3η Εργαστηριακή Άσκηση
Μέθοδος Μέγιστης Καθόδου με Προβολή

Δημήτριος Σουρλαντζής
Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Author Note

Αυτή η αναφορά είναι για την εργασία 3 από 3 για το μάθημα Τεχνικές Βελτιστοποίησης που διδάσκεται στο Τμήμα Ηλεκτρολόγων Μηχανικών Και Μηχανικών Υπολογιστών, του Αριστοτέλειου Πανεπιστημίου Θεσσαλονίκης. Το παρόν έγγραφο δεν αποτελεί κάποια επιστημονική έρευνα, αλλά μια ανάλυση των αλγορίθμων βελτιστοποίησης που ζητήθηκαν να αναλυθούν. Για οποιαδήποτε απορία στείλτε μήνυμα στην ηλεκτρονική διεύθυνση sourland@ece.auth.gr

1 Εισαγωγή

Σε αυτήν την εργασία έχουμε να αναλύσουμε την μέθοδο Μέγιστης Καθόδου με προβολή. Η μέθοδος μέγιστης καθόδου έχει ήδη αναλυθεί στην εργασία 2, και η βασική της φιλοσοφία είναι γνωστή. Εδώ έχουμε κάποιους παραπάνω περιορισμούς, οπότε αναγκαστικά θα πρέπει να τροποποιηθεί ο αλγόριθμος για να υπολογίζει τα αποτελέσματα σύμφωνα με τις νέες συνθήκες. Έχουμε κληθεί να βελτιστοποιήσουμε την παρακάτω συνάρτηση:

$f(x_1, x_2) = \frac{1}{2}\mathrm{x}_1^2 + 2x_2^2$

Υπό τους περιορισμούς:

$-15 \leq x_1 \leq 15$

$-20 \leq x_2 \leq 12$

Η εικόνα (σχήμα 1) και το διάγραμμα κλίσης (σχήμα 2) τής συνάρτησης μας είναι:

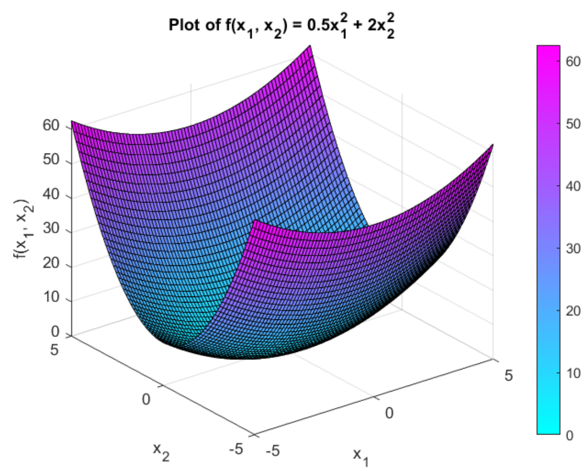

Figure 1: Εικόνα της Συνάρτησης $f(x_1, x_2) = \frac{1}{2}\mathrm{x}_1^2 + 2x_2^2$.

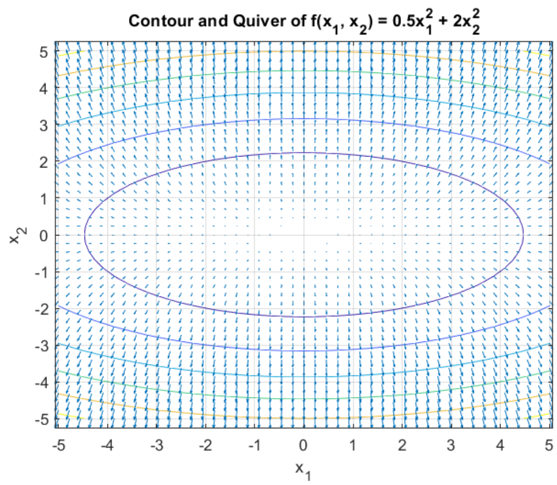

Figure 2: Η κλίση της $f(x_1, x_2) = \frac{1}{2}x_1^2 + 2x_2^2$.

