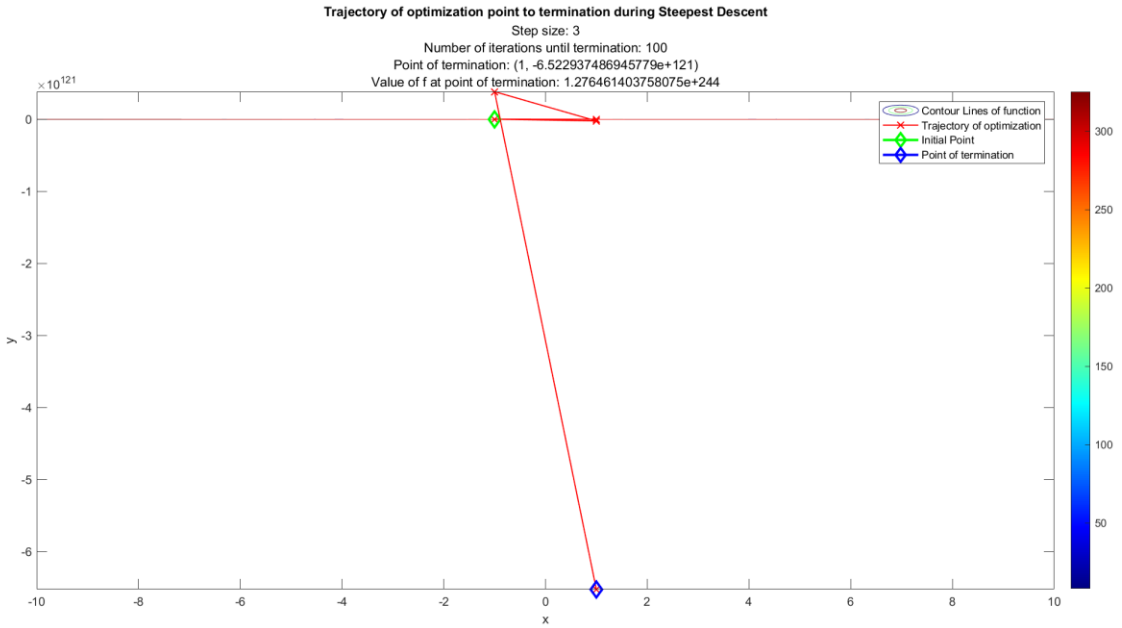
Σημείωση 1 Στο παραπάνω διάγραμμα ο άξονας τιμών της f είναι σε λογαριθμική κλίμακα!

