Εργασία: Τεχνικές Βελτιστοποίησης (Μέρος 2)

Panagiotis Koutris

November 2024

Εισαγωγή

Στην εργασία αυτή θα ασχοληθούμε με το πρόβλημα ελαχιστοποίησης της συνάρτησης:

$$f(x, y) = x^5 \cdot e^{-x^2-y^2}$$

μέσω της χρήσης των εξής αλγορίθμων αναζήτησης:

- **Μέθοδος Μέγιστης Καθόδου (Steepest Descent)**
- **Μέθοδος Newton**
- **Μέθοδος Levenberg-Marquardt**

Για κάθε μία από τις μεθόδους θα εξεταστούν οι εξής περιπτώσεις:
Τα αρχικά σημεία $(\mathbf{x_0}, \mathbf{y_0})$ θα είναι τα εξής:

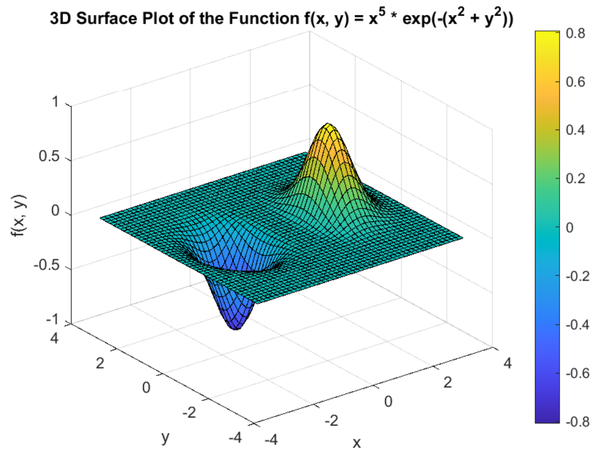
- $(\mathbf{0}, \mathbf{0})$,
- $(\mathbf{-1}, \mathbf{1})$,
- $(\mathbf{1}, \mathbf{-1})$.

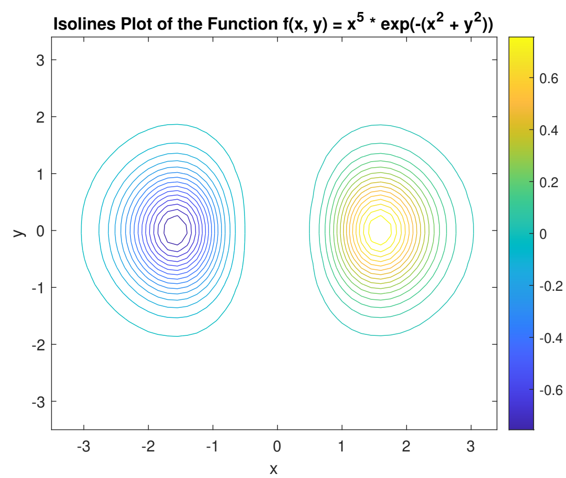
Το βήμα γ_k θα επιλεγεί ως εξής:

1. **Σταθερό** (της επιλογής μας),
2. **Με ελαχιστοποίηση της** $f(x_k + \gamma_k d_k)$ **μέσω της μεθόδου της διχοτόμου**,
3. **Βάσει του κανόνα Armijo**.

ΘΕΜΑ 1: Σχηματική απεικόνιση της f

Η συνάρτηση $f(x, y) = x^5 \cdot e^{-(x^2+y^2)}$ έχει την εξής μορφή:





Από τα παραπάνω διαγράμματα μπορούμε προσεγγιστικά να δούμε ότι το ελάχιστο βρίσκεται στη θέση (-1.5 , 0) και έχει τιμή περίπου ίση με -0,8.