2 Η Γνωστή Μέθοδος Μέγιστης Καθόδου

Έχουμε να αναλύσουμε την μέθοδο μέγιστης καθόδου με ακρίβεια $\varepsilon = 0.01$ και για τέσσερα διαφορετικά βήματα $\gamma_\kappa = 0.05, \gamma_\kappa = 0.5, \gamma_\kappa = 2$ και $\gamma_\kappa = 10$.
Το σημείο εκκίνησης μας θα είναι οποιοδήποτε σημείο εκτός του $(0, 0)$. Έχουμε επιλέξει το σημείο εκκίνησής μας το $(12, -12)$. Ξεκινώντας από την επιλογή $\gamma_\kappa = 0.05$ μπορούμε να δούμε στο σχήμα 3, ότι ο αλγόριθμος βρίσκει με επιτυχία το ελάχιστο της συνάρτησης, ωστόσο η διαδικασία παίρνει αρκετά βήματα.

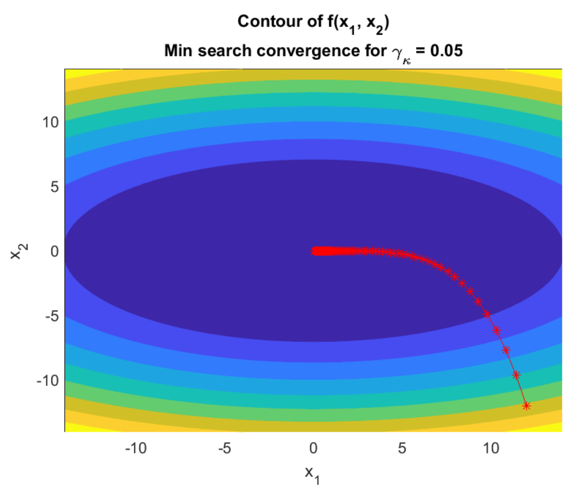

Figure 3: Ο αλγόριθμος συγκλίνει στο ελάχιστο με επιτυχία.

Για τα υπόλοιπα γ_κ, μπορούμε να παρατηρήσουμε ότι ο αλγόριθμος δεν συγκλίνει ποτέ. Θα αναλύσουμε με χρήση μαθηματικών τον λόγο που η επιλογή $\gamma_\kappa = 0.05$ συγκλίνει καθώς και γιατί οι άλλες τρεις επιλογές αποκλίνουν .

Αρχικά υπολογίζουμε την σχέση του $x_{\kappa+1}$ με το x_k. Υπολογίζουμε το gradient της συνάρτησής μας που είναι$\nabla f(x) = (x_1, 4x_2)$. Πολλαπλασιάζουμε το γ_κ με το gradient και έχουμε ότι το

$(x_{\kappa+1,1}, x_{\kappa+1,2}) = (x_{\kappa,1}, x_{\kappa,2}) - (\gamma_\kappa x_{\kappa,1}, 4\gamma_\kappa x_{\kappa,2})$
$= ((1-\gamma_\kappa)x_{\kappa,1}, (1-4\gamma_\kappa)x_{\kappa,2})$.
Τελικά έχουμε ότι $x_{\kappa+1} = Ax_k$ όπου A ο διαγώνιος πίνακας
$\begin{bmatrix} 1-\gamma_\kappa & 0 \\ 0 & 1-4\gamma_\kappa \end{bmatrix}$.
Η νόρμα του πίνακα A συμβολίζεται με $||A||$ και ορίζεται ως η μέγιστη, σε απόλυτη τιμή, ιδιοτιμή λ του πίνακα A. Επειδή, στην περίπτωση αυτή, ο πίνακάς μας είναι διαγώνιος, οι ιδιοτιμές του είναι τα στοιχεία της διαγωνίου του. Η ιδέα είναι ότι αν $||A||\geq 1$, τότε ο αλγόριθμος δεν μπορεί να συγκλίνει.

- $\gamma_\kappa = 0.05$: A $= \begin{bmatrix} 0.95 & 0 \\ 0 & 0.8 \end{bmatrix}$. $||A|| = 0.95 < 1$ άρα ο αλγόριθμος όπως φαίνεται και στο σχήμα 3, συγκλίνει κανονικά. Πιο αναλυτικά, ((x$_{\kappa+1,1}, x_{\kappa+1,2}) = (0.95x_{\kappa,1}, 0.8x_{\kappa,2})$, άρα και οι δύο συντεταγμένες συγκλίνουν στο ελάχιστο που είναι στο (0,0).