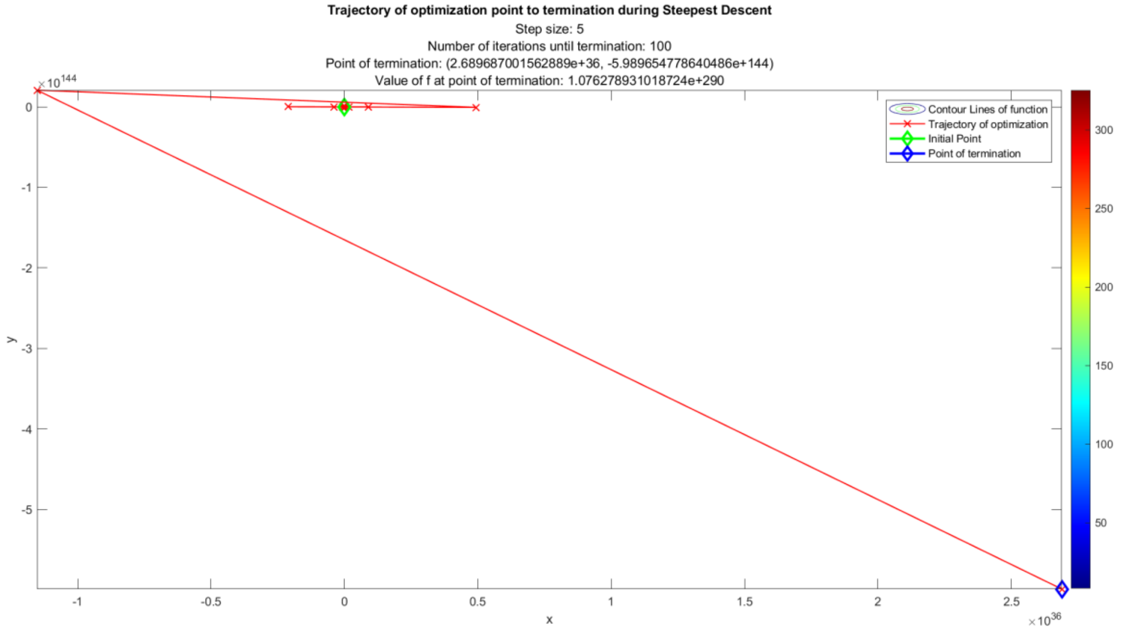


Σημείωση 2 Στο παραπάνω διάγραμμα ο άξονας τιμών της f είναι σε λογαριθμική κλίμακα!


Trajectory of optimization point to termination during Steepest Descent
Step size: 0.1
Number of iterations until termination: 96
Point of termination: (-0.0014239, 1.5693e-38)
Value of f at point of termination: 6.7587e-07
Contour Lines of function
Trajectory of optimization
Initial Point
Point of termination
x
y


Trajectory of optimization point to termination during Steepest Descent
Step size: 0.3
Number of iterations until termination: 41
Point of termination: (-0.00013292, 0.00013292)
Value of f at point of termination: 5.8895e-08
Contour Lines of function
Trajectory of optimization
Initial Point
Point of termination
x
y





Οι υποψίες μας επιβεβαιώνονται. Πράγματι για $\gamma > \frac{1}{3}$ η μέθοδος δεν συγκλίνει στο ελάχιστο, αντίθετα αποκλίνει σε υπερβολικά μεγάλες τιμές πολύ γρήγορα (παρατηρήστε ότι στα διαγράμματα τιμών της f για γ=3 και 5 ο άξονας y είναι σε λογαριθμική κλίμακα!).

Το πρόβλημα με περιορισμούς

Θέλουμε να ελαχιστοποιήσουμε την συνάρτηση $f\ :\ \mathbb{R}^2 \rightarrow \mathbb{R}$ με τύπο $f(x,y) = \frac{1}{3}x^2 + 3y^2,\ (x,y)\ \in X$ παρουσία περιορισμών, δηλαδή καλούμαστε να λύσουμε το πρόβλημα:

$$\min_{(x,y)\,\in X} f(x,y)$$

όπου το X είναι ένα μη-κενό, κλειστό κυρτό υποσύνολο του $\mathbb{R}^n$ και η συνάρτηση $f\colon \mathbb{R}^n \rightarrow \mathbb{R}$ είναι συνεχώς διαφορίσιμη. Μερικοί ορισμοί που θα μας χρησιμεύσουν στην πορεία είναι οι παρακάτω:

- Στάσιμο σημείο x^* θα λέγεται κάθε διάνυσμα που ικανοποιεί την σχέση $\nabla f^T(x^*)(x - x^*) \geq 0\ , \forall\ x \in X$
- Προβολή του διανύσματος $z \in \mathbb{R}^n$ στο X θα λέγεται οποιοδήποτε διάνυσμα $x^* \in X$ με $x^* = \underset{x \in X}{\operatorname{argmin}}|z - x|^2$ και θα το συμβολίζουμε $x^* = Pr_X\{z\}$

Μέθοδος μέγιστης καθόδου με προβολή

Η μέθοδος μέγιστης καθόδου με προβολή θα χρησιμοποιηθεί για το πρόβλημα ελαχιστοποίησης της f παρουσία των δοσμένων περιορισμών

$$-10 \leq x \leq 5$$

$$-8 \leq y \leq 12$$

οι οποίοι ορίζουν ένα ορθογώνιο παραλληλόγραμμο στον $\mathbb{R}^2$ το οποίο θα φανεί ξεκάθαρα στα διαγράμματα που θα ακολουθήσουν.