ΘΕΜΑ 2: Μέθοδος Μέγιστης Καθόδου

Η μέθοδος μέγιστης καθόδου αποτελεί μια μέθοδο βελτιστοποίησης, η οποία βασίζεται στον βαθμό κλίσης της συνάρτησης $f(x)$ για να βρει ένα τοπικό ελάχιστο. Σε κάθε βήμα, επιλέγεται ως κατεύθυνση αναζήτησης η κατεύθυνση της μέγιστης καθόδου:

$$d_k = -\nabla f(x_k),$$

δηλαδή η αντίθετη της κλίσης, καθώς αυτή εξασφαλίζει τη μέγιστη μείωση της $f(x)$ τοπικά. Η διαδικασία επαναλαμβάνεται μέχρι να ικανοποιηθεί μια συνθήκη τερματισμού, όπως:

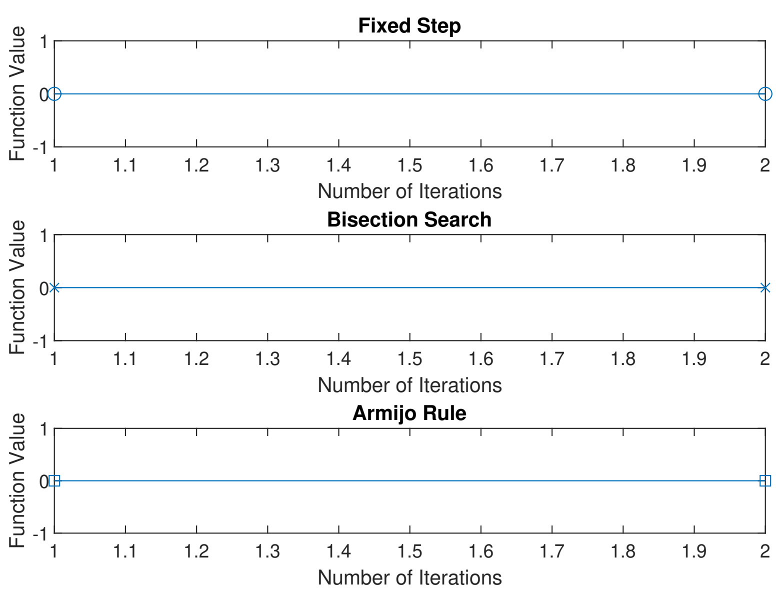
$$\|\nabla f(x_k)\| < \epsilon,$$

όπου $\epsilon > 0$ είναι μια προκαθορισμένη ανοχή. Παρακάτω παρατίθενται τα αποτελέσματα για τα διάφορα αρχικά σημεία και τους τρόπους υπολογισμού του βήματος:

Αρχικό σημείο (0,0)

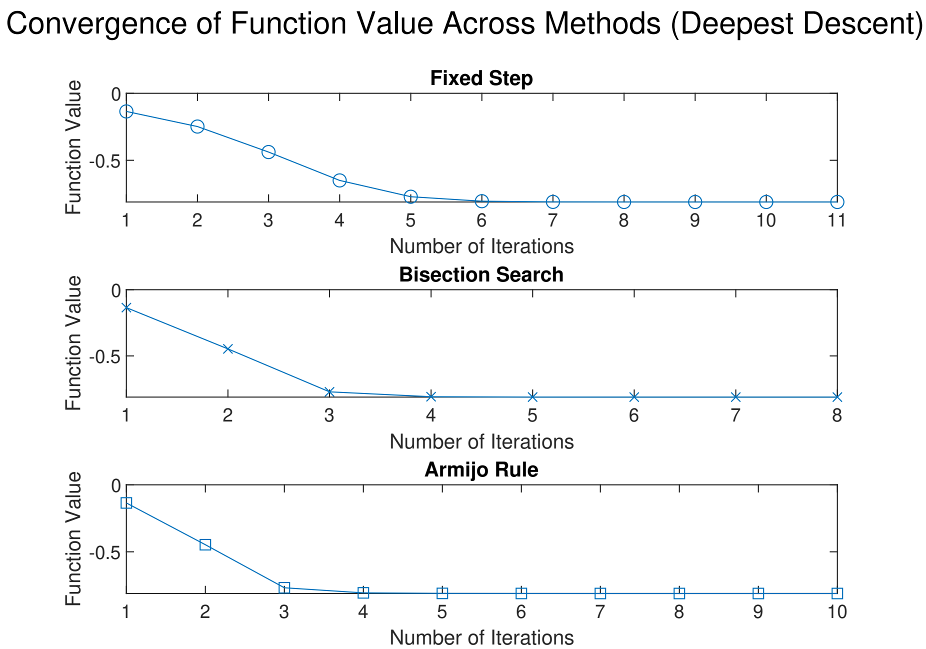
	Method	Final_x	Final_y	Final_f_value
1	Fixed Step	0	0	0
2	Bisection ...	0	0	0
3	Armijo Rule	0	0	0