- $\gamma_\kappa = 0.5$: A $= \begin{bmatrix} 0.5 & 0 \\ 0 & -1 \end{bmatrix}$. $||A|| = 1$ άρα ο αλγόριθμος δεν μπορεί να συγκλίνει στο ελάχιστο γιατί έχουμε ((x$_{\kappa+1,1}, x_{\kappa+1,2}) = (0.5x_{\kappa,1}, -x_{\kappa,2})$. Παρόλο που η $x_{\kappa+1,1}$ θα μειώνεται, η $x_{\kappa+1,2}$ θα κάνει ταλάντωση ,με πλάτος 1, με αποτέλεσμα να μην συγκλίνει ποτέ το σημείο στο ελάχιστο. Στην υλοποίηση που έχουμε κάνει με την χρήση του εργαλείου MATLAB, έχουμε θέσει ένα κατώφλι ασφαλείας στα βήματα που ορίζεται στα 10000. Για αυτήν την επιλογή του γ_κ, είναι προφανές ότι τα βήματα θα περάσουν το κατώφλι ασφαλείας. Η διαδικασία φαίνεται στο σχήμα 4.

- $\gamma_\kappa = 2$: A $= \begin{bmatrix} -1 & 0 \\ 0 & -7 \end{bmatrix}$. $||A|| = 7 > 1$ άρα ο αλ-

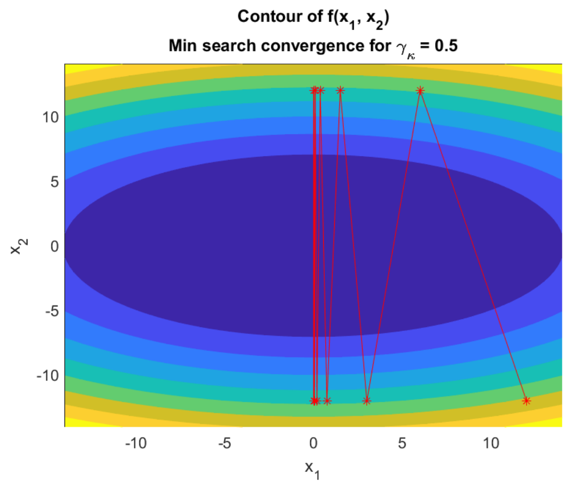

Figure 4: Ο αλγόριθμος καταλήγει να κάνει ταλαντώσεις.

γόριθμος δεν μπορεί να συγκλίνει στο ελάχιστο γιατί έχουμε $((\mathrm{x}_{\kappa+1,1}, x_{\kappa+1,2}) = (-x_{\kappa,1}, -7x_{\kappa,2})$. Το $x_{\kappa+1,1}$ θα κάνει ταλάντωση με πλάτος 1 ενώ το $x_{\kappa+1,2}$ θα αποκλίνει αφού η απόλυτη τιμή του θα αυξάνεται συνεχώς και θα καταλήξει να ταλαντώνεται από το $-\infty$ στο ∞ σε κάθε βήμα, άρα δεν θα καταλήξει στο (0,0) που είναι το ελάχιστο. Σε αυτήν την περίπτωση, ο αλγόριθμος περνάει το δεύτερο κατώφλι ασφαλείας που δεν επιτρέπει στο μέτρο της κλίσης της συνάρτησης να περάσει μία συγκεκριμένη τιμή. Αφού οι απόλυτες τιμές των $((\mathrm{x}_{\kappa+1,1}, x_{\kappa+1,2})$ αυξάνονται συνεχώς, θα αυξάνεται και το μέτρο της κλίσης της συνάρτησης, άρα θα αποκλίνει. Η διαδικασία φαίνεται στο σχήμα 5.

- $\gamma_\kappa = 10$: A $= \begin{bmatrix} -9 & 0 \\ 0 & -39 \end{bmatrix}$. $||A||= 39 > 1$ άρα ο αλγόριθμος δεν μπορεί να συγκλίνει στο ελάχιστο γιατί έχουμε $((\mathrm{x}_{\kappa+1,1}, x_{\kappa+1,2}) = (-9x_{\kappa,1}, -39x_{\kappa,2})$ άρα, στο $x_{\kappa+1,1}$ και στο $x_{\kappa+1,2}$ οι απόλυτες τιμές τους θα