Θα αξιοποιήσουμε την επαναληπτική σχέση $x_{k+1} = x_k + \gamma_k(\bar{x}_k - x_k), \gamma_k \in (0,1]$

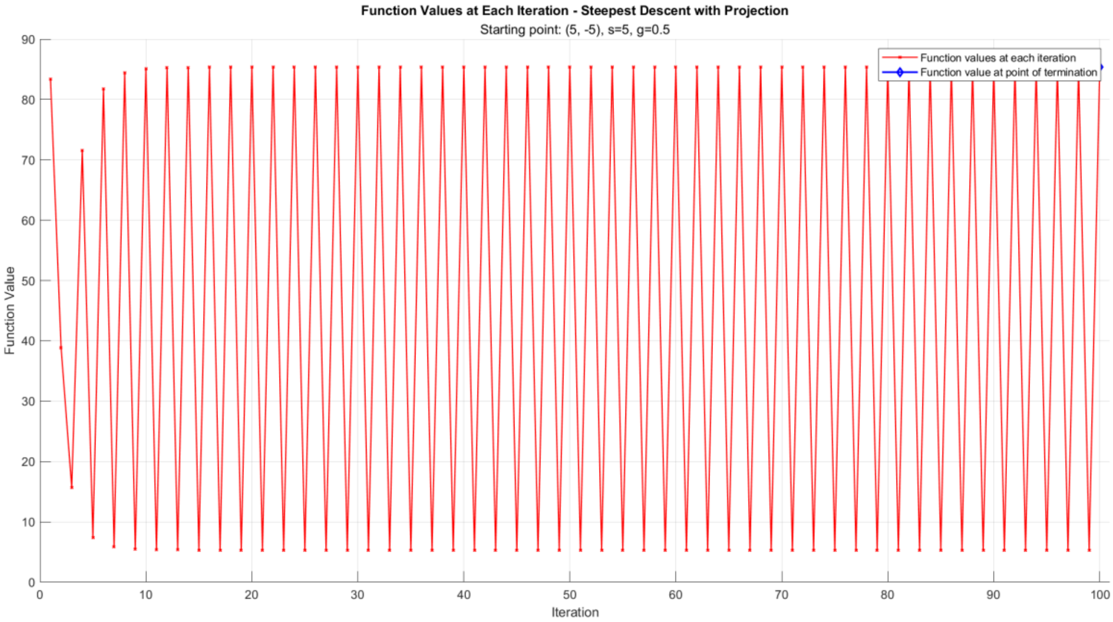
όπου $\bar{x}_k\ = Pr_X\{x_k - s_k \nabla f(x_k)\}$. Συνθήκη τερματισμού της μεθόδου θα είναι το να είναι το τρέχον x_k κρίσιμο σημείο με ακρίβεια ε=0.01. Δηλαδή συνθήκη τερματισμού είναι η $|\ x_k - Pr_X\{x_k - s_k \nabla f(x_k)\}\ | < \varepsilon$.

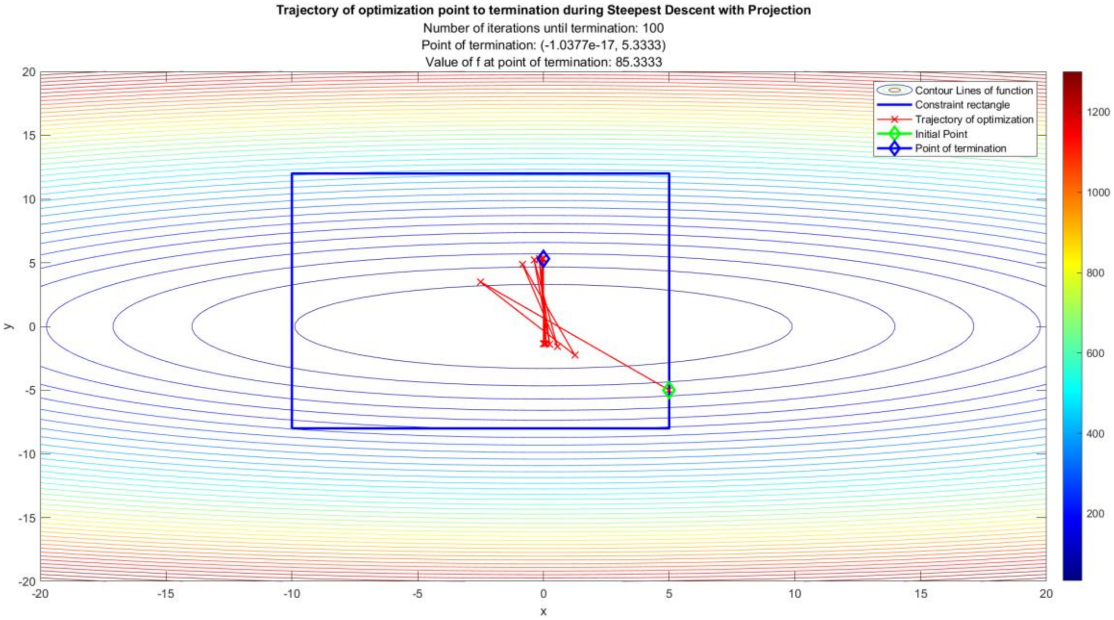
Θα εξετάσουμε τι συμβαίνει για διαφορετικά αρχικά σημεία (x_0, y_0) καθώς και διαφορετικές σταθερές s_k και γ_k με ακρίβεια ε=0.01 .

