3^{η} Εργασία Τεχνικών Βελτιστοποίησης

Αλεξόπουλος Δημήτριος ΑΕΜ 10091

16 Οκτωβρίου 2023

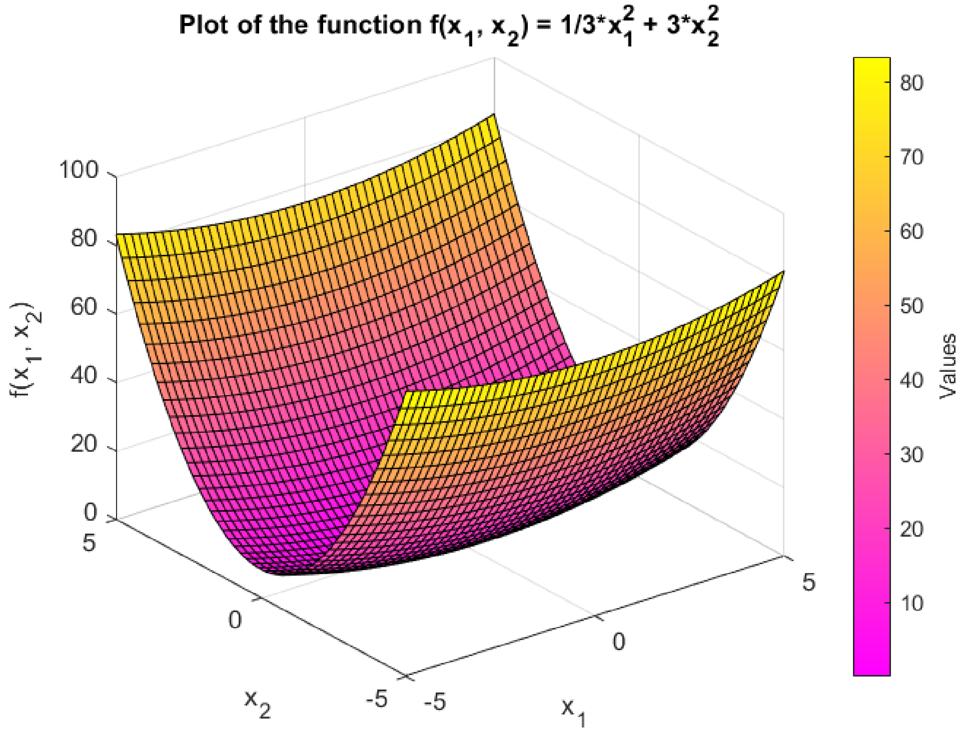
Εισαγωγή

Στην ακόλουθη εργασία θα διερευνήσουμε τις αδυναμίες της απλής μεθόδου μέγιστης καθόδου κι έπειτα θα την προσαρμόσουμε, ώστε να λειτουργεί υπό περιορισμούς. Για τον σκοπό αυτό θα χρησιμοποιηθεί η μέθοδος της μέγιστης καθόδου με προβολή.
Η αντικειμενική συνάρτηση που θα μελετήσουμε είναι η:

$$f(x) = \frac{1}{3}x_1^2 + 3x_2^2 \text{ , με } x = [x1\ x2]^T$$

1 Γραφική Παράσταση της Αντικειμενικής Συνάρτησης

Αρχικά, για να έχουμε μια καλύτερη εικόνα της συνάρτησης που πρόκειται να μελετήσουμε, την σχεδιάζουμε. Παρακάτω την βλέπουμε στις τρεις διαστάσεις και παρατηρούμε την μορφή της και την θέση, διαισθητικά, του ελαχίστου.



Σχήμα 1: Γραφική παράσταση της f

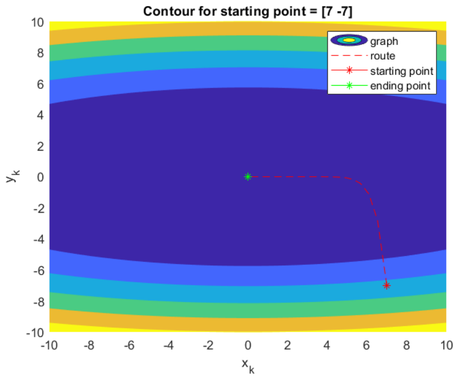
2 Γνωστή Μέθοδος Μέγιστης Καθόδου

Υπενθυμίζουμε ότι η μέθοδος της μέγιστης καθόδου (Steepest Descent) προκύπτει επιλέγοντας $\Delta_k = I, k = 1, 2, 3, ...$ στην σχέση $x_{k+1} = x_k - \gamma_k \Delta_k \nabla f(x_k)$, όπου I ο μοναδιαίος $n \times n$ πίνακας.