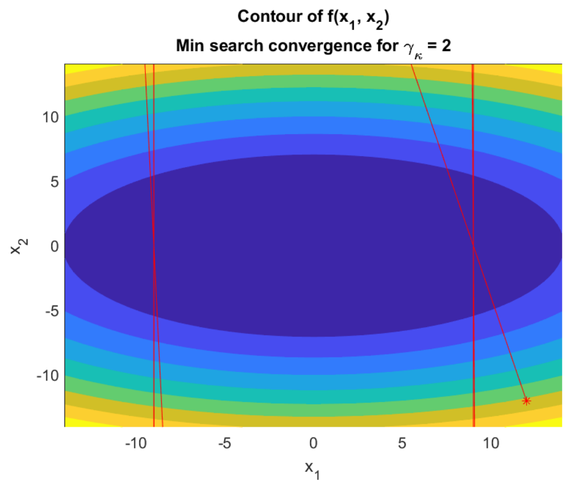

Figure 5: Ο αλγόριθμος αποκλίνει ως προς το $x_{\kappa,2}$ και ταλαντεύεται ως προς το $x_{\kappa,1}$

αυξάνονται. Και οι 2 συντεταγμένες θα κάνουν ταλάντωση από το $-\infty$ στο ∞ σε κάθε βήμα, άρα και πάλι δεν θα καταλήξει ο αλγόριθμος στο (0,0) που είναι το ελάχιστο. Σε αυτήν την περίπτωση, πάλι, ο αλγόριθμος περνάει το δεύτερο κατώφλι ασφαλείας που δεν επιτρέπει το μέτρο της κλίσης της συνάρτησης να περάσει μία συγκεκριμένη τιμή. Αφού οι απόλυτες τιμές των $((\mathrm{x}_{\kappa+1,1}, x_{\kappa+1,2})$ αυξάνονται συνεχώς, θα αυξάνεται και το μέτρο της κλίσης της συνάρτησης, άρα θα αποκλίνει. Η διαδικασία φαίνεται στο σχήμα 6.

2.1 Ποια είναι η συνθήκη για το γ_κ έτσι ώστε ο αλγόριθμος να βρίσκει με επιτυχία το ελάχιστο;

Έχουμε την σχέση του $x_{\kappa+1}$ με το x_κ η οποία είναι $x_{\kappa+1} = Ax_k$. Μεταβάλλουμε την πάνω σχέση ως $x_{\kappa+1} = A_k x_k$ με $A_\kappa = \begin{bmatrix} (1-\gamma_\kappa)^\kappa & 0 \\ 0 & (1-4\gamma_\kappa)^\kappa \end{bmatrix}$. Έχουμε ότι $||A||= max(|\lambda|)$ όπου λ οι ιδιοτιμές του πίνακα. Αφού ο πίνακάς

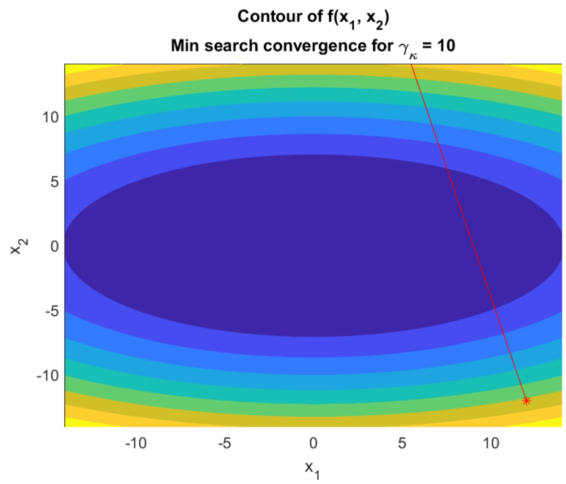

Figure 6: Ο αλγόριθμος αποκλίνει τελείως.

μας είναι διαγώνιος, τότε $\lambda_1 = 1 - \gamma_\kappa, \lambda_2 = 1 - \gamma_\kappa$.
Πρέπει να λύσουμε την ανισότητα $|\lambda_1|>|\lambda_2|$ και να δούμε για ποια γ_κ ισχύει:
$|\lambda_1|>|\lambda_2| \Leftrightarrow |1-\gamma_\kappa|>|1-4\gamma_\kappa|$. Βρίσκουμε τα κρίσιμα σημεία της ανισότητας λύνοντας την $|1-\gamma_\kappa|=|1-4\gamma_\kappa|$. Έχουμε ότι $1-\gamma\kappa = 1-4\gamma_\kappa$ ή $1-\gamma\kappa = -1+4\gamma_\kappa$. Η πρώτη εξίσωση βγάζει $\gamma_\kappa = 0$ ενώ η δεύτερη $\gamma_\kappa = \frac{2}{5}$. Άρα για $0 < \gamma_\kappa < \frac{2}{5}$, $||A||= 1-\gamma_\kappa$ άρα $\frac{3}{5}< ||A||< 1$ και ο αλγόριθμος συγκλίνει. Για $\gamma_\kappa > \frac{2}{5}$ έχουμε ότι $||A||=|1-4\gamma_\kappa|$. Πρέπει να βρούμε τα γ_κ έτσι ώστε $||A||=|1-4\gamma_\kappa|< 1$. Λύνουμε $|1-4\gamma_\kappa|< 1 \Leftrightarrow -1 < 1-4\gamma_\kappa < 1 \Leftrightarrow -2 < -4\gamma_\kappa < 0 \Leftrightarrow 0 < \gamma_\kappa < \frac{1}{2}$. Συμψηφίζοντας τις λύσεις μας για το γ_κ τελικά έχουμε ότι για να συγκλίνει ο αλγόριθμός μας στο ελάχιστο, με την συνάρτηση που μας έχει δοθεί, πρέπει $0 < \gamma_\kappa < \frac{1}{2}$.