Κάνουμε τα διαγράμματα της τιμής της αντικειμενικής συνάρτησης σε κάθε επανάληψη, καθώς και της τροχιάς του σημείου της μεθόδου πάνω στις ισοσταθμικές καμπύλες της συνάρτησης.

Παρατήρηση: *Σε όσα διαγράμματα αναφέρεται ο αριθμός επαναλήψεων έως τον τερματισμό να είναι 100, τότε η μέθοδος έχει εξαναγκαστεί να τερματίσει με έναν «κόφτη» επαναλήψεων max_iter=100 που έχουμε ορίσει για να μην υπάρχουν άσκοπες ταλαντώσεις. Εάν ο αριθμός επαναλήψεων είναι μικρότερος του 100 τότε ο αλγόριθμος τερμάτισε λόγω της συνθήκης τερματισμού.*

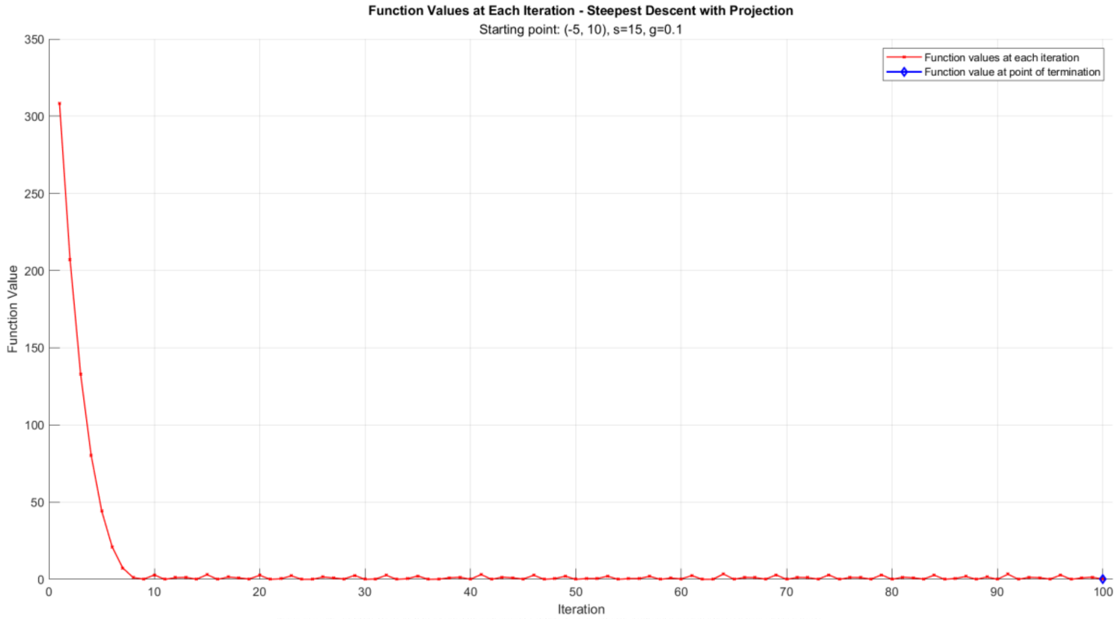
Για s=5, γ=0.5, αρχικό σημείο (5,-5)

