Τεχνικές Βελτιστοποίησης

Μέθοδος Μέγιστης Καθόδου με Προβολή

3η Εργαστηριακή Άσκηση

Ονοματεπώνυμο: **Ρεπάνης Γεώργιος – Δημήτριος**

Δεκέμβριος 2025

ΑΕΜ: 10588

Θέμα 1

Στο πρώτο θέμα μελετάμε την εφαρμογή της Μέγιστης Καθόδου σε μια κυρτή τετραγωνική συνάρτηση, $f(x)=\frac{1}{3}x_1^2+3x_2^2$, χωρίς περιορισμούς. Στόχος είναι να εξετάσω πώς η επιλογή ενός σταθερού βήματος γ επηρεάζει την σύγκλιση της μεθόδου προς το μοναδικό τοπικό ελάχιστο στο σημείο (0,0), ξεκινώντας από οποιοδήποτε σημείο διαφορετικό του μηδενός. Για κάθε τιμή βήματος υπολογίζουμε αριθμητικά τον αλγόριθμο, απεικονίζουμε σε διάγραμμα την σύγκλιση της αντικειμενικής συνάρτησης ως προς τον αριθμό των επαναλήψεων και στη συνέχεια σχολιάζουμε τα αποτελέσματα.

Θεωρώ:

$$f:\Re\rightarrow\Re, \qquad f(x)=\frac{1}{3}x_1^2+3x_2^2, \qquad x=\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

<u>Μέθοδος Μεγίστης Καθόδου (με σταθερό βήμα γ)</u>

Γενικά:

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

όπου $\gamma_k>0$ το βήμα

Υπολογίζω τα Gradient, Hessian και ελάχιστο της f:

Gradient: Υπολογίζω $\frac{df}{dx_1}=\frac{2}{3}x_1$ και $\frac{df}{dx_2}=6x_2$ άρα $\nabla f(x)=\begin{bmatrix}\frac{2}{3}x_1\\6x_2\end{bmatrix}$.

Hessian: $\nabla^2 f(x)=\begin{bmatrix}\frac{2}{3} & 0\\0 & 6\end{bmatrix}=Q$

Ο πίνακας Q είναι θετικά ορισμένος με ιδιοτιμές $\lambda_1=\frac{2}{3}>0\,,\lambda_2=6>0$ άρα η f είναι γνήσια κυρτή και έχει μοναδικό ελάχιστο στο σημείο όπου $\nabla f(x)=0$ δηλαδή x*=(0,0).

i) Για γ=0,1 έχω:

Γενικός τύπος της μεθόδου:

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

Επειδή Q είναι διαγώνιος, η μέθοδος σπάει σε συνιστώσες:

$$x_{1,k+1}=x_{1,k}-0.1\frac{2}{3}x_{1,k}=\frac{14}{15}x_{1,k}$$

$$x_{2,k+1}=x_{2,k}-0.1\,6\,x_{2,k}=\frac{2}{5}x_{2,k}$$

Άρα

$$x_{1,k}=\left(\frac{14}{15}\right)^k x_{1,0} \text{ και } x_{2,k}=\left(\frac{2}{5}\right)^k x_{2,0}$$

Για οποιοδήποτε αρχικό σημείο $x_0 \neq (0,0)$.

Για την σύγκλιση της ακολουθίας x_k. Αρχικά ισχύει ότι $\frac{14}{15} < 1$ και $\frac{2}{5} < 1$ άρα:

$$\lim_{k \to \infty} x_{1,k} = \lim_{k \to \infty} \left(\frac{14}{15}\right)^k x_{1,0} = 0 \text{ και } \lim_{k \to \infty} x_{2,k} = \lim_{k \to \infty} \left(\frac{2}{5}\right)^k x_{2,0} = 0$$

άρα $\lim_{k \to \infty} x_k = (0,0) = x^{¿}$

Ισοδύναμα έχω:

$$x_{k+1} = (I - \gamma Q) x_k, \text{ όπου } I - \gamma Q = \begin{bmatrix} 1 - \gamma \frac{2}{3} & 0 \\ 0 & 1 - 6\gamma \end{bmatrix} = \begin{bmatrix} \frac{14}{15} & 0 \\ 0 & \frac{2}{5} \end{bmatrix}$$

Οι ιδιοτιμές του πίνακα είναι $\mu_1 = \frac{14}{15}, \mu_2 = \frac{2}{5}$

Εφόσον $\mu_i < 1, i = 1,2$ η ακολουθία x_k συγκλίνει στο μηδέν για κάθε αρχικό σημείο.

Γενικά μια τετραγωνική συνάρτηση $f(x) = \frac{1}{2} x^T Q x$ η μέθοδος με σταθερό βήμα συγκλίνει όταν:

$$0 < \gamma < \frac{2}{\lambda_{max}(Q)}$$

Εδώ $\lambda_{max}(Q) = 6$, άρα

$$0 < \gamma < \frac{2}{6} = \frac{1}{3}$$

Το γ εδώ είναι 0.1 ικανοποιεί την ανισότητα άρα θα έχουμε σύγκλιση.

Υπολογισμός βημάτων για ακρίβεια ε=0.001.

Αν ως κριτήριο τερματισμού πάρουμε:

$$\|x_k - x^{¿}\| = \|x_k\| \leq \varepsilon$$

Τότε:

$$\|x_k\|^2 = x_{1,k}^2 + x_{2,k}^2 = \left(\frac{14}{15}\right)^{2k} x_{1,0}^2 + \left(\frac{2}{5}\right)^{2k} x_{2,0}^2 \leq \left(\frac{14}{15}\right)^{2k} (x_{1,0}^2 + x_{2,0}^2) = \left(\frac{14}{15}\right)^{2k} \|x_0\|^2$$

Επειδή $\frac{14}{15} > \frac{2}{5}$

Άρα αρκεί

$$\left(\frac{14}{15}\right)^k \leq \frac{\varepsilon}{\|x_0\|} = \frac{0.001}{\|x_0\|}$$

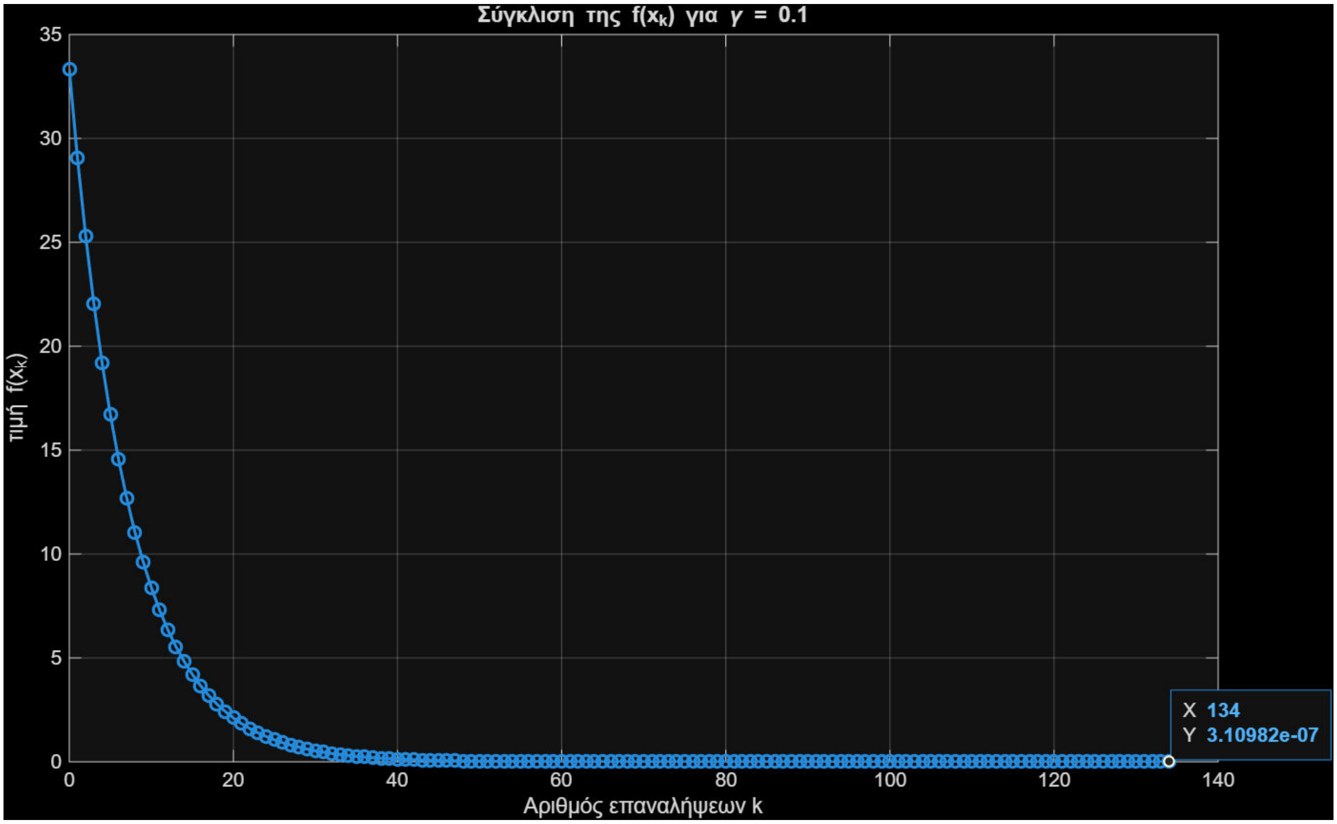
Λύνοντας ως προς k:

$$k \geq \frac{\ln(\varepsilon / \|x_0\|)}{\ln(14/15)}$$

για $\gamma = 0.1, x_0 = [10,0]^T$

$$k \geq \frac{\ln(10^{-4})}{\ln(14/15)} \approx 133.5$$

Ακολουθεί το διάγραμμα για γ=0.1, k=134, ε=0.001, $x_0 = (10,0)$



<u>Σχολιασμός:</u>

Στο k=0 η τιμή είναι $f(x_0) = \frac{1}{3} 10^2 \approx 33.3$. Στις πρώτες 20 επαναλήψεις βλέπουμε μια γρήγορη πτώση της f το οποίο είναι λογικό εξαιτίας του μεγάλου συντελεστή σύγκλισης της συνάρτησης που είναι 14/15 δηλαδή περίπου 0.93.

Μετά από 30-40 επαναλήψεις η f γίνεται μικρότερη από 1 και πλησιάζοντας στο τελικό ζητούμενο σημείο σύγκλισης χρειαζόμαστε όλο και περισσότερα βήματα για μικρές βελτιώσεις.

Το γράφημα φτάνει μέχρι k=134 επαναλήψεις όπου ο αλγόριθμος σταματά διότι ικανοποιείται το κριτήριο $\|x_{134}\| \leq 0.001$. Στο ίδιο σημείο η f έχει την τιμή $3.1 * 10^{-7}$ δηλαδή πρακτικά πολύ κοντά στο θεωρητικό ελάχιστο 0.

Συμπέρασμα το διάγραμμα για γ=0.1 επιβεβαιώνει την θεωρητική ανάλυση, η μέθοδος συγκλίνει μονοτονικά στο ολικό ελάχιστο με σταθερή γραμμική σύγκλιση και χρειάζεται 134 επαναλήψεις για να εκπληρώσει την ακρίβεια ε=0.001.

ii) Για γ=0.3 έχω:

$$1 - \frac{2}{3}\gamma = 1 - \frac{2}{3}0.3 = 1 - 0.2 = 0.8$$

και

$$1 - 6\gamma = 1 - 6\,0.3 = 1 - 1.8 = -0.8$$

Άρα ο αλγόριθμος επανάληψης γράφεται:

$$x_{1,k+1}=0.8\,x_{1,k} \text{ και } x_{2,k+1}=-0.8\,x_{2,k}$$

Λύνω τις αναδρομές

$$x_{1,k}=0.8\,x_{1,0} \text{ και } x_{2,k}=-0.8\,x_{2,0}$$

Άρα

$$\lim_{k\to\infty} x_{1,k}=0 \text{ και } \lim_{k\to\infty} x_{2,k}=0$$

οπότε

$$\lim_{k\to\infty} x_k=(0,0)=x^{¿}$$

Παρατηρώ ότι και οι δύο συντελεστές έχουν μέτρο <1 άρα η μέθοδος συγκλίνει για οποιοδήποτε αρχικό σημείο $x_0\neq 0$.

Η αντικειμενική συνάρτηση στο βήμα k:

$$f(x_k)=\frac{1}{3}x_{1,k}^2+3x_{2,k}^2=\frac{1}{3}\left(0.8^{2k}x_{1,0}^2\right)+3\left((-0.8)^{2k}x_{2,0}^2\right)=0.8^{2k}\left(\frac{1}{3}x_{1,0}^2+3x_{2,0}^2\right)$$

Κριτήριο ακρίβειας ε=0.001:

$$\|x_k\|^2=x_{1,k}^2+x_{2,k}^2=0.8^{2k}\left(x_{1,0}^2+x_{2,0}^2\right)=0.8^{2k}\|x_0\|^2$$

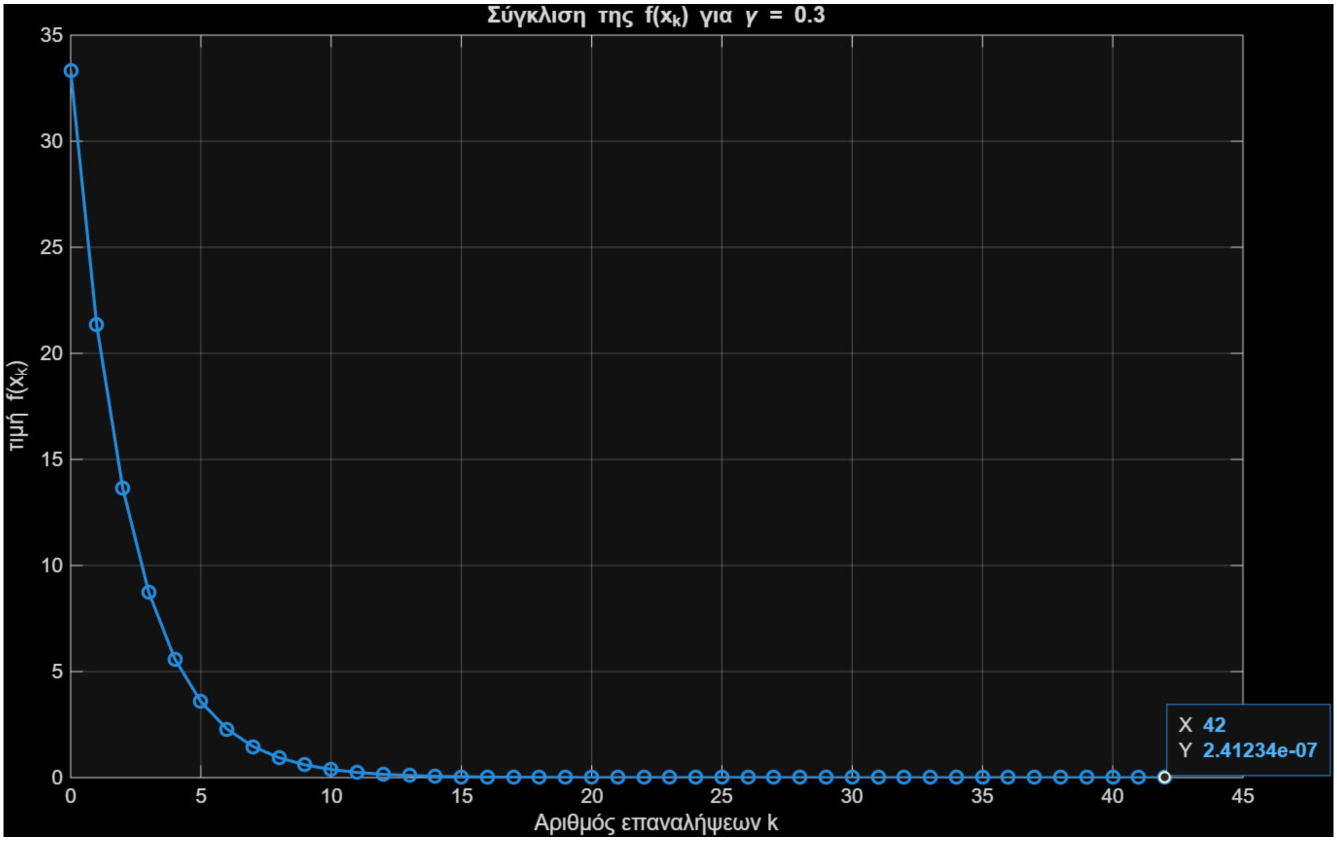
Για να ισχύει $\|x_k\|\leq\varepsilon$, αρκεί:

$$0.8^k\leq\frac{\varepsilon}{\|x_0\|}\Rightarrow k\geq\frac{\ln(\varepsilon/\|x_0\|)}{\ln(0.8)}$$