Convergence of Function Value Across Methods (Deepest Descent)



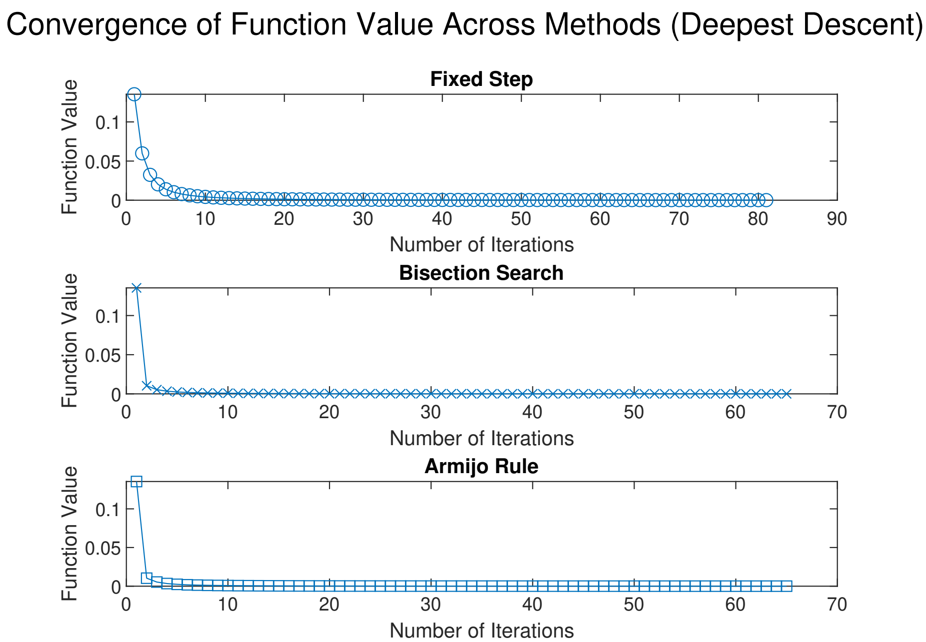
Αρχικό σημείο (-1,1)

	Method	Final_x	Final_y	Final_f_value
1	Fixed Step	-1.5811	4.6098e-04	-0.8112
2	Bisection ...	-1.5811	3.1380e-04	-0.8112
3	Armijo Rule	-1.5812	2.1280e-04	-0.8112



Αρχικό σημείο (1,-1)

	Method	Final_x	Final_y	Final_f_value
1	Fixed Step	0.2313	-1.3021	1.1517e-04
2	Bisection ...	0.1902	-1.3649	3.7221e-05
3	Armijo Rule	0.1902	-1.3649	3.7221e-05



Η επιλογή του αρχικού σημείου έναρξης είναι καθοριστική για τη σύγκλιση της μεθόδου μέγιστης καθόδου και επηρεάζει την επίτευξη του ελαχίστου. Για τα τρία διαφορετικά αρχικά σημεία καταλήξαμε στα εξής συμπεράσματα:

- **Σημείο** $(0,0)$: Το αρχικό σημείο $(0,0)$ δεν παρουσίασε καμία μεταβολή ανεξάρτητα από τη μέθοδο επιλογής βήματος. Το τελικό σημείο παρέμεινε το ίδιο με το αρχικό, δείχνοντας την ιδιαιτερότητα του σημείου αυτού ως ένα κρίσιμο σημείο όπου ο βαθμός κλίσης είναι μηδενικός.

- **Σημείο** $(-1,1)$: Η εκκίνηση από το $(-1,1)$ οδήγησε σε διαφορετικό αποτέλεσμα. Πήραμε ικανοποιητικά αποτελέσματα όσον αφορά την εύρεση του ελαχίστου καθώς συγκριτικά με τα άλλα σημεία δεν ξεκινήσαμε από κρίσιμο σημείο/τοπικό ελάχιστο δηλαδή ο βαθμός κλίσης δεν ήταν μηδενικός και το αρχικό σημείο βρίσκονταν ήδη σχετικά κοντά στο ελάχιστο.