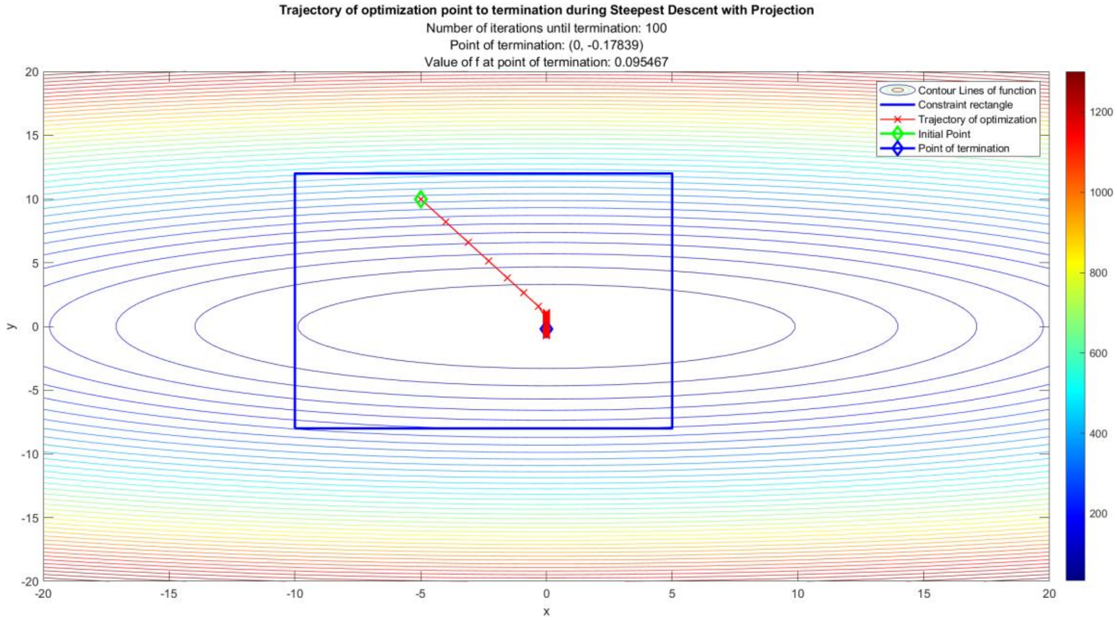




Παρατηρούμε πως η μέθοδος Μέγιστης Καθόδου με προβολή δεν συγκλίνει αλλά ταλαντώνεται γύρω από το σημείο ελαχίστου. Η ταλάντωση αυτή είναι αρκετά μεγάλη καθώς γίνεται εξ 'ολοκλήρου εντός του ορθογωνίου παραλληλογράμμου που ορίζουν οι περιορισμοί, οπότε η μέθοδος σε κάθε επανάληψη κάνει βήμα s=5 που είναι αρκετά μεγάλο. Οπότε αντίστοιχα θα είναι μεγάλο και το πλάτος της ταλάντωσης.

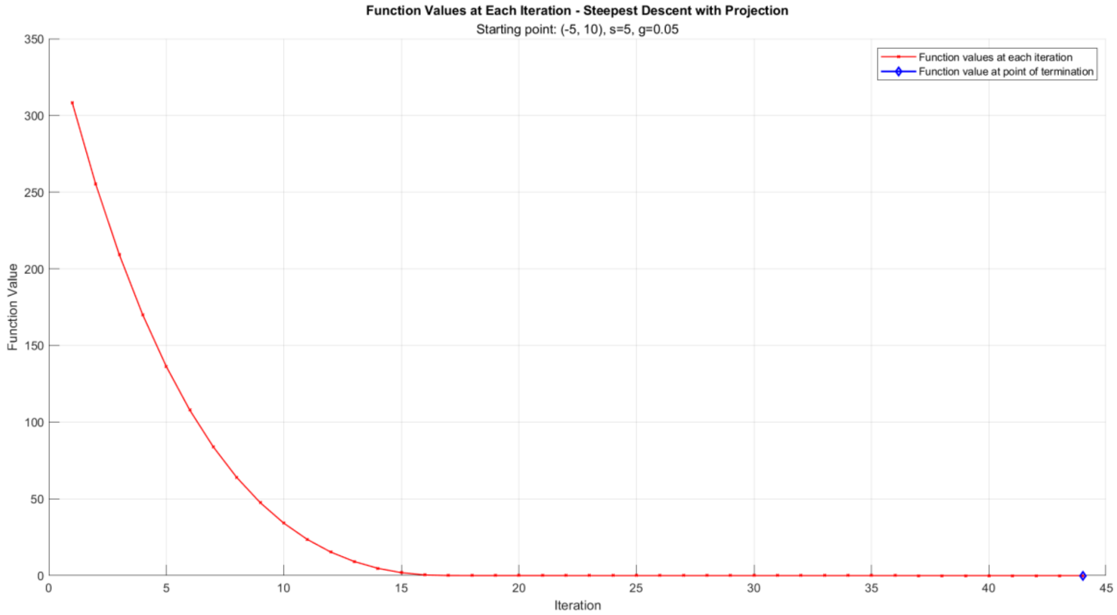
Για s=15, γ=0.1, αρχικό σημείο (-5,10)