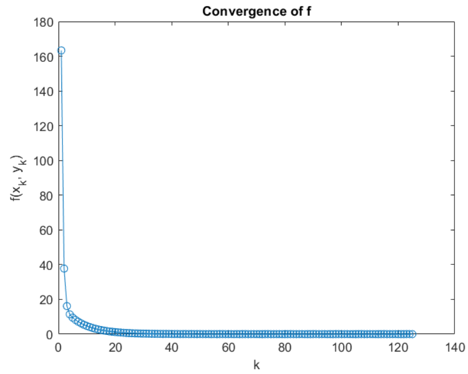
Σε αυτό το σημείο θα ελέγξουμε την αποτελεσματικότητα του αλγορίθμου με ακρίβεια $\epsilon = 0.001$ και σημείο εκκίνησης οποιοδήποτε πέραν του $(0, 0)$. Επιλέγουμε αρχικό σημείο το $(7, -7)$, ενώ για το βήμα γ_k κάνουμε τις εξής δοκιμές:

2.1 Σταθερό $\gamma_k = 0.1$

Η τιμή της συνάρτησης συγκλίνει στο ελάχιστο για $(x_{1k}, x_{2k}) = (0.001348, -0.000000)$ με τιμή ελαχίστου 0.000001. Ωστόσο, ο αριθμός των επαναλήψεων k είναι αρκετά μεγάλος όπως φαίνεται στο παρακάτω διάγραμμα.



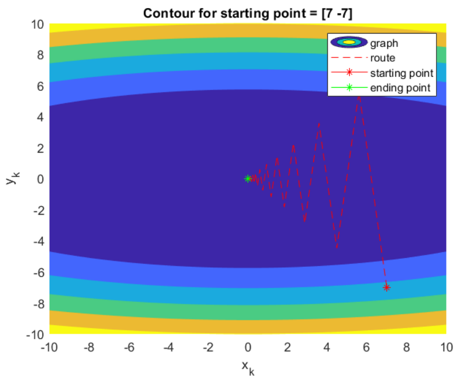
Σχήμα 2: Οπτικοποίηση της εύρεσης ελαχίστου για σταθερό $\gamma_k = 0.1$



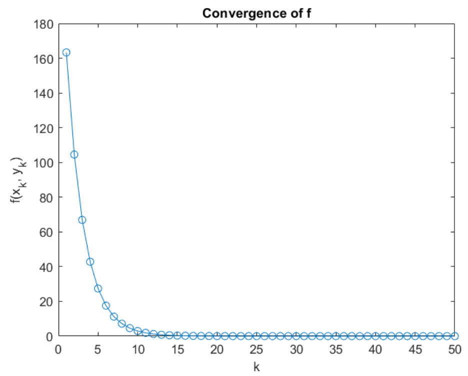
Σχήμα 3: Σύγκλιση της αντικειμενικής για σταθερό $\gamma_k = 0.1$

2.2 Σταθερό $\gamma_k = 0.3$

Η τιμή της συνάρτησης συγκλίνει στο ελάχιστο για $(x_{1k}, x_{2k}) = (0.000125, 0.000125)$ με τιμή ελαχίστου 0.000000. Ο αριθμός των επαναλήψεων k είναι αισθητά μικρότερος από πριν, ωστόσο είναι και πάλι μεγάλος, ενώ παράλληλα παρατηρούμε ότι στην σύγκλιση της αντικειμενικής συνάρτησης είναι εμφανείς ταλαντώσεις.



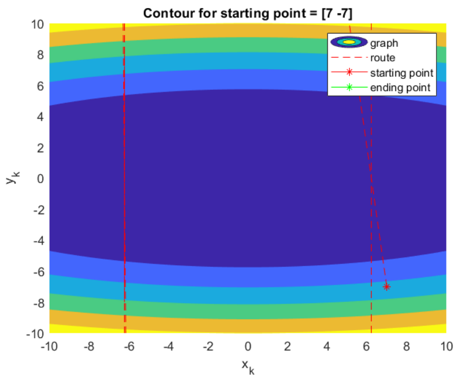
Σχήμα 4: Οπτικοποίηση της εύρεσης ελαχίστου για σταθερό $\gamma_k = 0.3$