3 Η Μέθοδος Μέγιστης Καθόδου με Προβολή

Σε αυτό το κεφάλαιο θα αναλύσουμε την μέθοδο μέγιστης καθόδου με προβολή. Η λογική είναι ότι εξετάζουμε ένα νέο $\tilde{x}$ αν θα βγει εκτός των περιορισμών. Επειδή ο περιορισμός μας είναι ένα παραλληλόγραμμο στο επίπεδο $z = 0$, αν κάποιο $\tilde{x}$ βγει εκτός του περιορισμού, το προβάλλουμε πάνω στο παραλληλόγραμμο. Έστω κάποιο $\tilde{x} = (x_1, x_2)$. με ένα από τα x_1, x_2 να βρίσκεται εκτός περιορισμού. Στην περίπτωση αυτή, απλά θέτουμε την συντεταγμένη από το σημείο που βγήκε εκτός περιορισμού ίσο με το φράγμα από το οποίο ξέφυγε. Η συγκεκριμένη μέθοδος δουλεύει μόνο και μόνο επειδή ο περιορισμός μας είναι παραλληλόγραμμο. Σε κάθε άλλη περίπτωση, όπως για παράδειγμα στην περίπτωση που ο περιορισμός μας ήταν τα x_1, x_2 να ανήκουν σε εντός ενός κυκλικού δίσκου, θα κατασκευάζαμε διαφορετικά την μέθοδο προβολής. Υπενθυμίζεται ότι το $\tilde{x} = x_\kappa - s_\kappa \nabla f(x_\kappa)$ με $x, x_k, \tilde{x} \in R^2$.

- $\gamma_\kappa = 0.05, s_\kappa = 8$
  Με τις παραπάνω επιλογές των γ_κ, s_κ τρέχοντας τον αλγόριθμο μέγιστης καθόδου με προβολή μπορούμε να δούμε στο σχήμα 7 ότι ο αλγόριθμος συγκλίνει στο ελάχιστο με 20 βήματα. Έχουμε ξεκινήσει από το σημείο (10,-5). Για το (Θέμα 1, i), αν συγκρίνουμε το σχήμα 7 με το σχήμα 8, θα δούμε ότι ενώ η χρήση προβολής παρουσιάζει κάποιες ελάχιστες ταλαντώσεις, κάνει πολύ λιγότερα βήματα. Η αντίστοιχη εύρεση ελαχίστου χωρίς προβολή θέλει 137 βήματα. Από την άλλη για το (Θέμα 1, iv), η χρήση προβολής κάνει τον αλγόριθμο να συγκλίνει στο 0, ενώ αν δεν είχαμε προβολή θα απέκλινε τελείως (σχήμα 9). Ο αλγόριθμος μας θα είναι σαν να

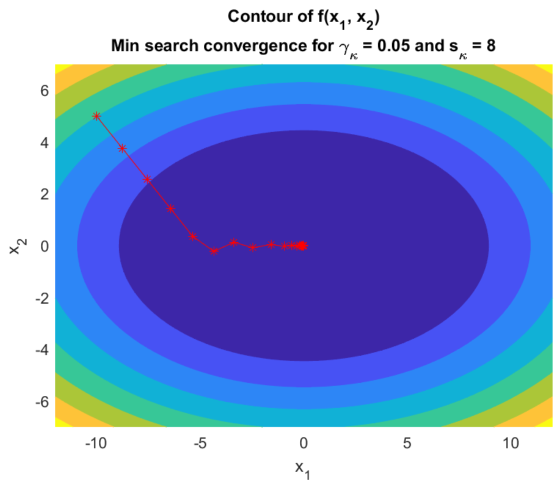

Figure 7: Η εύρεση ελαχίστου για $\gamma_\kappa = 0.05, s_\kappa = 0.8$

τρέχει με βήμα 0.4 που είναι μέσα στην συνθήκη σύγκλισης. Ο αλγόριθμος του θέματος (Θέμα 1, i) θα περιμένουμε να κάνει περίπου τα δεκαπλάσια βήματα από αυτόν τον αλγόριθμο.