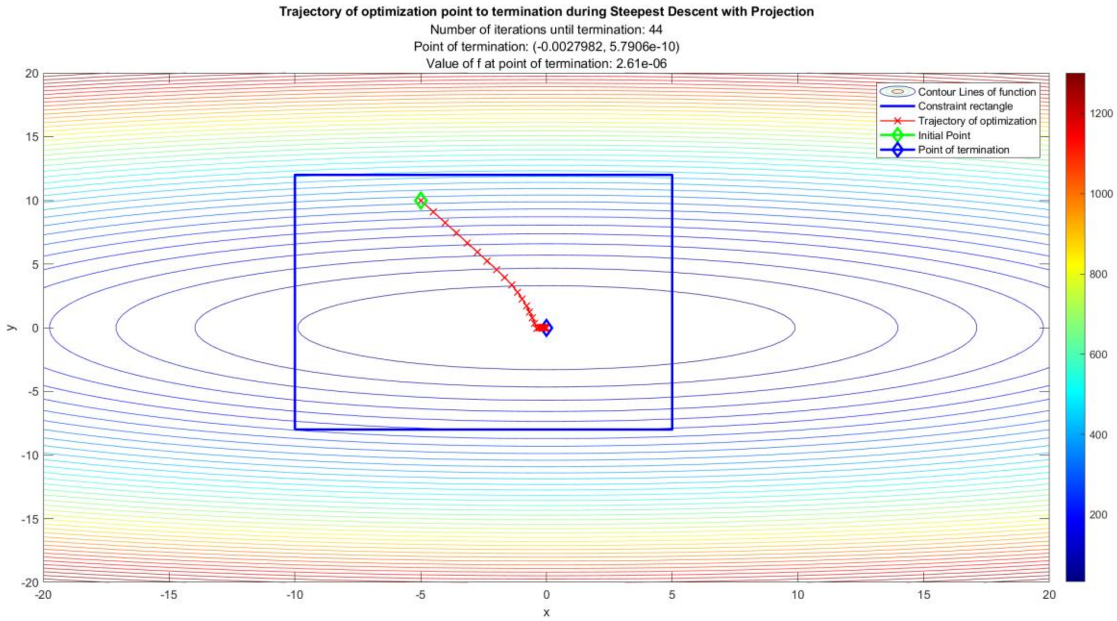




Σε αυτήν την περίπτωση φαίνεται να έχουμε καλύτερα αποτελέσματα: πιο γρήγορη σύγκριση στο ελάχιστο και παρότι υπάρχει πάλι αδυναμία σύγκλισης και ταλάντωση γύρω από το ελάχιστο, αυτήν τη φορά η ταλάντωση έχει πολύ μικρότερο πλάτος. Αυτό οφείλεται στο πολύ μεγάλο s=15 το οποίο «στέλνει» το $\bar{x}_k$ εκτός του μπλε παραλληλόγραμμου που ορίζουν οι περιορισμοί. Οπότε από την μία, σε αντίθεση με το Θέμα 1, λόγω της προβολής η μέθοδος παραμένει εντός των ορίων των περιορισμών παρόλο που το βήμα s είναι μεγάλο, ενώ επίσης λόγω του μικρού βήματος γ, η ταλάντωση που γίνεται γύρω από το σημείο ελαχίστου είναι σημαντικά μικρότερη σε σύγκριση με το Θέμα 2.