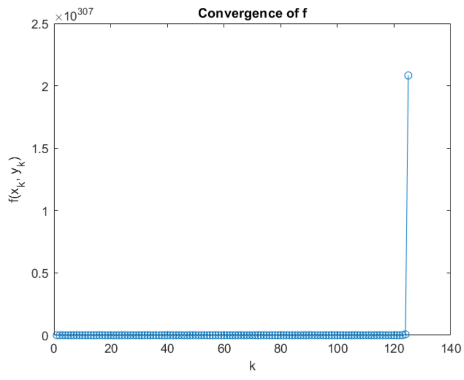
Σχήμα 5: Σύγκλιση της αντικειμενικής για σταθερό $\gamma_k = 0.3$

2.3 Σταθερό $\gamma_k = 3$

Η τιμή της συνάρτησης δεν συγκλίνει στο ελάχιστο. Παρατηρούμε ταλαντώσεις οι οποίες μάλιστα απομακρύνουν την αντικειμενική συνάρτηση κατά πολύ από το ελάχιστο, ενώ ο αλγόριθμος τερματίζει μετά από πολλές επαναλήψεις με απροσδιόριστη τιμή NaN.



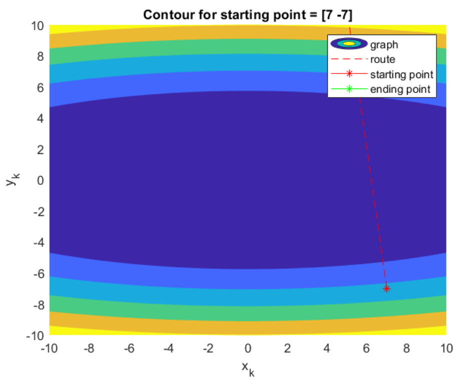
Σχήμα 6: Οπτικοποίηση της εύρεσης ελαχίστου για σταθερό $\gamma_k = 3$



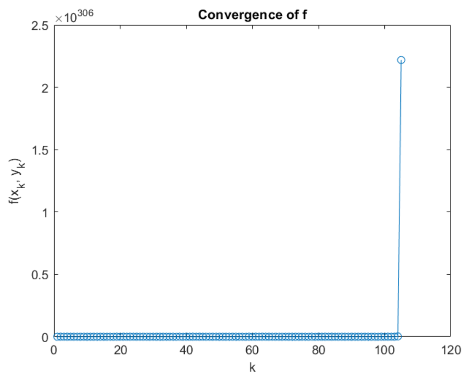
Σχήμα 7: Σύγκλιση της αντικειμενικής για σταθερό $\gamma_k = 3$

2.4 Σταθερό $\gamma_k = 5$

Όπως και πριν, η τιμή της συνάρτησης δεν συγκλίνει στο ελάχιστο. Παρατηρούμε και πάλι ταλαντώσεις οι οποιές μάλιστα απομακρύνουν την αντικειμενική συνάρτηση κατά πολύ από το ελάχιστο, ενώ ο αλγόριθμος τερματίζει μετά από πολλές επαναλήψεις με απροσδιόριστη τιμή NaN.



Σχήμα 8: Οπτικοποίηση της εύρεσης ελαχίστου για σταθερό $\gamma_k = 5$



Σχήμα 9: Σύγκλιση της αντικειμενικής για σταθερό $\gamma_k = 5$

2.5 Μαθηματική Ερμηνεία

Στη συνέχεια θα ερμηνεύσουμε με μαθηματική αυστηρότητα τα παραπάνω πειραματικά αποτελέσματα. Βλέπουμε, αρχικά, ότι η αντικειμενική μας συνάρτηση είναι τετραγωνικής μορφής με θετικά ορισμένο εσσιανό πίνακα Q. Έχουμε δηλαδή:

$$f(x) = \frac{1}{2}x^T Qx, \quad \nabla f(x) = Qx, \quad \nabla^2 f(x) = Q,$$

όπου $x = [x1 \;\; x2]^T$ και $Q = \begin{bmatrix} \frac{2}{3} & 0 \\ 0 & 6 \end{bmatrix}$.
Η μέθοδος της μέγιστης καθόδου έχει τη μορφή:

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

Επομένως,

$$|x_{k+1}|^2 = x_{k+1}^T x_{k+1} = x_k^T (I - \gamma_k Q)^2 x_k \leq \lambda_{max}((I - \gamma_k Q)^2)|x_k|^2,$$

όπου λ_{max} η μέγιστη ιδιοτιμή του πίνακα. Συνεπώς,

$$\lambda_{max}((I - \gamma_k Q)^2) = max\{(1 - \gamma_k m)^2, (1 - \gamma_k M)^2\}$$