για $x_0=(10,0)$

$$k\geq\frac{\ln(10^{-4})}{\ln(0.8)}\approx 42$$

Ακολουθεί το διάγραμμα για γ=0.3, k=42, ε=0.001, $x_0=(10,0)$



Σχολιασμός:
Για k=0 ξεκινάω πάλι από το 33.3.
Οι πρώτες 7-10 επαναλήψεις δείχνουν πολύ απότομη πτώση της f. Η καμπύλη πέφτει πιο γρήγορα σε σχέση με γ=0.1).
Η μέθοδος σταματά όταν ικανοποιείται το κριτήριο $\|x_k\| \leq \varepsilon$ και αυτό συμβαίνει στην επανάληψη 42.
Δηλαδή για μεγαλύτερο γ η μέθοδος συγκλίνει πιο γρήγορα στο ελάχιστο κοντά στο 0

iii) Για γ=3 έχω:

$$1-\frac{2}{3}\gamma=1-\frac{2}{3}3=1-2=1$$

και

$$1-6\gamma=1-6*3=1-18=-17$$

άρα

$$x_{1,k+1}=-x_{1,k} \Rightarrow x_{1,k}=(-1)^k x_{1,0}$$

και

$$x_{2,k+1}=-17x_{2,k} \Rightarrow x_{2,k}=(-17)^k x_{2,0}$$

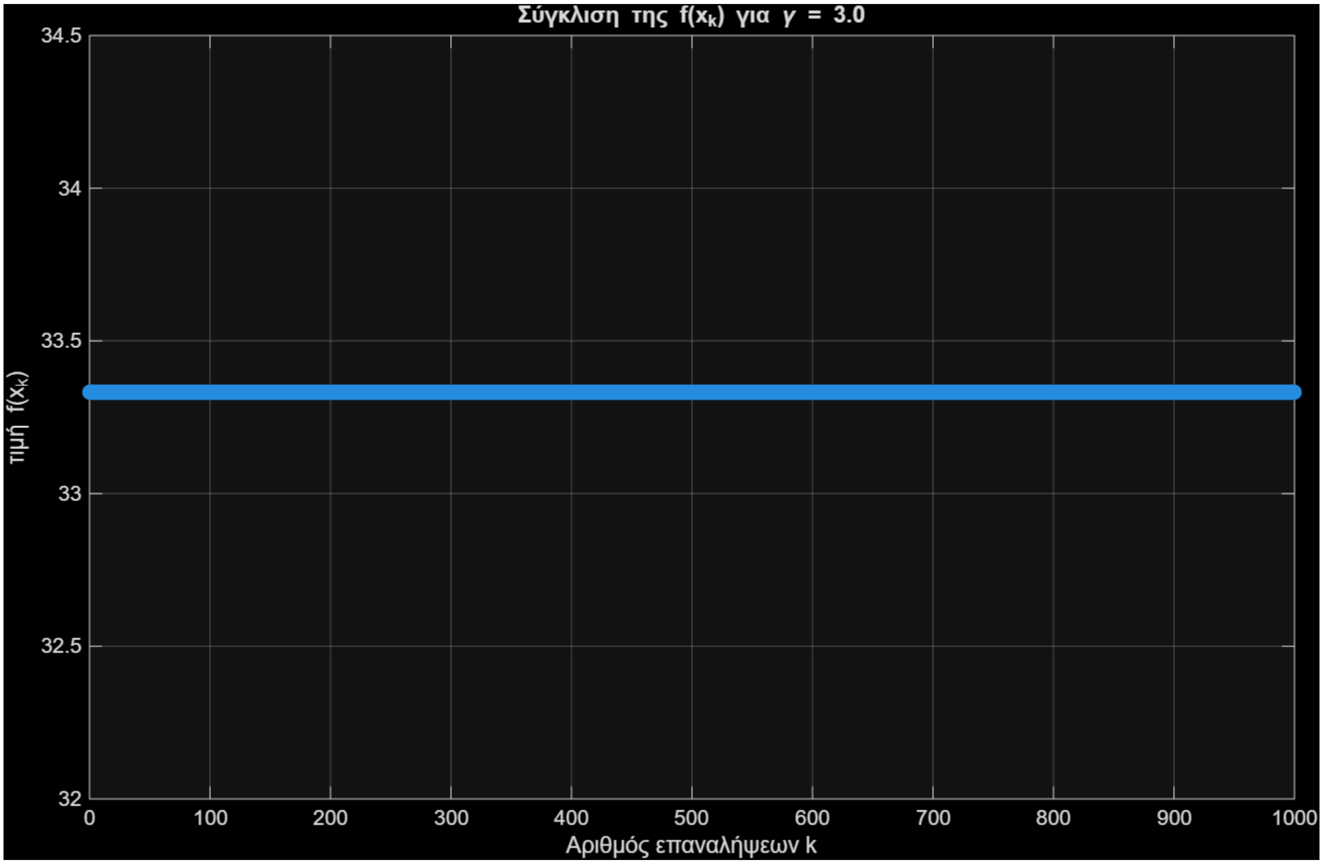
Τότε αν $x_{2,0} \neq 0$ το μέτρο του $x_{2,k}$ μεγαλώνει λόγω του συντελεστή 17 άρα

$$\|x_k\| \rightarrow \infty, f(x_k) \rightarrow \infty$$

Οπότε για γ=3 η μέθοδος ΔΕΝ συγκλίνει στο ελάχιστο για κανένα αρχικό σημείο.
Αυτό συμφωνεί και με τον γενικό κανόνα

$$0<\gamma<\frac{2}{\lambda_{max}(Q)}=\frac{2}{6}=\frac{1}{3} \text{ όπου εδώ έχω γ=3>1/3}$$

Ακολουθεί το διάγραμμα για γ=3, $k \rightarrow \infty$, ε=0.001, $x_0=(10{,}0)$



<u>Σχολιασμός:</u>
Βλέπουμε ότι η καμπύλη της f είναι οριζόντια στο αρχικό σημείο 33.3 για κάθε k. Δηλαδή η μέθοδος δεν μειώνει καθόλου την τιμή της συνάρτησης ακόμη και για 1000 επαναλήψεις. Ειδικότερα:
Με $x_0=(10{,}0) \Rightarrow x_{2,k}=0$ και $x_{1,k}=(-1)^k 10$ άρα το σημείο ταλαντώνεται μεταξύ (-10,0) και (10,0) χωρίς να πλησιάζει το (0,0) και

$$f(x_k)=\frac{1}{3}x_{1,k}^2=\frac{100}{3}\approx 33{,}3 \text{ παραμένει σταθερή}$$

Επιπλέον το κριτήριο ακρίβειας $\|x_k\| \leq \varepsilon$ δεν ικανοποιείται ποτέ αφού η $\|x_k\|$ μένει ίση με 10 σε όλες τις επαναλήψεις.
Άρα ο αλγόριθμος δεν συγκλίνει ποτέ στο ελάχιστο.

iv) Για γ=5 έχω:

$$1-\frac{2}{3}\gamma=1-\frac{2}{3}*5=1-\frac{10}{3}=-\frac{7}{3}$$

και

$$1-6\gamma=1-6*5=1-30=-29$$

Άρα

$$x_{1,k+1}=-\frac{7}{3}x_{1,k} \Rightarrow x_{1,k}=\left(-\frac{7}{3}\right)^k x_{1,0}$$