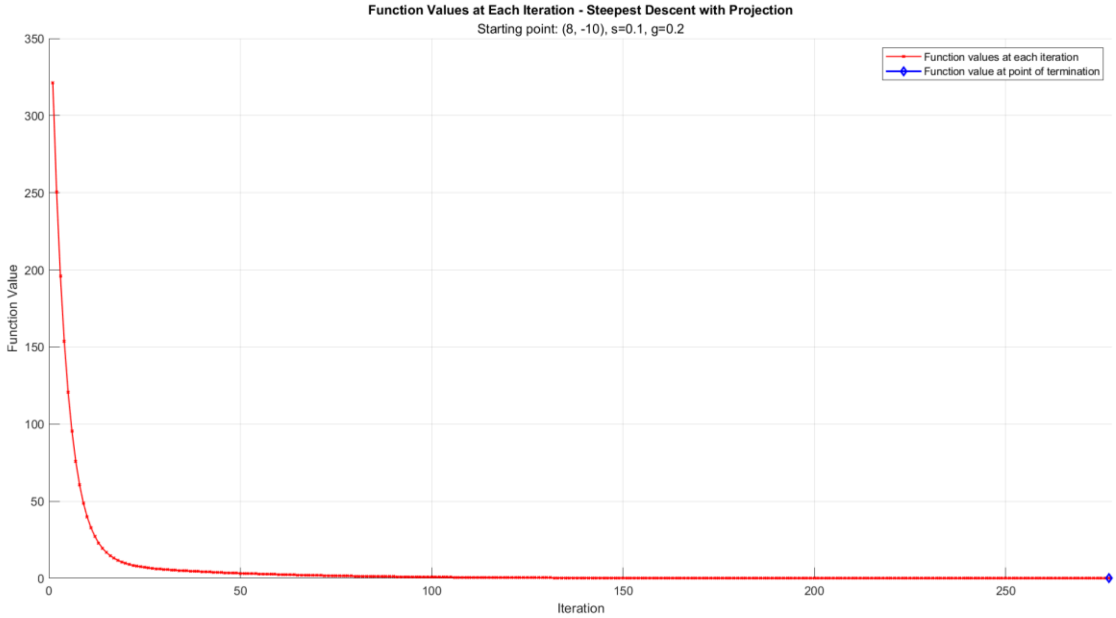
Θέλουμε να βρούμε κάποιον πρακτικό τρόπο να βοηθήσουμε την μέθοδο να συγκλίνει στο σημείο ελαχίστου και να μην ταλαντώνεται γύρω από αυτό. Θέλουμε η μέθοδος να μην «ξεφεύγει» από τα σύνορα των περιορισμών και να εξαναγκάζεται μέσω της προβολής να επιστρέψει εντός αυτών, οπότε μπορούμε να κάνουμε την μέθοδο να συγκλίνει πολύ απλά επιλέγοντας ένα s αρκετά μικρό ώστε να μην την στέλνει εκτός. Ταυτόχρονα θέλουμε ένα αρκετά μικρό γ ώστε να μην υπάρχει ταλάντωση γύρω από το σημείο ελαχίστου, αλλά να μπορεί να βρεθεί επαρκώς κοντά σε αυτό ώστε να τερματίσει. Οπότε επιλέγουμε τον συνδυασμό παραμέτρων s=5, γ=0.05 και βλέπουμε στα διαγράμματα πως η επιλογή αυτή πράγματι μας δικαιώνει καθώς ο αλγόριθμος τερματίζει σε 44 επαναλήψεις συγκλίνοντας πολύ κοντά στο ελάχιστο (0,0).