όπου m και M η μικρότερη και η μεγαλύτερη ιδιοτιμή, αντίστοιχα, του πίνακα Q. Από τα παραπάνω συμπεραίνουμε πως $\forall x_k \neq 0$ θα ισχύει:

$$\frac{|x_{k+1}|}{|x_k|} \leq max\{|1-\gamma_k m|, |1-\gamma_k M|\}$$

Επομένως, οι τιμές των βημάτων γ_k που εγγυώνται την ευστάθεια του αλγορίθμου είναι εκείνες που αντιστοιχούν σε $\frac{|x_{k+1}|}{|x_k|} \leq 1$. Άρα σε:

$$\gamma_k M - 1 \leq 1 \text{ ή } \boxed{\gamma_k \leq \frac{2}{M}}$$

Πράγματι, στην περίπτωσή μας από τον πίνακα Q βλέπουμε ότι $M = 6$ και επομένως η συνθήκη σύγκλισης της αντικειμενικής συνάρτησης είναι $\gamma_k \leq \frac{1}{3} = 0.33$. Γι' αυτό και παρατηρούμε ότι στις περιπτώσεις που $\gamma_k = 0.1$ ή 0.3 έχουμε σύγκλιση, ενώ στις περιπτώσεις που $\gamma_k = 3$ ή 5 η αντικειμενική αποκλίνει από το ελάχιστο.

3 Μέθοδος Μέγιστης Καθόδου με Προβολή

Εδώ θα αναλύσουμε την μέθοδο μέγιστης καθόδου με προβολή. Η μέθοδος αυτή ακολουθεί την εξής λογική:

1. Ξεκινούμε από ένα εφικτό σημείο $x_0 \in X$.
2. Ακολουθούμε τον γνωστό αλγόριθμο της μέγιστης καθόδου χωρίς περιορισμούς.
3. Αν το νέο σημείο είναι εφικτό, συνεχίζουμε με τον ίδιο αλγόριθμο.
4. Αν το νέο σημείο δεν είναι εφικτό, βρίσκουμε την προβολή του στο Q και πηγαίνουμε και πάλι το βήμα 2 μέχρι να καταλήξουμε σε στάσιμο σημείο.

Στην περίπτωσή μας, επειδή ο περιορισμός είναι ένα παραλληλόγραμμο στο επίπεδο ζ = 0, αν κάποιο $\bar{x}$ βγει εκτός του περιορισμού, το προβάλλουμε πάνω στο παραλληλόγραμμο. Εάν δηλαδή ένα σημείο βρίσκεται εκτός περιορισμού, θέτουμε την συντεταγμένη, η οποία βγήκε εκτός περιορισμού, ίση με το φράγμα το οποίο ξεπέρασε. Η τεχνική αυτή, λειτουργεί μόνο και μόνο επειδή ο περιορισμός μας είναι παραλληλόγραμμο.

Θα εξετάσουμε τώρα κάποιες ειδικές περιπτώσεις. Για την μαθηματική τεκμηρίωση των συμπερασμάτων μας, η λογική που θα ακολουθήσουμε είναι η ακόλουθη.

Όταν η τιμή της αντικειμενικής μας συνάρτησης καταλήγει να ταλαντώνεται εντός των περιορισμών για την προβολή του σημείου x_k θα ισχύει:

$$\bar{x_k} = Pr_X\{x_k - s_k \nabla f(x_k)\} = x_k - s_k \nabla f(x_k)$$

Επομένως, θα έχουμε:

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

Παρατηρούμε, δηλαδή, ότι εντός των περιορισμών η μέθοδος μέγιστης καθόδου με προβολή εκφυλίζεται στην απλή μέθοδο μέγιστης καθόδου με $\gamma_k' = \gamma_k s_k$.

Από την ανάλυση που προηγήθηκε συμπεραίνουμε ότι ο αλγόριθμος συγκλίνει για:

$$0 < \gamma_k' \leq \frac{1}{3} \Leftrightarrow \boxed{0 < \gamma_k s_k \leq \frac{1}{3}}$$

Επιπλέον, για την ταχύτητα σύγκλισης του αλγορίθμου, από την σχέση

$$\frac{|x_{k+1}|}{|x_k|} \leq max\{|1-\gamma_k' m|, |1-\gamma_k' M|\}$$

είναι προφανές ότι το άνω φράγμα ελαχιστοποιείται όταν

$$1 - \gamma_k' m = \gamma_k' M - 1,$$

δηλαδή όταν

$$\gamma'_{k_{max}} = \frac{2}{m+M} \Leftrightarrow \gamma'_{k_{max}} = 0.3 \Leftrightarrow \boxed{\gamma_k s_k = 0.3}$$