- **Σημείο** $(1, -1)$: Από το σημείο $(1, -1)$, οι κλίσεις οδήγησαν τη μέθοδο σε τελικά σημεία που απείχαν σημαντικά από το ελάχιστο καθώς η μέθοδος εγκλωβίστηκε σε τοπικά ελάχιστα όπου ο βαθμός κλήσης είναι μηδενικός.

Η επιλογή του τρόπου υπολογισμού του βήματος παίζει άμεσο ρόλο στον αριθμό των επαναλήψεων και επηρεάζει σημαντικά την ταχύτητα και την αποτελεσματικότητα της σύγκλισης. Για τους τρεις διαφορετικούς τρόπους καταλήξαμε στα εξής συμπεράσματα:

- **Σταθερό Βήμα (Fixed Step):** Η επιλογή σταθερού βήματος παρουσίασε την πιο αργή σύγκλιση μεταξύ των μεθόδων. Παρόλο που η απλότητα υλοποίησης είναι πλεονέκτημα, η αποτελεσματικότητα εξαρτάται έντονα από την τιμή του βήματος. Ένα μεγάλο βήμα μπορεί να προκαλέσει αστάθεια, ενώ ένα πολύ μικρό βήμα οδηγεί σε υπερβολικό αριθμό επαναλήψεων.
- **Βήμα με Ελαχιστοποίηση (Minimum Step):** Ο υπολογισμός του βήματος που ελαχιστοποιεί τη συνάρτηση κατά μήκος της κατεύθυνσης αναζήτησης είχε τον μικρότερο αριθμό επαναλήψεων. Παρόλα αυτά, το αυξημένο υπολογιστικό κόστος λόγω της ελαχιστοποίησης το καθιστά λιγότερο αποδοτικό σε επίπεδο κλοστους παρόλου που είναι ο πιο αποδοτικός σε επίπεδο αριθμού επαναλήψεων.
- **Βάσει του Κανόνα Armijo (Armijo Rule):** Η μέθοδος του Armijo προσφέρει μια καλή ισορροπία μεταξύ ταχύτητας σύγκλισης και υπολογιστικού κόστους.

ΘΕΜΑ 3: Μέθοδος Newton

Η μέθοδος Newton αποτελεί μια μέθοδο βελτιστοποίησης δεύτερης τάξης, η οποία χρησιμοποιεί τον βαθμό κλίσης $\nabla f(x)$ και τον πίνακα Hessian $\nabla^2 f(x)$ της συνάρτησης $f(x)$ για την εύρεση ενός τοπικού ελαχίστου. Απαραίτητη προϋπόθεση για την εφαρμογή της μεθόδου είναι ο πίνακας Hessian $\nabla^2 f(x)$ να είναι θετικά ορισμένος. Σε κάθε βήμα, υπολογίζεται η κατεύθυνση αναζήτησης:

$$d_k = -\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k),$$

η οποία λαμβάνει υπόψη την καμπυλότητα της συνάρτησης, εξασφαλίζοντας ταχύτερη σύγκλιση συγκριτικά με μεθόδους πρώτης τάξης. Το νέο σημείο δίνεται από τη σχέση:

$$x_{k+1} = x_k + \gamma_k d_k,$$

όπου γ_k είναι το βήμα που μπορεί να υπολογιστεί με σταθερό βήμα ή γραμμική αναζήτηση. Η διαδικασία συνεχίζεται έως ότου ικανοποιηθεί μια συνθήκη τερματισμού, όπως:

$$\|\nabla f(x_k)\| < \epsilon,$$

όπου $\epsilon > 0$ είναι μια προκαθορισμένη ανοχή.

Στην περίπτωση της δικιάς μας f ο εσσιανός της πίνακας δεν είναι θετικά ορισμένος στα αρχικά σημεία οπότε δεν μπορεί να εφαρμοστεί η μέθοδος Newton. Ωστόσο παρατίθεται κώδικας της μεθόδου για λόγους πληρότητας. Για τέτοιες περιπτώσεις συναρτήσεων υπάρχουν καλύτερες παραλλαγές της μεθόδου όπως η επόμενη.