και

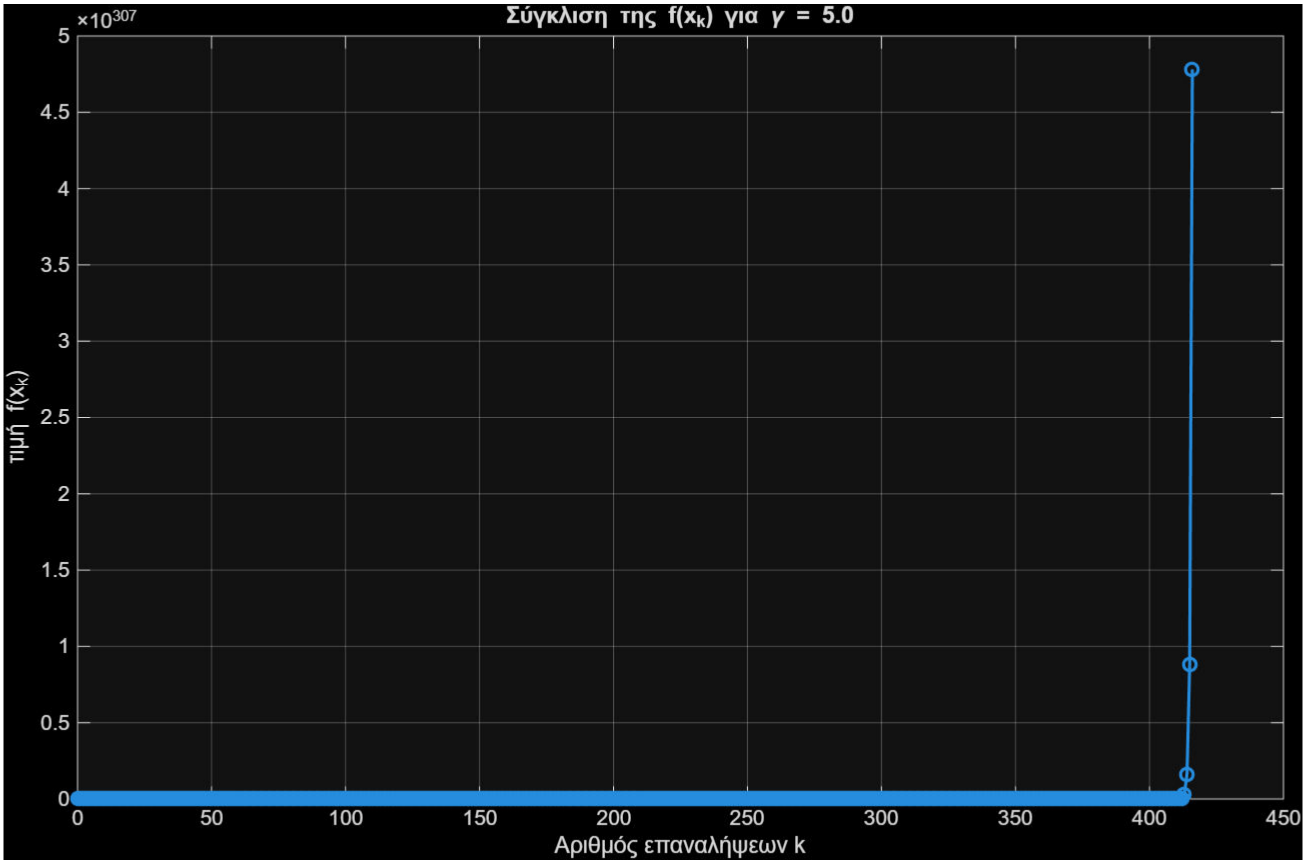
$$x_{2,k+1}=-29\,x_{2,k} \Rightarrow x_{2,k}=(-29)^k x_{2,0}$$

Και οι δύο συντελεστές έχουν μέτρο μεγαλύτερο από 1, άρα για οποιοδήποτε $x_0 \neq 0$, ισχύει:

$$\|x_k\| \rightarrow \infty \text{ και } f(x_k) \rightarrow \infty$$

Συμπέρασμα για γ=5 η μέθοδος εκτρέπεται τελείως χωρίς καμία περίπτωση να συγκλίνει με το (0,0).

Ακολουθεί το διάγραμμα για γ=5, $k \rightarrow \infty$, ε=0.001, $x_0=(10{,}0)$



<u>Σχολιασμός:</u>

Με αρχικό σημείο $x_0=(10{,}0)$ μένει $x_{2,k}=0$ και $x_{1,k}=\left(-\frac{7}{3}\right)^k 10$ και επομένως $f(x_k)=\frac{100}{3}\left(\frac{49}{9}\right)^k$.

Προφανώς για k=0, f=33.3 αλλά λόγω κλίμακας φαίνεται πρακτικά 0.

Η f αυξάνεται εκθετικά από το πρώτο βήμα.

Στην φαινομενικά τελευταία επανάληψη ουσιαστικά σπάει το πρόγραμμα και το σημείο πρακτικά απειρίζει. (τάξη 10^{307} και πολύ κοντά στο $double \approx 10^{308}$

Δεν ικανοποιείται ποτέ το κριτήριο $\|x_k\| \leq \varepsilon$ διότι $\|x_k\| = 10\, \forall\, k$ και δεν παίζει ρόλο το αρχικό σημείο αφού με οποιοδήποτε θα απειρίσει.

Γενικό συμπέρασμα:

Από θεωρία γνωρίζουμε ότι μία τετραγωνική πραγματική συνάρτηση $f(x)=\frac{1}{2}x^T Q x$, το σταθερό βήμα πρέπει να ικανοποιεί : $1<\gamma<\frac{2}{\lambda_{max}(Q)}=\frac{1}{3}$, άρα

γ=0,1: σύγκλιση αλλά σχετικά αργή.

γ=0.3: σύγκλιση αλλά πιο γρήγορη.

γ=3 και γ=5: $|1-\gamma\lambda_i|\geq 1$ για κάποια ιδιοτιμή λ_i άρα η μέθοδος αποκλίνει.

ΘΕΜΑ 2

Στο δεύτερο θέμα εξετάζουμε την Μέθοδο Μεγίστης Καθόδου με Προβολή για την ίδια κυρτή τετραγωνική συνάρτηση $f(x)=\frac{1}{3}x_1^2+3x_2^2$ με περιορισμούς $-10\leq x_1\leq 5$ και $-8\leq x_2\leq 12$. Ξεκινώντας από το σημείο που μα δίνεται (5,-5) και σταθερές παραμέτρους $s_k=5$ και $\gamma_k=0.5$ εφαρμόζω gradient projection δηλαδή κάνω ένα βήμα κατά μήκους του αρνητικού gradient και μετά προβάλουμε το σημείο στην περιοχή που μας δίνει ο περιορισμός. Στο τέλος θα συγκρίνουμε τα αποτελέσματα και πώς το βήμα και οι περιορισμοί επηρέασαν την συμπεριφορά του αλγορίθμου. Με το διάγραμμα f συναρτήση του k θα συγκρίνουμε τα ποιοτικά χαρακτηριστικά με την μέθοδο χωρίς περιορισμούς.

Περιορισμός:

$$X=\{x\in\Re^2 : -10\leq x_1\leq 5, -8\leq x_2\leq 12\}$$

Ο χωρίς περιορισμός ελάχιστος είναι το (0,0). Αυτό ανήκει στο εσωτερικό του Χ, άρα το ελάχιστο με περιορισμούς είναι επίσης (0,0).

Gradient:

$$\nabla f(x)=\begin{bmatrix}\frac{2}{3}x_1\\ 6x_2\end{bmatrix}$$

Προβολή στην περιοχή Χ:
Η προβολή $P_X(z)$ ενός σημείου $z=(z_1,z_2)$ στην περιοχή Χ είναι:

$$P_X(z)=\begin{bmatrix}min\{5, max\{-10, z_1\}\}\\ min\{12, max\{-8, z_2\}\}\end{bmatrix}$$

Δηλαδή αν μια συνιστώσα βγει εκτός ορίων την επαναφέρουμε στο πλησιέστερο άκρο ενώ αν είναι εντός την αφήνουμε όπως είναι.

Μέθοδος Μέγιστης Καθόδου με Προβολή:
Υπολογίζω δοκιμαστικού βήματος καθόδου:

$$y_k=x_k-s\nabla f(x_k)$$

Προβάλω το σημείο στην περιοχή Χ:

$$z_k = P_X(y_k)$$

Ορίζω την διεύθυνση:

$$d_k = z_k - x_k$$

Κάνω το τελικό βήμα:

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

<u>Αριθμητική Εφαρμογή:</u>
Μας δίνονται

$$s_k = 5,\ \ \gamma_k = 0.5,\ x_0 = (5, -5) \text{ και } \varepsilon = 0.01$$

Πρώτα βήματα του αλγορίθμου από το $x_0 = (5, -5)$:

Βήμα $0 \rightarrow 1$
Gradient στο x_0:

$$\nabla f(x_0) = \begin{bmatrix} \frac{2}{3} * 5 \\ 6 * (-5) \end{bmatrix} = \begin{bmatrix} \frac{10}{3} \\ -30 \end{bmatrix}$$

Δοκιμαστικό βήμα:

$$y_0 = x_0 - 5\nabla f(x_0) = \begin{bmatrix} 5 \\ -5 \end{bmatrix} - 5 \begin{bmatrix} \frac{10}{3} \\ -30 \end{bmatrix} \approx (-11.67, 145)$$

Προβολή στο X:

Στο x_1: -11.67<10 άρα κόβεται στο 10
Στο x_2: 145>12 άρα κόβεται στο 12

$$z_0 = P_X(y_0) = (-10, 12)$$

Διεύθυνση:

$$d_0 = z_0 - x_0 = (-10 - 5, 12 - (-5)) = (-15, 17)$$

Τελικό βήμα:

$$x_1 = x_0 + 0.5 d_0 = (5, -5) + 0.5(-15, 17) = (-2.5, 3.5)$$

Το x_1 είναι εντός του X.