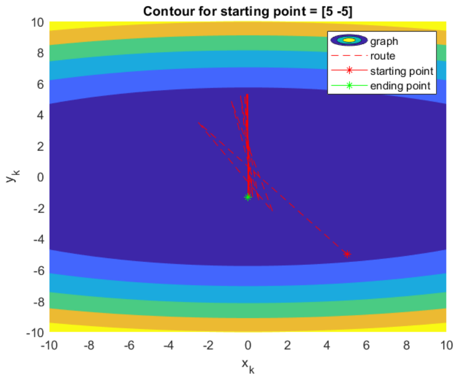
Θεωρητικά, λοιπόν, όσο πιο κοντά βρίσκεται το γινόμενο $\gamma_k s_k$ στην παραπάνω τιμή, τόσο μεγαλύτερη είναι η ταχύτητα σύγκλισης.

Ακολουθεί η μελέτη των ειδικών περιπτώσεων.

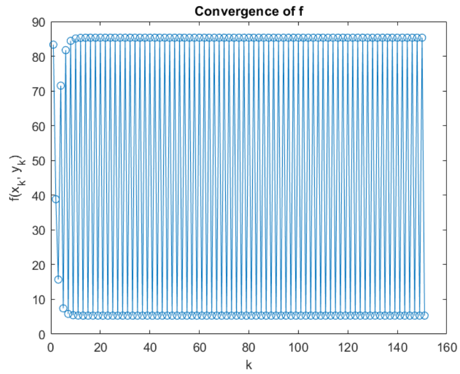
3.1 Για $s_k = 5$, $\gamma_k = 0.5$ και $(x_{1_0}, x_{2_0}) = (5, -5)$

Η τιμή της συνάρτησης δεν συγκλίνει στο ελάχιστο. Παρατηρούμε ταλαντώσεις οι οποιές βρίσκονται εντός των περιορισμών και κινούνται σε μια περιοχή σχετικά κοντά στο ελάχιστο, όχι όμως ικανοποιητικά κοντά. Ο αλγόριθμος δεν τερματίζει ποτέ, παρά μόνο με δική μας παρέμβαση στον κώδικα. Το αποτέλεσμα αυτό συνάδει με τους παραπάνω θεωρητικούς υπολογισμούς, καθώς το γινόμενο $\gamma_k s_k$ στην προκειμένη περίπτωση ισούται με $2.5 > 0.3$.

Σε σχέση με τους υπολογισμούς της ενότητας 2 (θέμα 1) παρατηρούμε ότι η ταλάντωση της τιμής της αντικειμενικής συνάρτησης είναι πιο περιορισμένη και τερματίζοντας τον αλγόριθμος δεν φτάνουμε σε απροσδιοριστία, αλλά παίρνουμε μία, έστω και λανθασμένη, τιμή ελαχίστου.



Σχήμα 10: Οπτικοποίηση της εύρεσης ελαχίστου για $s_k = 5, \gamma_k = 0.5$

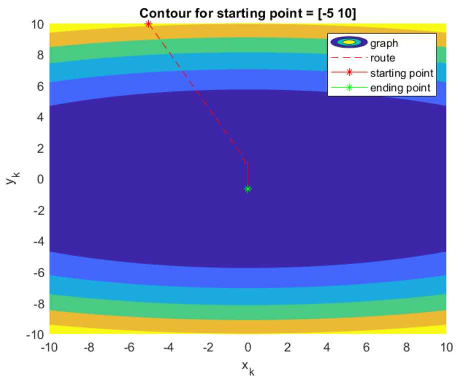


Σχήμα 11: Σύγκλιση της αντικειμενικής για $s_k = 5, \gamma_k = 0.5$

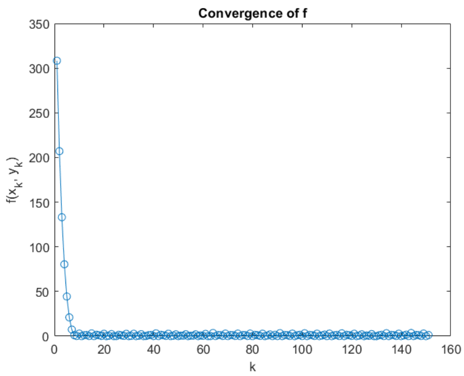
3.2 Για $s_k = 15$, $\gamma_k = 0.1$ και $(x_{1_0}, x_{2_0}) = (-5, 10)$