- $\gamma_\kappa = 0.3, s_\kappa = 10$
  Όπως φαίνεται και στο σχήμα 10, ο αλγόριθμος δεν μπορεί να τερματίσει γιατί κάνει ταλαντώσεις. Ξεκινώντας από το (-7,5) ο αλγόριθμος κάνει άλματα πάνω από το ελάχιστο, και παραβιάζεται η συνθήκη ασφαλείας για τα βήματα, οπότε τερματίζεται αυτόματα. Αυτό συμβαίνει γιατί έχουμε ότι $\tilde{x} = x_\kappa - s_\kappa \nabla f(x_\kappa)$, οπότε $x_{\kappa+1} = x_\kappa - \gamma_\kappa s_\kappa \nabla f(x_\kappa)$. Τελικά η μέθοδος μέγιστης καθόδου με χρήση προβολής, γίνεται η μέθοδος μέγιστης καθόδου με βήμα $\gamma'_\kappa = s_\kappa \gamma_\kappa$ για το οποίο ισχύει $0 < \gamma'_\kappa < \frac{1}{2}$. Θα εξετάσουμε δύο πρακτικούς τρόπους για να κάνουμε τον αλγόριθμο να συγκλίνει.
  - Κρατάμε ίδιο το γ_κ και λύνουμε ως προς s_κ

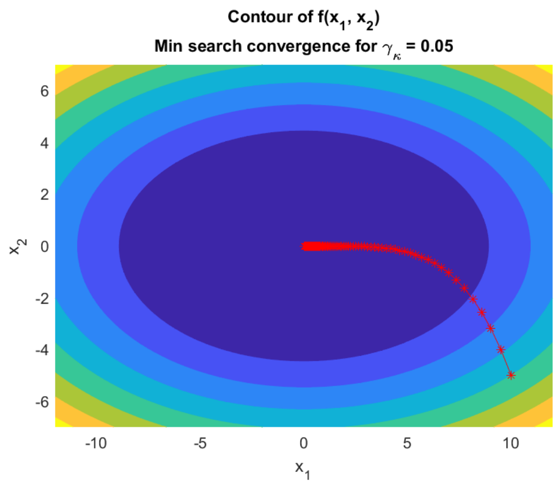

Figure 8: Η εύρεση ελαχίστου για $\gamma_\kappa = 0.05$

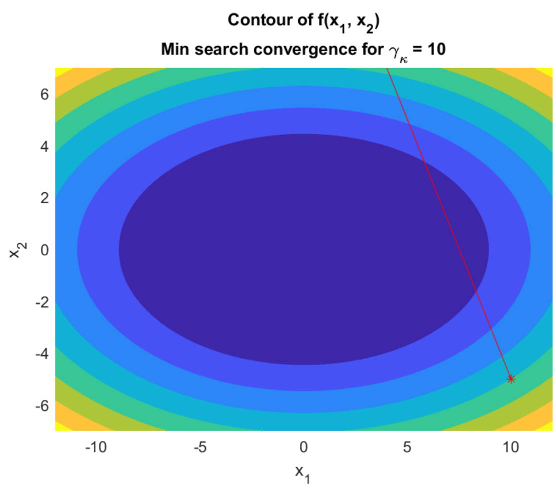

Figure 9: Η αποτυχία εύρεσης ελαχίστου για $\gamma_\kappa = 10$

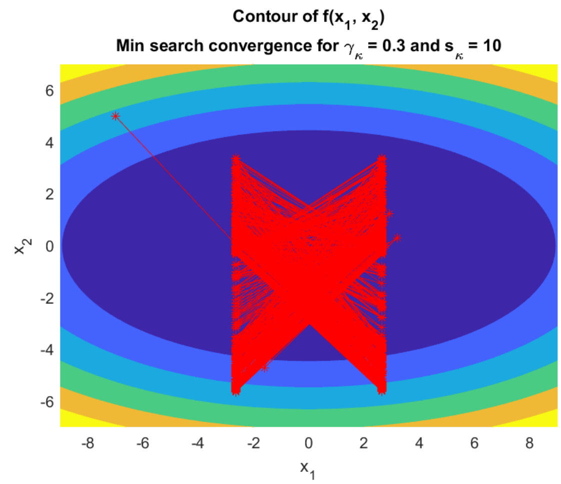

Figure 10: Οι ταλαντώσεις του αλγορίθμου για $\gamma_\kappa = 0.3, s_\kappa = 10$