Βήμα $1 \rightarrow 2$
Gradient στο x_1:

$$\nabla f(x_1) = \begin{bmatrix} \frac{2}{3} * (-2.5) \\ 6 * 3.5 \end{bmatrix} = \begin{bmatrix} -\frac{5}{3} \\ 21 \end{bmatrix}$$

Δοκιμαστικό βήμα:

$$y_1 = x_1 - 5\nabla f(x_1) = \begin{bmatrix} -2.5 \\ 3.5 \end{bmatrix} - 5\begin{bmatrix} -\frac{5}{3} \\ 21 \end{bmatrix} \approx (5.85, -101.5)$$

Προβολή στο X:
Στο x_1: 5.85>5 άρα κόβεται στο 5
Στο x_2: -101.5←8 άρα κόβεται στο -8

Διεύθυνση:

$$d_1 = z_1 - x_1 = (5 - (-2.5), -8 - 3.5) = (7.5, -11.5)$$

Τελικό βήμα:

$$x_2 = x_1 + 0.5\, d_1 = (-2.5, 3.5) + 0.5(7.5, -11.5) = (1.25, -2.25)$$

Το x_2 είναι εντός του X.

Αν συνεχίσουμε τις επαναλήψεις τα $x_{1,k}$ τείνουν στο 0 ενώ τα $x_{2,k}$ ταλαντώνονται μεταξύ δύο τιμών:

$$x_{2,2n} \rightarrow -\frac{4}{3} \text{ και } x_{2,2n+1} \rightarrow \frac{16}{3}$$

Δηλαδή τα σημεία:

$$x^{(A)} = \left(0, -\frac{4}{3}\right) \text{ και } x^{(B)} = \left(0, \frac{16}{3}\right)$$

Δηλαδή η ακολουθία x_k εγκλωβίζεται σε μία επαναλαμβανόμενη ταλάντωση από $x^{(A)} \rightarrow x^{(B)} \rightarrow x^{(A)} \rightarrow x^{(B)} \ldots$ μέσα στην περιοχή X.

Η αντικειμενική συνάρτηση :

$$f\left(0, -\frac{4}{3}\right) = 3\left(\frac{4}{3}\right)^2 = \frac{16}{3} \approx 5.33$$

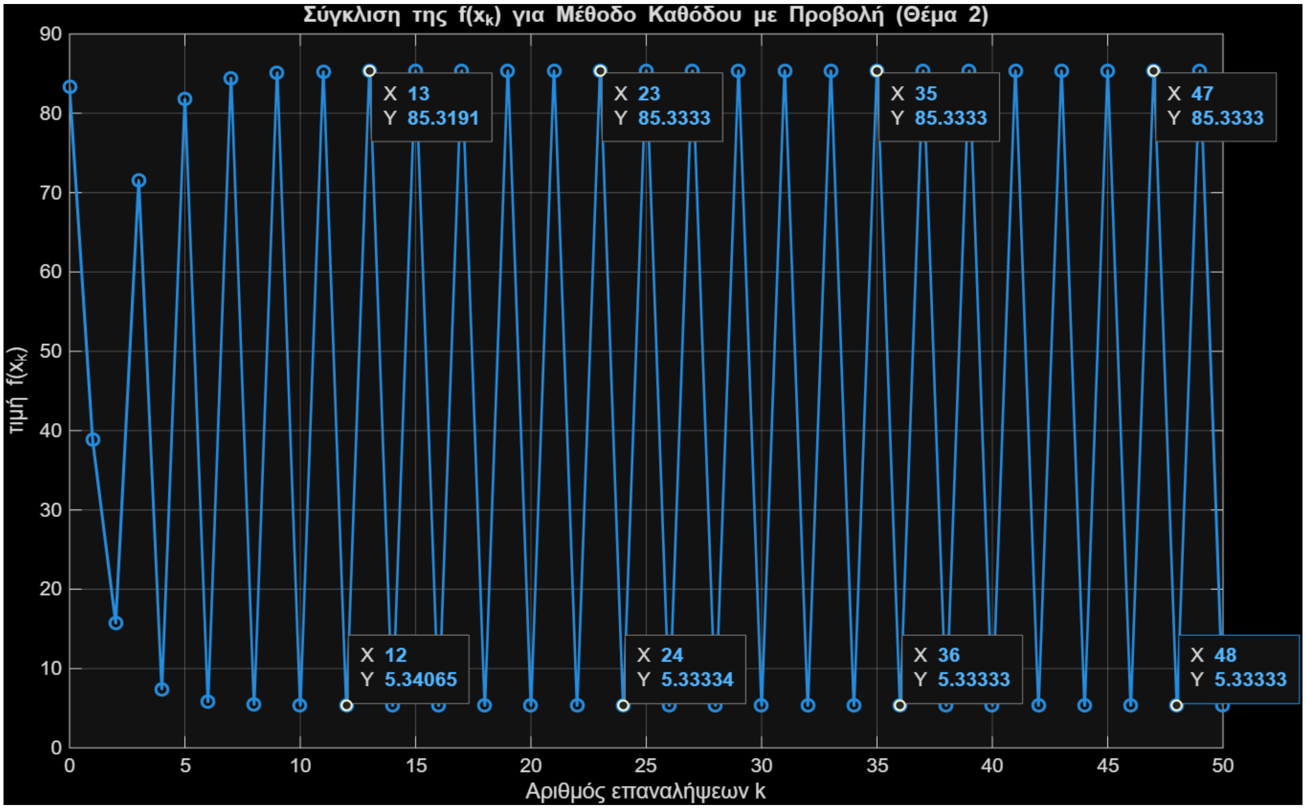
και

$$f\left(0, \frac{16}{3}\right) = 3\left(\frac{16}{3}\right)^2 = \frac{768}{3} \approx 85.33$$

οπότε το γράφημα της f θα πηγαίνει από την μια θετική τιμή στην άλλη και δεν θα πέφτει προς το 0.

<u>Κριτήριο ακρίβειας ε=0.01</u>
Το κριτήριο ακρίβειας προφανώς δεν ικανοποιείται ποτέ αφού ως προς το πραγματικό ελάχιστο η f δεν συγκλίνει ποτέ.

Ακολουθεί η προσομοίωση με Μέθοδο Μέγιστης Καθόδου με Προβολή με $s_k = 5, \gamma_k = 0.5, x_0 = (5, -5), \varepsilon = 0.01$ της f συναρτήσει των επαναλήψεων k.



Σχολιασμός:

Το διάγραμμα είναι απόλυτα συμβατό με την θεωρητική μας ανάλυση. Η Μέθοδος με προβολή δεν συγκλίνει στο ελάχιστο. Μετά τα πρώτα 10 βήματα η τιμή της f σταθεροποιείται σε μία περιοδική ταλάντωση ανάμεσα στα 5.33 (χαμηλή τιμή) και 85.33 (υψηλή τιμή). Βλέπουμε τα δύο όρια και την f να πηγαίνει πάνω-κάτω μεταξύ τους και τον αλγόριθμο να παγιδεύεται μέσα στην περιοχή X χωρίς να μπορεί να πλησιάσει την f=0.

Σύγκριση Μεθόδου Μέγιστης Καθόδου χωρίς περιορισμούς και Μεθόδου Μέγιστης Καθόδου με Προβολή και περιορισμούς

Στην Μέθοδο Μέγιστης Καθόδου χωρίς περιορισμούς, για καλά βήματα (γ=0.1 , γ=0.3) τα διαγράμματα της f είναι μονοτονικά φθίνοντα και συγκλίνουν στο 0 και η απόσταση από το ελάχιστο $\|x_k\|$ τελικά γίνεται μικρότερη από την ακρίβεια ε=0.001. Δηλαδή το κριτήριο ικανοποιείται μετά από πεπερασμένο αριθμό επαναλήψεων (πχ k=134 για γ=0.1) και ο αλγόριθμος σταματά πολύ κοντά στο μηδέν.

Στην Μέθοδο Μέγιστης Καθόδου με Προβολή και περιορισμούς αν και το (0,0) είναι εφικτό-ανήκει δηλαδή στο X- η επιλεγμένη παράμετρος s=5 και γ=0.5 κάνει την μέθοδο να ταλαντώνεται μεταξύ δύο τιμών f=5.33 και f=85.33. Η x_k παραμένει αρκετά μεγαλύτερη από το ε=0.01 και το κριτήριο ακρίβειας δεν ικανοποιείται ποτέ, άρα ο αλγόριθμος δεν συγκλίνει ποτέ στο ελάχιστο σε αντίθεση με την επιτυχημένη σύγκλιση του πρώτου θέματος.