Η τιμή της συνάρτησης και πάλι δεν συγκλίνει στο ελάχιστο. Παρατηρούμε ταλαντώσεις οι οποιές βρίσκονται εντός των περιορισμών και κινούνται σε μια περιοχή σχετικά πιο κοντά στο ελάχιστο, όχι όμως και πάλι ικανοποιητικά κοντά. Ο αλγόριθμος δεν τερματίζει ποτέ, παρά μόνο με δική μας παρέμβαση στον κώδικα. Το αποτέλεσμα αυτό συνάδει με τους παραπάνω θεωρητικούς υπολογισμούς, καθώς το γινόμενο $\gamma_k s_k$ στην προκειμένη περίπτωση ισούται με $1.5 > 0.3$.

Σε σχέση με τα προηγούμενα παρατηρούμε ότι η ταλάντωση της τιμής της αντικειμενικής συνάρτησης είναι πιο περιορισμένη και η τελική τιμή ελαχίστου που παίρνουμε τερματίζοντας τον αλγόριθμο είναι αρκετά πιο κοντά στο πραγματικό ελάχιστο.



Σχήμα 12: Οπτικοποίηση της εύρεσης ελαχίστου για $s_k = 15, \gamma_k = 0.1$



Σχήμα 13: Σύγκλιση της αντικειμενικής για $s_k = 15, \gamma_k = 0.1$

Όπως προέκυψε από την παραπάνω θεωρητική ανάλυση η συνθήκη σύγκλισης της αντικειμενικής συνάρτησης είναι η:

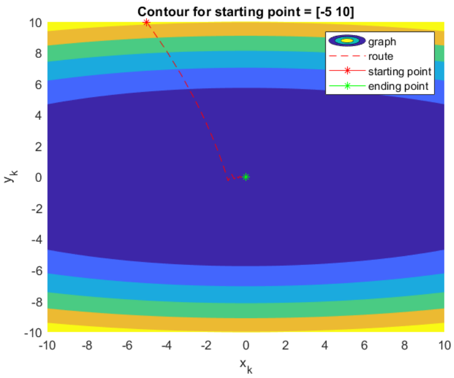
$$0 < \gamma_k s_k \le \tfrac{1}{3},$$

επομένως μπορούμε να προτείνουμε δύο απλούς πρακτικούς τρόπους, ώστε η μέθοδος να συγκλίνει στο ελάχιστο. Συγκεκριμένα:

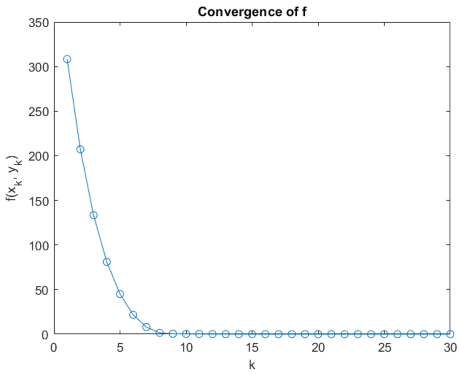
1. Διατηρούμε ως έχει το γ_k και μεταβάλλουμε το s_k.
   Τότε, από την σχέση της συνθήκης σύγκλισης θα έχουμε:

$$0.1s_k \leq \frac{1}{3} \Leftrightarrow s_k \leq \frac{10}{3}$$

   Επιλέγουμε $s_k = 3$ και βλέπουμε ότι η μέθοδος συγκλίνει στο ελάχιστο σε σχετικά μικρό αριθμό βημάτων. Μάλιστα, σύμφωνα με την συνθήκη μεγιστοποίησης της ταχύτητας σύγκλισης, αυτός ο αριθμός βημάτων είναι και ο μικρότερος δυνατός για τις συγκεκριμένες συνθήκες του προβήματος.