ΘΕΜΑ 4: Μέθοδος Levenberg-Marquardt

Η μέθοδος Levenberg-Marquardt είναι μια υβριδική μέθοδος βελτιστοποίησης, η οποία συνδυάζει χαρακτηριστικά της μεθόδου Newton και της μεθόδου μέγιστης καθόδου και χρησιμποποιείτια και στην περίπτωση που ο εσσιανός πίνακας δεν είναι θετικά ορισμένος. Η κατεύθυνση αναζήτησης υπολογίζεται ως εξής:

$$d_k = -\big(\nabla^2 f(x_k) + \mu_k I\big)^{-1} \nabla f(x_k),$$

όπου $\mu_k > 0$ είναι μια παράμετρος που προσαρμόζεται δυναμικά ώστε ο $\big(\nabla^2 f(x_k) + \mu_k I\big)$ να είναι θετικά ορισμένος. Όταν το μ_k είναι μεγάλο, η μέθοδος συμπεριφέρεται σαν μέθοδος μέγιστης καθόδου, ενώ όταν το μ_k είναι μικρό, προσεγγίζει τη μέθοδο Newton. Το νέο σημείο δίνεται από τη σχέση:

$$x_{k+1} = x_k + \gamma_k d_k,$$

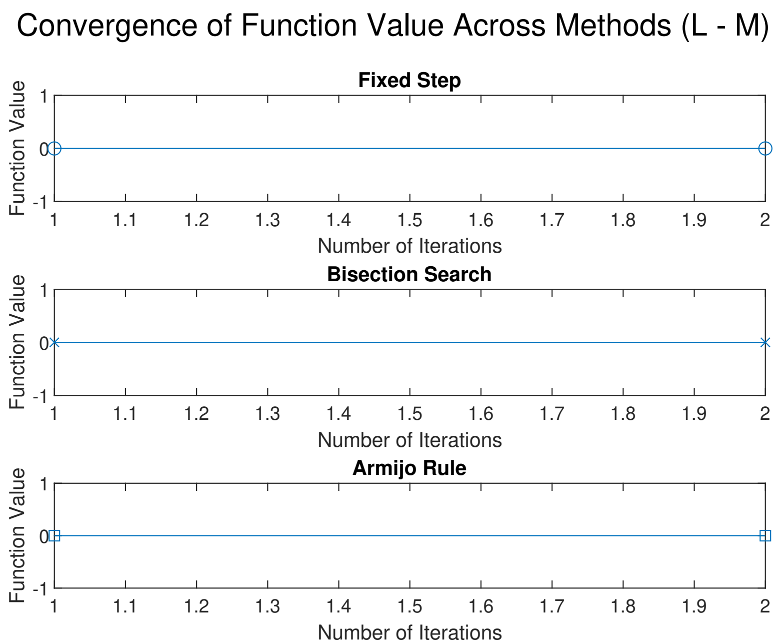
Η διαδικασία συνεχίζεται έως ότου ικανοποιηθεί μια συνθήκη τερματισμού, όπως:

$$\|\nabla f(x_k)\| < \epsilon,$$

όπου $\epsilon > 0$ είναι μια προκαθορισμένη ανοχή. Παρακάτω παρατίθενται τα αποτελέσματα για τα διάφορα αρχικά σημεία και τους τρόπους υπολογισμού του βήματος:

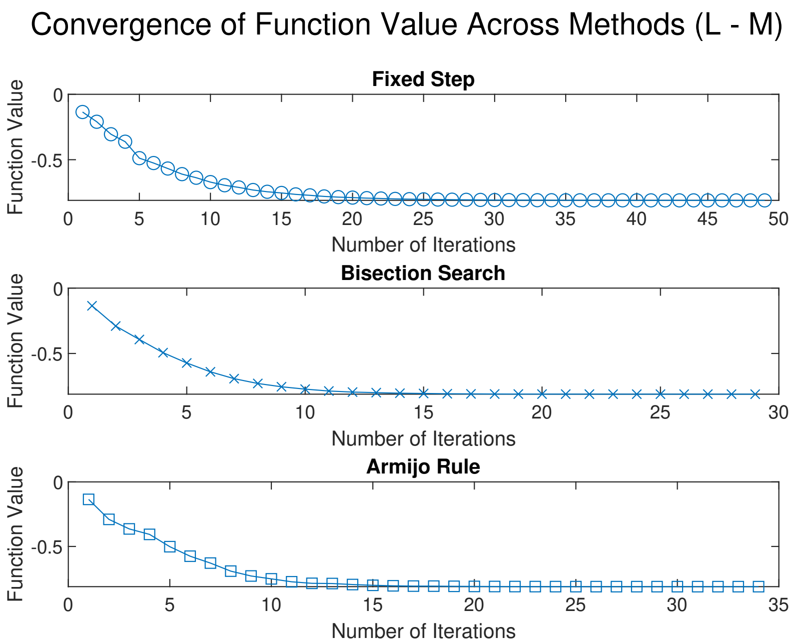
Αρχικό σημείο (0,0)