ΘΕΜΑ 3

Στο τρίτο θέμα μελετάμε ξανά την Μέθοδο Μεγίστης Καθόδου με Προβολή για την ίδια τετραγωνική συνάρτηση $f(x)=\frac{1}{3}x_1^2+3x_2^2$ με τους ίδιους περιορισμούς $-10\leq x_1\leq 5$ και $-8\leq x_2\leq 12$ αλλά με διαφορετικές παραμέτρους $s_k=15, \gamma_k=0.1$ και άλλο αρχικό σημείο $x_0=(-5{,}10)$ με ζητούμενη ακρίβεια ε=0.01. Στόχος είναι να δούμε αν η επιλογή ενός μεγαλύτερου s και πιο μικρού τελικού βήματος γ επιτρέπει στην μέθοδο να ξεπεράσει τα προβλήματα ταλάντωσης που είχαμε στην προηγούμενη περίπτωση -θεμα2- και να οδηγήσει τελικά σε προσέγγιση του ελαχίστου (0,0) μέσα στην περιοχή X.

Βήμα 0➔1
Αρχικό σημείο:

$$x_0=(-5{,}10)$$

Gradient:

$$\nabla f(x_0)=\begin{bmatrix}\frac{2}{3}*(-5)\\6*10\end{bmatrix}=\begin{bmatrix}-\frac{10}{3}\\60\end{bmatrix}$$

Δοκιμαστικό βήμα:

$$y_0=x_0-15\nabla f(x_0)=\begin{bmatrix}-5\\10\end{bmatrix}-15\begin{bmatrix}-\frac{10}{3}\\60\end{bmatrix}=\begin{bmatrix}45\\-890\end{bmatrix}$$

Προβολή στο X:

$$y_0=(45,-890)\Rightarrow z_0=P_X(y_0)=(5,-8)$$

διότι 45>5 και -890<8.

Διεύθυνση:

$$d_0=z_0-x_0=(5-(-5),-8-10)=(10,-18)$$

Τελικό βήμα :

$$x_1=x_0+0.1d_0=(-5{,}10)+0.1(10,-18)=(-4{,}8.2)$$

Βήμα 1➔2
Αρχικό σημείο:

$$x_1=(-4{,}8.2)$$

Gradient:

$$\nabla f(x_1)=\begin{bmatrix}\frac{2}{3}*(-4)\\6*8.2\end{bmatrix}=\begin{bmatrix}-\frac{8}{3}\\49.2\end{bmatrix}$$

Δοκιμαστικό βήμα:

$$y_1=x_1-15\nabla f(x_1)=\begin{bmatrix}-5\\10\end{bmatrix}-15\begin{bmatrix}-\frac{8}{3}\\49.2\end{bmatrix}=\begin{bmatrix}36\\-729.8\end{bmatrix}$$

Προβολή στο X:

$$y_1=(36,-729.8)\Rightarrow z_1=P_X(y_1)=(5,-8)$$

Διεύθυνση:

$$d_1 = z_1 - x_1 = (5-(-4), -8-8.2) = (9, -16.2)$$

Τελικό βήμα :

$$x_2 = x_1 + 0.1 d_1 = (-4, 8.2) + 0.1(9, -16.2) = (-3.1, 6.58)$$

Βήμα 2→3

Αρχικό σημείο:

$$x_2 = (-3.1, 6.58)$$

Gradient:

$$\nabla f(x_2) = \begin{bmatrix} \frac{2}{3} * (-3.1) \\ 6 * 6.58 \end{bmatrix} = \begin{bmatrix} -2.066 \\ 39.48 \end{bmatrix}$$

Δοκιμαστικό βήμα:

$$y_2 = x_2 - 15 \nabla f(x_2) \approx (27.9, -585.62)$$

Προβολή στο X:

$$y_2 = (27.9, -585.62) \Rightarrow z_2 = P_X(y_2) = (5, -8)$$

Διεύθυνση:

$$d_2 = z_2 - x_2 = (5-(-3.1), -8-6.58) = (8.1, -14.58)$$

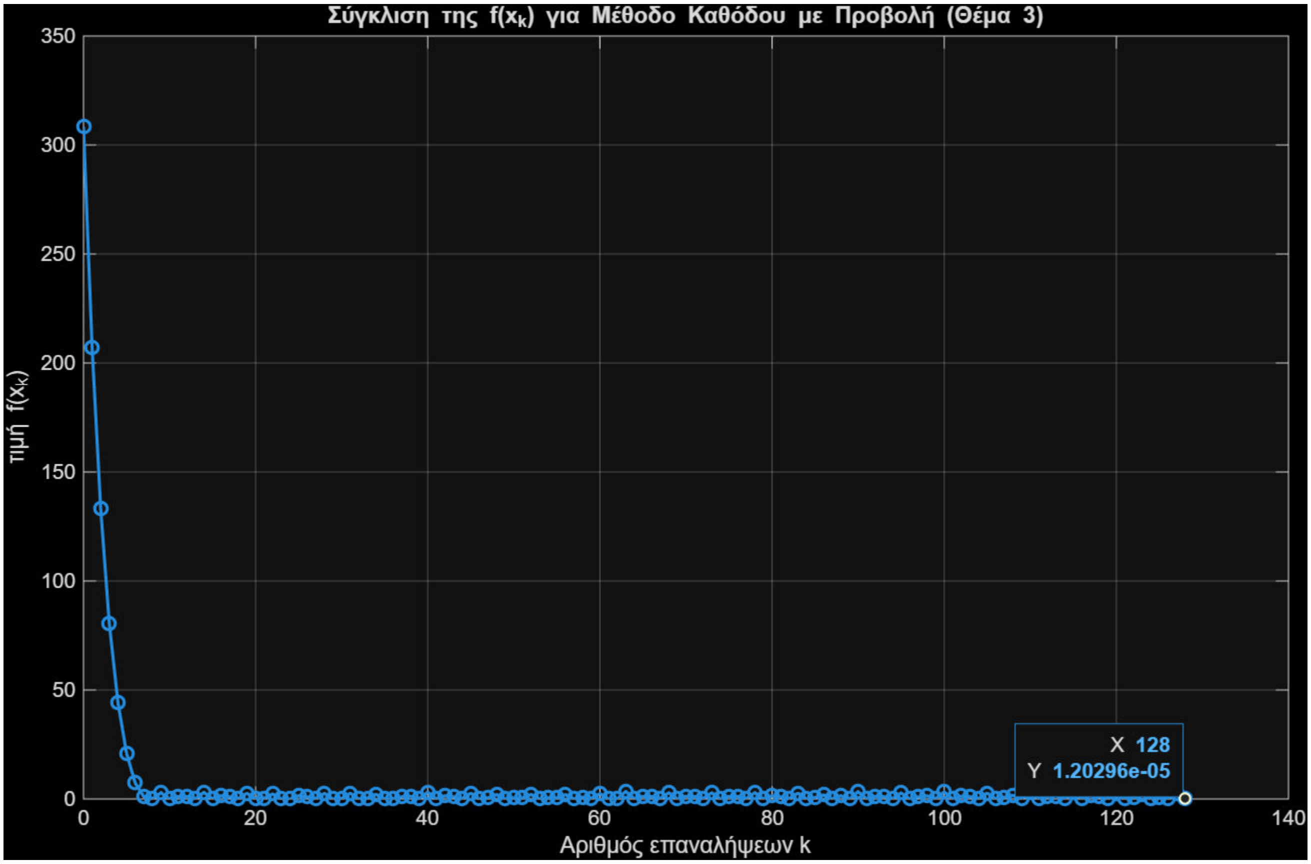
Τελικό βήμα :

$$x_3 = x_2 + 0.1 d_2 \approx (-2.29, 5.122)$$

Από τις επαναλήψεις βλέπουμε ότι το x_k κινείται σιγά σιγά προς το εσωτερικό της περιοχής X. Όταν φτάσει εκεί η προβολή αλλάζει χαρακτήρα και τελικά το x_1 θα τείνει στο 0.

Μέσα από την προσομοίωση βλέπουμε ότι: $\|x_k\| < \varepsilon = 0.01$ όταν k= 128.

Ακολουθεί η προσομοίωση με Μέθοδο Μέγιστης Καθόδου με Προβολή με $s_k = 15, \gamma_k = 0.1, x_0 = (-5, 10), \varepsilon = 0.01$ της f συναρτήσει των επαναλήψεων k=129.