Σχήμα 14: Οπτικοποίηση της εύρεσης ελαχίστου για $s_k = 3, \gamma_k = 0.1$

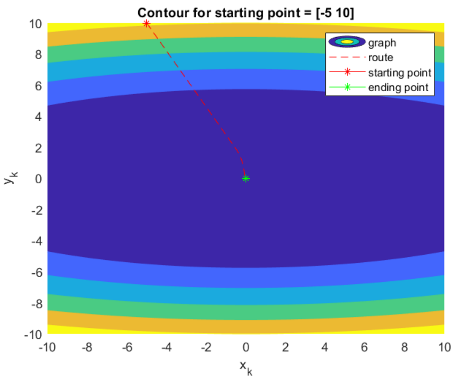


Σχήμα 15: Σύγκλιση της αντικειμενικής για $s_k = 3, \gamma_k = 0.1$

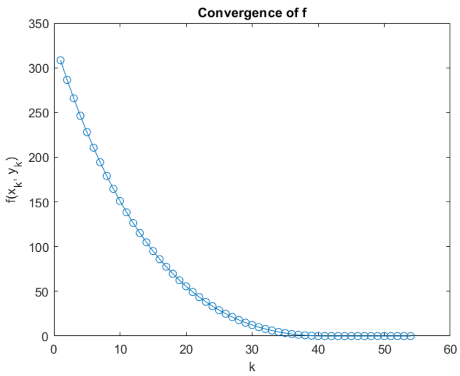
2. Διατηρούμε ως έχει το s_k και μεταβάλλουμε το γ_k.
   Τότε, από την σχέση της συνθήκης σύγκλισης θα έχουμε:

$$15\gamma_k \leq \frac{1}{3} \Leftrightarrow \gamma_k \leq \frac{1}{45}$$

   Επιλέγουμε $\gamma_k = 0.02$ και βλέπουμε ότι η μέθοδος συγκλίνει στο ελάχιστο σε σχετικά μικρό αριθμό βημάτων.



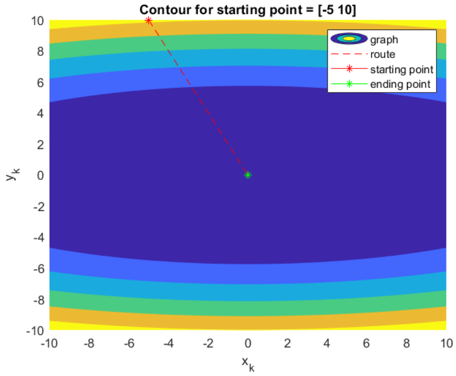
Σχήμα 16: Οπτικοποίηση της εύρεσης ελαχίστου για $s_k = 15, \gamma_k = 0.02$



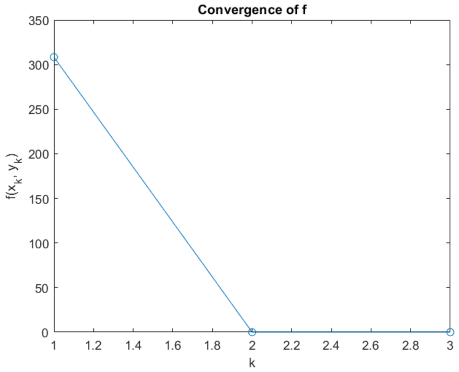
Σχήμα 17: Σύγκλιση της αντικειμενικής για $s_k = 15, \gamma_k = 0.02$

Παρατηρούμε πως εάν μεταβάλουμε το s_k έχουμε λιγότερα βήματα, αλλά έχουμε και ταλαντώσεις, ενώ αν μεταβάλουμε το γ_k, δεν έχουμε ταλαντώσεις αλλά έχουμε περισσότερα βήματα.

Σημείωση: Εάν για τον ρυθμό σύγκλισης γράψουμε για το βήμα x_k μια σχέση της μορφής $x_k = (1 - \alpha\gamma_k s_k)^k x_0$ με $\alpha > 0$ και λύσουμε την εξίσωση $\frac{df(x_{k+1})}{d\gamma s_k} = 0$ μπορούμε να βρούμε μία σχέση για το $\gamma_k s_k$ κάθε μίας από τις δύο μεταβλητές x_1, x_2, τέτοια ώστε η μέθοδος να συγκλίνει σε ένα μόνο βήμα. Συγκεκριμένα, βρίσκουμε $\gamma_{k_1} s_{k_1} = 1.5$ και $\gamma_{k_2} s_{k_2} = \frac{1}{6}$ και εφαρμόζοντάς τα στον αλγόριθμό μας έχουμε την ταχύτερη δυνατή σύγκλιση, όπως φαίνεται στα παρακάτω σχήματα.



Σχήμα 18: Οπτικοποίηση της εύρεσης ελαχίστου με την ταχύτερη δυνατή σύγκλιση.

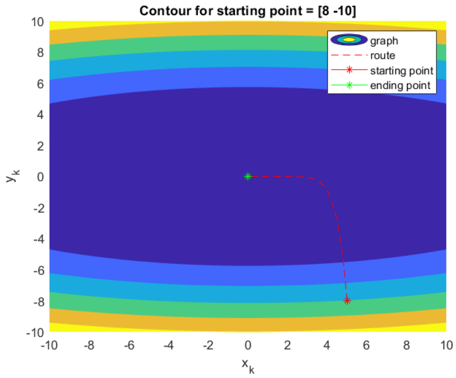


Σχήμα 19: Σύγκλιση της αντικειμενικής σε ένα μόνο βήμα.