Πρέπει $0 < \gamma'_\kappa < \frac{1}{2} \Leftrightarrow 0 < \gamma_\kappa s_\kappa < \frac{1}{2}$. Με $\gamma_\kappa = 0.3$ έχουμε $0 < s_\kappa < \frac{5}{3}$.
Θέτουμε το $s_\kappa = \frac{4}{3}$ και θα έχουμε σύγκλιση όπως στο σχήμα 11.

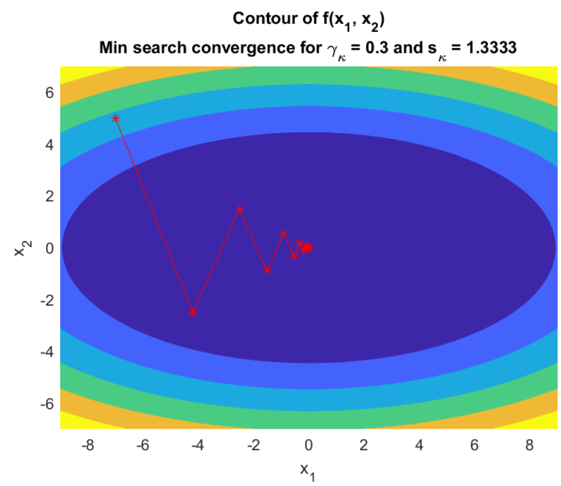

Figure 11: Η εύρεση του ελαχίστου ολοκληρώθηκε με επιτυχία για σταθερό γ_κ.

– Κρατάμε ίδιο το s_κ και λύνουμε ως προς γ_κ

Με την ίδια λογική θα έχουμε ότι $0 < \gamma_\kappa < \frac{1}{20}$ άρα για $\gamma_\kappa = 0.01$ έχουμε σύγκλιση όπως στο σχήμα 12.

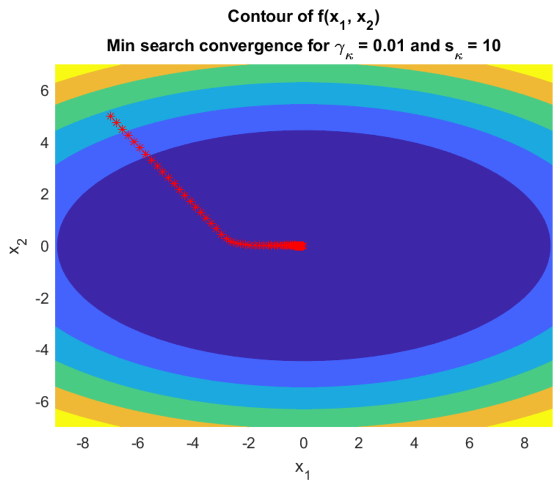

Figure 12: Η εύρεση του ελαχίστου ολοκληρώθηκε με επιτυχία για σταθερό s_κ.

Αν μεταβάλουμε το s_κ έχουμε λιγότερα βήματα, συγκεκριμένα 16, αλλά έχουμε ταλαντώσεις, ενώ αν μεταβάλουμε το γ_κ, δεν έχουμε ταλαντώσεις αλλά έχουμε περισσότερα βήματα, συγκεκριμένα 72. Όσο αφορά το θέμα 1, (i) μπορούμε να δούμε ότι ο αλγόριθμος μας θα έχει βήμα 3, που σπάει την συνθήκη σύγκλισης για το βήμα. Για το ερώτημα (iv), το γ_κ έχει την ίδια τιμή με το s_κ. Χρησιμοποιώντας προβολή, με $\gamma_\kappa = 0.3$, αποτρέπουμε τον αλγόριθμο από το να αποκλίνει στο άπειρο (θα είναι σαν να έχει βήμα 3). Ωστόσο και πάλι ο αλγόριθμος κάνει ταλαντώσεις και δεν θα συγκλίνει.