Σχολιασμός:

Στα πρώτα λίγα βήματα (k = 0-5) η f πέφτει πολύ απότομα από την τιμή 300 κοντά στο 0.Αυτό οφείλεται στο μεγάλο βήμα s = 15 που σπρώχνει το σημείο κατά μήκος του $-\nabla f(x_k)$ και στην προβολή που το φέρνει κοντά στο εσωτερικό της περιοχής X.

Για επόμενες επαναλήψεις (k=5-30) η f είναι ήδη κοντά στην σύγκλιση και συνεχίζει να μειώνεται με μικρές ταλαντώσεις με την βοήθεια του βήματος γ=0.1 που κάνει την ήπια σύγκλιση της προβολής.

Από εκεί και πέρα (μέχρι k=128) η f ταλαντώνεται γύρω από το μηδέν σε πολύ μικρές τιμές και στο τέλος φτάνει σε $f(x_{128}) = 1.2 * 10^{-5}$ όπου η $\|x_{128}\|$ φτάνει την επιθυμητή ακρίβεια ε=0.01. Δηλαδή το σημείο είναι πρακτικά πάνω στο (0,0).

Σύγκριση Μεθόδου Μέγιστης Καθόδου χωρίς Περιορισμούς
Μέθοδος Μέγιστης Καθόδου με Προβολή και περιορισμούς $(s_k = 5, \gamma_k = 0.5, x_0 = (5, -5), \varepsilon = 0.01)$
Μέθοδος Μέγιστης Καθόδου με Προβολή και περιορισμούς $(s_k = 15, \gamma_k = 0.1, x_0 = (-5{,}10), \varepsilon = 0.01)$

Στην Μέθοδο Μέγιστης Καθόδου χωρίς Προβολή (Θέμα 1) για γ=0.1 και 0.3 η f πέφτει μονοτονικά και η $\|x_k\|$ πέφτει κάτω από την ακρίβεια ε=0.001. Έχουμε καθαρή σύγκλιση στο (0,0) άλλοτε αργά και άλλοτε γρήγορα. Όμως για γ=3 και 5 η μέθοδος είναι ασταθής και δεν ικανοποιείται το κριτήριο για την ακρίβεια.

Αυτό συνολικά μας δίνει να καταλάβουμε ότι με κατάλληλο βήμα πετυχαίνουμε την σύγκλιση και την ακρίβεια ακόμη και σε λίγες επαναλήψεις πάντα στην περίπτωση όπου δεν έχουμε περιορισμούς.

Στην Μέθοδο Μέγιστης Καθόδου με Προβολή και περιορισμούς (Θέμα 2) οι παράμετροι με τους οποίους δουλεύουμε καθώς και το αρχικό σημείο $(s_k=5, \gamma_k=0.5, x_0=(5,-5))$ εγκλωβίζουν την μέθοδο και η f πηγαινοέρχεται μεταξύ των τιμών 5.33 και 85.33. Εδώ η προβολή κάνει την δουλειά της αλλά αυτός ο συνδυασμός των βημάτων είναι αρκετά επιθετικός και δεν συγκλίνει το σύστημα στο ελάχιστο.
Συμπαιρένουμε λοιπόν, ότι ο συγκεκριμένος συνδυασμός s,γ δεν είναι καλή επιλογή ελέγχου δίνει μεν σταθερά φραγμένα αποτελέσματα αλλά δεν συγκλίνει ποτέ στο (0,0).

Στην Μέθοδο Μέγιστης Καθόδου με Προβολή και περιορισμούς (Θέμα 3) οι παράμετροι με τους οποίους δουλεύουμε καθώς και το αρχικό σημείο $(s_k=15, \gamma_k=0.1, x_0=(-5,10))$ ευνοούν την σύγκλιση. Η f έχει πολύ γρήγορη αρχική πτώση προς το 0 και στο k=128 πετυχαίνουμε και την ζητούμενη ακρίβεια. Από τις μεθόδους με περιορισμούς αυτός ο συνδυασμός (s μεγάλο, γ μικρό) είναι ο πιο ισορροπημένος. Σέβεται τους περιορισμούς, συγκλίνει και δίνει την ζητούμενη ακρίβεια.

Συνολικά για το συγκεκριμένο πρόβλημα ελαχιστοποίησης η Μέθοδος Μέγιστης Καθόδου χωρίς περιορισμούς συγκλίνει γρήγορα και ομαλά και είναι η πιο αξιόπιστη επιλογή αν δεν υπάρχουν περιορισμοί. Αν τώρα υπάρχουν περιορισμοί ο καλύτερος συνδυασμός βημάτων είναι εκείνος του θέματος 3 που εξασφαλίζει τη ζητούμενη σύγκλιση, ακρίβεια και σέβεται τους περιορισμούς.

<u>ΘΕΜΑ 4</u>

Στο τέταρτο θέμα εξετάζουμε ξανά την Μέθοδο Μεγίστης Καθόδου με Προβολή για τους ίδιους περιορισμούς με πιο μικρό προ-βήμα $s_k=0.1$ και τελικό βήμα $\gamma_k=0.2$ ξεκινώντας από το σημείο $x_0=(8,-10)$ και ακρίβεια ε=0.01. Στόχος είναι να μελετήσουμε αν με τόσο συντηρητικά βήματα μπορούμε να καταφέρουμε σύγκλιση μέσα στο X, αν μπορούμε να το κάνουμε πιο γρήγορα από πριν (spoiler alert = δεν μπορούμε) και αν μπορούμε να πετύχουμε την ακρίβεια ε.

Πληροφορία που μπορούμε να δούμε εκ των προτέρων για την σύγκλιση ότι αν δεν υπήρχε προβολή και παίρναμε κλασσικό βήμα:

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

Η γνωστή συνθήκη για σύγκλιση τετραγωνικής f είναι:

$$0<s<\frac{2}{\lambda_{max}(Q)}=\frac{1}{3}$$

Εδώ $s=0.1<\frac{1}{3}$ άρα ο απλός gradient descent θα ήταν σίγουρα συγκλίνων.
Η προβολή P_X είναι μη επεκτατικός τελεστής σε κυρτό και κλειστό σύνολο X. Για κυρτή f, η σύνθεση βήμα καθόδου + προβολή με τόσο μικρό βήμα διατηρεί την σύγκλιση προς το ελάχιστο στο X που εδώ είναι το (0,0).

Επίσης με αρχικό σημείο το $x_0=(8,-10)$ και η επιτρεπτή περιοχή
$X=\{x\in\Re^2: -1\leq x_1\leq 5, -8\leq x_2\leq 12\}$ αν ελέγξουμε τις πρώτες συνιστώσες
x_1: $-10\leq 8\leq 5$ 8>5 δεν ισχύει και
x_2 : $-2\leq -10\leq 12$ που δεν ισχύει επίσης γιατί -10<8 άρα $x_0\notin X$.
Άρα κάνω προβολή.

Όμως αμέσως μετά την πρώτη προβολή:

$$y_0 \rightarrow z_0 = P_X(y_0) = (5, -4)$$ που είναι μέσα στο X.

Και από εκεί και πέρα επειδή τα βήματα s,γ είναι μικρά και είμαι ήδη μέσα στο X.
Οπότε η σύγκλιση φαίνεται εκ των προτέρων, πριν τελειώσω τα βήματα είμαι σίγουρος για την ισορροπία του αλγορίθμου σε κάποια επανάληψη k στο (0,0).