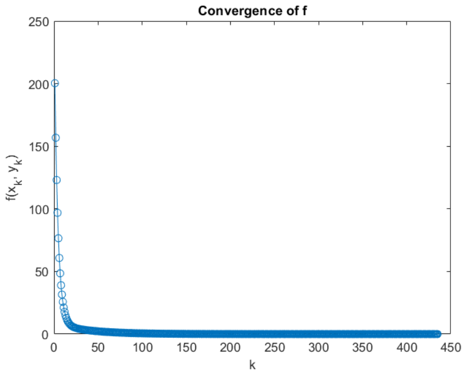
3.3 Για $s_k = 0.1$, $\gamma_k = 0.2$ και $(x_{1_0}, x_{2_0}) = (8, -10)$

Η τιμή της συνάρτησης συγκλίνει στο ελάχιστο για $(x_{1k}, x_{2k}) = (0.014755, -0.000000)$ με τιμή ελαχίστου 0.000073. Ο αριθμός των επαναλήψεων k είναι, ωστόσο, πολύ μεγάλος και αρκετά μεγαλύτερος από τις προηγούμενες περιπτώσεις στις οποίες είχαμε σύγκλιση. Παρατηρούμε, ακόμη, ότι δεν έχουμε καθόλου ταλαντώσεις.

Τα παραπάνω αποτελέσματα θα μπορούσαν να προβλεφθούν με βάση την θεωρητική ανάλυση που προηγήθηκε. Συγκεκριμένα, βλέπουμε ότι το γινόμενο $\gamma_k s_k$ ισούται με $0.02 < \frac{1}{3}$, επομένως είμαστε βέβαιοι ότι η μέθοδος θα συγκλίνει. Επιπλέον, βλέπουμε ότι $0.02 \ll 0.3$ κι άρα από την συνθήκη μεγιστοποίησης της ταχύτητας σύγκλισης της μεθόδου περιμένουμε πως ο αριθμός των βημάτων θα είναι πολύ μεγάλος. Τέλος, λόγω της μικρής τιμής του γ_k δεν περιμένουμε ταλαντώσεις.



Σχήμα 20: Οπτικοποίηση της εύρεσης ελαχίστου για $s_k = 0.1, \gamma_k = 0.2$



Σχήμα 21: Σύγκλιση της αντικειμενικής για $s_k = 0.1, \gamma_k = 0.2$

4 Συμπεράσματα - Σχόλια

Η μέθοδος μέγιστης καθόδου με προβολή είναι αποτελεσματική στην περίπτωση περιορισμών, οι οποίοι όμως δημιουργούν στον διανυσματικό χώρο έναν υπο-χώρο απλής μορφής, τέτοιας ώστε να υπάρχει αναλυτική λύση της προβολής των σημείων στον υπο-χώρο αυτό. Σε αντίθετη περίπτωση, σε κάθε επανάληψη με κάποιο x_k εκτός περιορισμών θα έπρεπε να λύνουμε ένα πρόβλημα βελτιστοποίησης της μορφής $min_{x \in X}|z - x|^2$, το οποίο είναι εξαιρετικά ασύμφωρο υπολογιστικά.

Σε περιπτώσεις, ωστόσο, όπως αυτή που μελετήσαμε παραπάνω, όπου ο περιορισμός είναι ένα παραλληλόγραμο στο επίπεδο, η προβολή υπολογίζεται αναλυτικά και η μέθοδος είναι υπολογιστικά συμφέρουσα. Επιπλέον, έχουμε αρκετή ελευθερία, με βάση την θεωρητική ανάλυση που διεκπαιρεώσαμε, να τροποποιήσουμε τις παραμέτρους sk και γ_k, έτσι ώστε να πετύχουμε την ταχύτερη δυνατή σύγκλιση με τις λιγότερες δυνατές ταλαντώσεις. Οι ταλαντώσεις σε άλλο παράδειγμα όπου μπορεί να υπήρχαν γειτονικά τοπικά ελάχιστα, θα μπορούσαν να εγκλωβίσουν τον αλγόριθμο σε κάποιο μη επιθυμητό σημείο. Επομένως, η θεωρητική ανάλυση είναι ένα πολύ βοηθητικό εργαλείο.