- $\gamma_\kappa = 0.5, s_\kappa = 0.1$
  Με αυτές τις συνθήκες ο αλγόριθμος συγκλίνει στο ελάχιστο όπως στο σχήμα 13. Παρόλο που ξεκινάμε

με εκτός του περιορισμού μας, τα γ_κ, s_κ είναι αρκετά μικρά έτσι ώστε να μην κάνει τεράστια βήματα, οπότε δεν θα αποκλίνει. Σύμφωνα με την σχέση που βρήκαμε στο θέμα 3 για το βήμα, το συνολικό μας βήμα θα είναι $\gamma'_\kappa = 0.05$ που είναι ίδιο με το βήμα από το θέμα 1 ερώτημα i. Ωστόσο, και σε αυτήν την περίπτωση, αφού έχουμε μικρό βήμα, έχουμε πολλά βήματα για να φτάσουμε στο ελάχιστο και θα χρειαστεί παραπάνω χρόνος. Επειδή εδώ τυχαίνει να έχουμε ένα ελάχιστο που είναι και το ολικό, προφανώς, δεν υπάρχει κάποιος ιδιαίτερος κίνδυνος στο γεγονός ότι ξεκινάμε εκτός περιορισμών. Ωστόσο αν είχαμε κάποια άλλη συνάρτηση με πολλαπλά ελάχιστα που δεν ήταν ολικά, θα υπήρχε κίνδυνος να τερματίσουμε σε λύση που δεν θα ήταν η βέλτιστη, οπότε η παραβίαση του περιορισμού θα είχε σημαντικές επιπτώσεις.

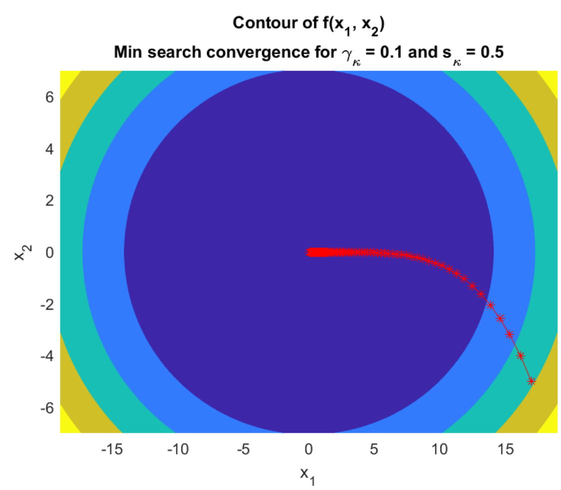

Figure 13: Η εύρεση του ελαχίστου ολοκληρώθηκε με επιτυχία για σταθερό γ_κ.

4 Αποτελέσματα και Σχόλια

Από την ανάλυση και τα αποτελέσματα των αλγορίθμων μπορούν να βγουν διάφορα συμπεράσματα. Για την συγκεκριμένη συνάρτηση που δίνεται, βρέθηκαν τα κριτήρια σύγκλισης για το βήμα ενώ είδαμε ότι υπάρχει μια αντιστρόφως ανάλογη σχέση ανάμεσα στα βήματα και στις ταλαντώσεις που κάνει ο αλγόριθμος, όταν αυτός συγκλίνει.

- Τι σημαίνει να έχουμε πολλά βήματα στον αλγόριθμό μας;
  Είναι προφανές ότι αν έχουμε περισσότερα βήματα, ο αλγόριθμος θα χρειαστεί παραπάνω ώρα για να ολοκληρωθεί. Το πόσο είναι αυτός ο χρόνος σε διάρκεια, είναι καίριο στην επιλογή του να έχουμε πολλά βήματα, αφού είναι ασφαλέστερο.

- Τι σημαίνει όταν ο αλγόριθμος παρουσιάζει ταλαντώσεις;
  Η πιο ενδιαφέρουσα περίπτωση είναι όταν ο αλγόριθμος μας παρουσιάζει ταλαντώσεις ενώ συγκλίνει, όπως στα σχήματα 7 και 11. Ενώ παρατηρούμε ότι ο αλγόριθμος κάνει λιγότερα βήματα για να βρει το ελάχιστο, υπάρχουν συγκεκριμένοι κίνδυνοι με αυτήν την τεχνική. Αν είχαμε κάποια άλλη συνάρτηση με πολλά τοπικά ελάχιστα, θα υπήρχε μεγάλη πιθανότητα να τερματίσουμε σε ένα από αυτά, οπότε να μην βρίσκαμε την βέλτιστη λύση, που είναι παρόμοια περίπτωση με το αν ξεκινούσαμε εκτός περιορισμών, όπως στο θέμα 4.