Πρώτα βήματα αλγορίθμου από $x_0 = (8, -10)$

Βήμα $0 \rightarrow 1$
Αρχικό σημείο:

$$x_0 = (8, -10)$$

Gradient:

$$\nabla f(x_0) = \begin{bmatrix} \frac{2}{3} * 8 \\ 6 * (-10) \end{bmatrix} = \begin{bmatrix} \frac{16}{3} \\ -60 \end{bmatrix} \approx \begin{bmatrix} 5.333 \\ -60 \end{bmatrix}$$

Δοκιμαστικό βήμα:

$$y_0 = x_0 - 0.1 \nabla f(x_0) = \begin{bmatrix} 8 \\ -10 \end{bmatrix} - 0.1 \begin{bmatrix} 5.333 \\ -60 \end{bmatrix} \approx \begin{bmatrix} 7.466 \\ -4 \end{bmatrix}$$

Προβολή στο X:

$$y_0 = (7.466, -4) \Rightarrow z_0 = P_X(y_0) = (5, -4)$$

Διεύθυνση:

$$d_0 = z_0 - x_0 = (5 - 8, -4 - (-10)) = (-3, 6)$$

Τελικό βήμα :

$$x_1 = x_0 + 0.2 d_0 \approx (7.4, -8.8)$$

Βήμα $1 \rightarrow 2$
Αρχικό σημείο:

$$x_1 = (7.4, -8.8)$$

Gradient:

$$\nabla f(x_1) = \begin{bmatrix} \frac{2}{3} * 7.4 \\ 6 * (-8.8) \end{bmatrix} \approx \begin{bmatrix} 4.933 \\ -52.8 \end{bmatrix}$$

Δοκιμαστικό βήμα:

$$y_1 = x_1 - 0.1 \nabla f(x_1) \approx \begin{bmatrix} 6.907 \\ -3.52 \end{bmatrix}$$

Προβολή στο X:

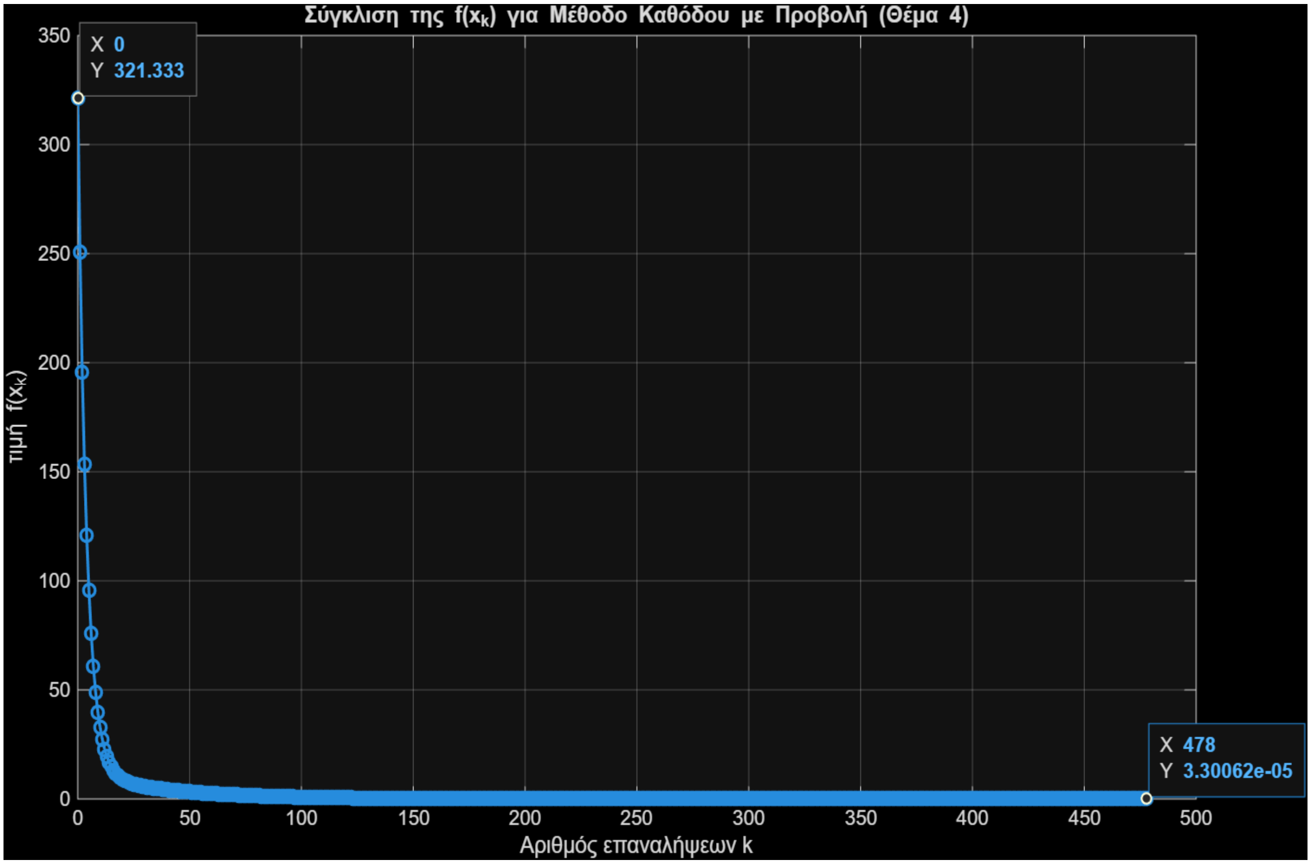
$$z_1 = P_X(y_1) = (5, -3.52)$$

Διεύθυνση:

$$d_1 = z_1 - x_1 = (5 - 7.4, -3.52 - (-8.8)) = (-2.4, 5.28)$$

Τελικό βήμα :

$$x_2 = x_1 + 0.2 d_1 \approx (6.92, -7.744)$$



Σχολιασμός:
Στο k=0 η f έχει την τιμή 321.33.
Στις πρώτες 20 επαναλήψεις η καμπύλη πέφτει πολύ απότομα κοντά στην τιμή 10 και μετά κάτω από 5. Αυτό είναι ένα γρήγορο αρχικό στάδιο της μεθόδου με πολύ γρήγορη σύγκλιση προς το σημείο ισορροπίας.
Στη συνέχεια η f κάνει αργά μικρά βήματα μέχρι την επίτευξη της ακρίβειας που έχει οριστεί. Αυτό συμβαίνει εξαιτίας της μικρής τιμής των βημάτων s και γ τα οποία διορθώνουν συνεχώς την f κοντά στο άξονα του μηδέν.
Κοντά στις 478 επαναλήψεις η f φτάνει σε πολύ μικρή τιμή περίπου $3*10^{-5}$ και η $\|x_k\|$ πέφτει κάτω από το ε=0.01 και όπου ικανοποιεί το κριτήριο ακρίβειας και ο αλγόριθμος σταματά.

Σύγκριση Μεθόδου Μέγιστης Καθόδου χωρίς Περιορισμούς
Μεθόδου Μέγιστης Καθόδου με Προβολή και περιορισμούς $(s_k=5, \gamma_k=0.5, x_0=(5,-5), \varepsilon=0.01)$
Μεθόδου Μέγιστης Καθόδου με Προβολή και περιορισμούς $(s_k=15, \gamma_k=0.1, x_0=(-5,10), \varepsilon=0.01)$
Μεθόδου Μέγιστης Καθόδου με Προβολή και περιορισμούς
$(s_k=0.1, \gamma_k=0.2, x_0=(8,-10), \varepsilon=0.01)$

Όπως αναφέραμε και στην παραπάνω αναλυτική σύγκριση, για το συγκεκριμένο πρόβλημα ελαχιστοποίησης η Μέθοδος Μέγιστης Καθόδου χωρίς περιορισμούς συγκλίνει γρήγορα και ομαλά και είναι η πιο αξιόπιστη επιλογή αν δεν υπάρχουν περιορισμοί. Αν τώρα υπάρχουν περιορισμοί ο καλύτερος συνδυασμός βημάτων είναι εκείνος του θέματος 3 που εξασφαλίζει τη ζητούμενη σύγκλιση, ακρίβεια και σέβεται τους περιορισμούς.

Τώρα για την Μέθοδο Μέγιστης Καθόδου με Προβολή και περιορισμούς με βήματα $s_k=0.1, \gamma_k=0.2$ και σημείο εκκίνησης το (8,-10), ξεκινάμε με σημείο εκτός περιοχής X. Η προβολή φέρνει σταδιακά τα σημεία μέσα στην επιτρεπόμενη περιοχή και εξαιτίας των πολύ μικρών βημάτων η f μειώνεται ομαλά χωρίς αστάθεια και συγκλίνει σιγά σιγά στο ζητούμενο ελάχιστο με την ζητούμενη ακρίβεια. Ωστόσο η σύγκλιση είναι αρκετά αργή (k=478). Άρα ο συνδυασμός αυτός των σ,γ είναι πολύ ασφαλής αλλά συντηρητικός εξασφαλίζει δηλαδή σίγουρη σύγκλιση αλλά με μεγάλο κόστος σε αριθμό επαναλήψεων.

Συνολικά για το συγκεκριμένο πρόβλημα μπορούμε να πούμε ότι χωρίς περιορισμούς η κλασσική μέθοδος του θέματος 1 με σωστό γ είναι η πιο αποδοτική. Όταν όμως υπάρχουν περιορισμοί, η μέθοδος με προβολή είναι μονόδρομος και ο καλύτερος συνδυασμός βημάτων είναι αυτός του θέματος 3 που πετυχαίνει τόσο ταχύτητα σύγκλισης όσο και σταθερότητα σε αντίθεση με τον πολύ επιθετικό συνδυασμό του θέματος 2 και τον υπερβολικά αργό αλλά ασφαλή συνδυασμό του θέματος 4.