	Method	Final_x	Final_y	Final_f_value
1	Fixed Step	0	0	0
2	Bisection ...	0	0	0
3	Armijo Rule	0	0	0



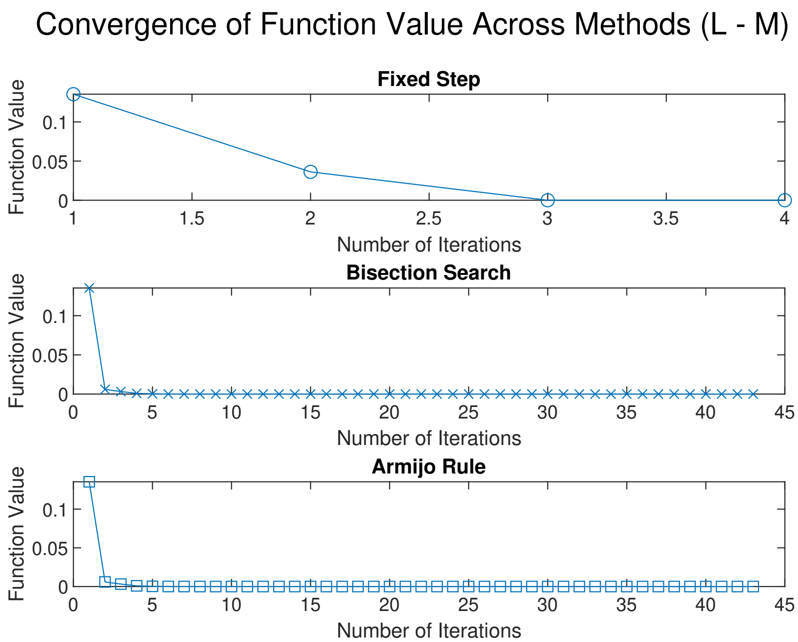
Αρχικό σημείο (-1,1)

	Method	Final_x	Final_y	Final_f_value
1	Fixed Step	-1.5811	0.0088	-0.8111
2	Bisection ...	-1.5811	0.0038	-0.8112
3	Armijo Rule	-1.5812	0.0022	-0.8112



Αρχικό σημείο (1,-1)

	Method	Final_x	Final_y	Final_f_value
1	Fixed Step	1.3206	-4.3248	5.2915e-09
2	Bisection ...	0.0781	-3.8024	1.5139e-12
3	Armijo Rule	0.0781	-3.8024	1.5138e-12



Οι παρατηρήσεις όσον αφορά τα διαφορετικά αρχικά σημεία και τους διαφορετικούς τρόπους επιλογής βήματος είναι σε γενικές γραμμές ίδιες με αυτές της μεθόδου μέγιστης καθόδου. Αξίζει βέβαια να σημειωθεί ότι για το σημείο (-1,1) που είναι το μόνο για το οποίο έχουμε επιθυμητά αποτελέσματα ,φαίνεται ότι η μέθοδος Levenberg-Marquardt χρειάζεται περισσότερες επαναλήψεις για να συγκλίνει συγκριτικά με τη μέθοδο μέγιστης καθόδου αλλά αυτό οφείλεται στη συμπεριφορά της μορφής της συνάρτησης που ευνοεί τη δεύτερη μέθοδο περισσότερο για αυτό το αρχικό σημείο. Σε γενικές γραμμές ωστόσο δεν ισχύει αυτό και η μέθοδος L-M είναι πιο αποδοτική.