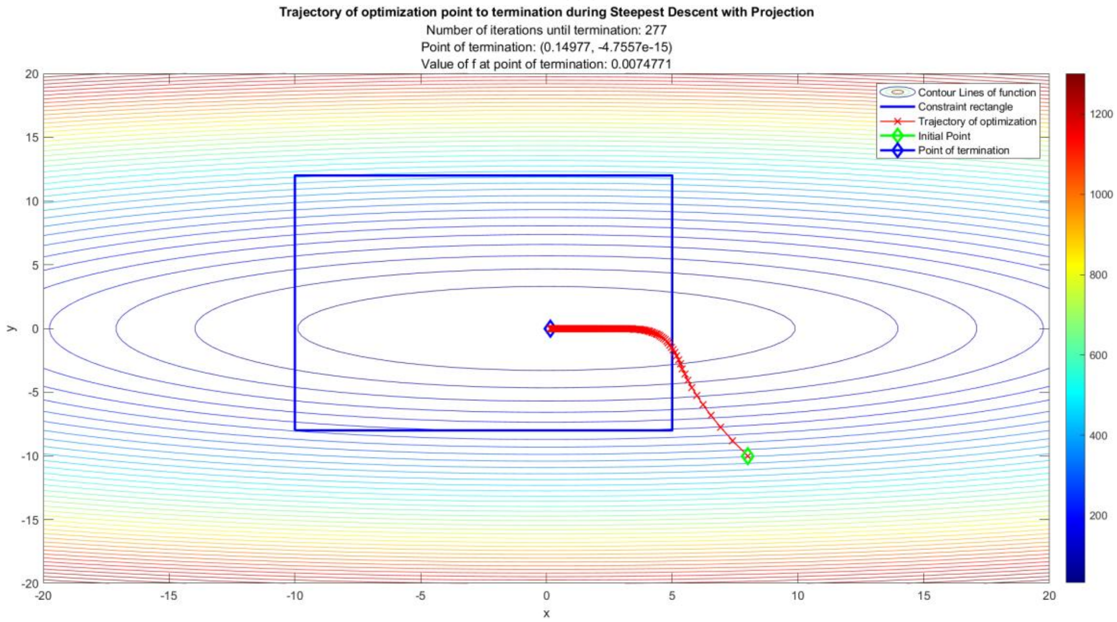




Για s=0.1, γ=0.2, αρχικό σημείο (8,-10)

Καταρχάς παρατηρούμε πως το αρχικό σημείο βρίσκεται εκτός του παραλληλόγραμμου που ορίζουν οι περιορισμοί. Αυτό σημαίνει πως λόγω της προβολής η μέθοδος θα κατευθυνθεί προς το εσωτερικό του παραλληλογράμμου, αλλά αυτό θα συμβεί αργά καθώς το γ=0.2 είναι αρκετά μικρό (το βήμα που γίνεται προς το παραλληλόγραμμο). Επίσης το s=0.1 είναι πολύ μικρό οπότε θα απαιτούνται πολλές επαναλήψεις από την στιγμή που η μέθοδος θα βρεθεί εντός του παραλληλογράμμου μέχρι να συγκλίνει στο ελάχιστο. Τελικά όμως επειδή και τα δύο (s, γ) είναι αρκετά μικρά η μέθοδος θα συγκλίνει στο ελάχιστο απλά θα χρειαστεί πολλές επαναλήψεις. Για αυτόν τον λόγο αυξάνουμε το όριο του «κόφτη» max_iter σε 300. Βλέπουμε στα διαγράμματα πως όντως η μέθοδος τερματίζει απλά σε αρκετές περισσότερες επαναλήψεις σε σύγκριση με το προηγούμενο θέμα καθώς τώρα χρειάζεται επιπλέον επαναλήψεις μέχρι να μπει στο παραλληλόγραμμο των περιορισμών.





Για πιο γρήγορη σύγκλιση δεδομένου ότι ξεκινάμε εκτός των περιορισμών θα θέλαμε ένα αρκετά μεγάλο γ, ώστε να κατευθυνθούμε γρήγορα πίσω στο παραλληλόγραμμο των περιορισμών. Επίσης θα χρειαστεί ένα αρκετά μικρό s ώστε όταν θα βρεθούμε πάλι εντός των περιορισμών να μην υπάρχει ταλάντωση γύρω από το σημείο ελαχίστου αλλά να συγκλίνουμε κατευθείαν σε αυτό έστω και με μικρό βήμα. Για να επιβεβαιώσουμε τις υποθέσεις μας επιλέγουμε γ=2, s=0.1 και τα αποτελέσματα ξανά μας δικαιώνουν καθώς έχουμε την δραματική μείωση των επαναλήψεων από σε 277 σε μόλις 21 μέχρι την σύγκλιση!