Τεχνικές Βελτιστοποίησης 2η Εργαστηριακή Άσκηση

Σπύρος Κούγιας, ΑΕΜ: 10124

I. ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

Καλούμαστε να ελαχιστοποιήσουμε μία μη γραμμική, συνάρτηση δύο μεταβλητών:

$$f(x,y) = x^3 e^{-x^2-y^4}$$

Οι μέθοδοι που χρησιμοποιούμε για την ελαχιστοποίηση είναι:

- Η μέθοδος της **Μέγιστης Καθόδου**
- Η μέθοδος **Newton**
- Η μέθοδος **Levenberg-Marquardt**

Οι αλγόριθμοι που θα χρησιμοποιηθούν στηρίζονται στην ιδέα της επαναληπτικής καθόδου. Ξεκινάμε από ένα αρχικό σημείο x0 και, με διαδοχικές επαναλήψεις, μετακινούμαστε σε νέα σημεία xk για τα οποία ισχύει $f(x_{k+1}) < f(x_k)$ (θεωρητικά τουλάχιστον). Έτσι, μετακινούμαστε διαρκώς προς μικρότερες τιμές της συνάρτησης, χρησιμοποιώντας πληροφορίες της κλίσης της. Η γενική αναδρομική σχέση που περιγράφει τη διαδικασία είναι:

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

Όπου:

- **dk** είναι η κατεύθυνση έρευνας. Για τους συγκεκριμένους αλγορίθμους, η κατεύθυνση προκύπτει γενικά ως $d_k = -\Delta_k \nabla f(x_k)$, όπου Δk είναι ένας συμμετρικός και θετικά ορισμένος πίνακας. Η βασική διαφορά μεταξύ των τριών μεθόδων προέρχεται από τον ορισμό αυτού του πίνακα:
  - Στη Μέγιστη Κάθοδο, $\Delta_k = I$
  - Στη μέθοδο Newton, $\Delta_k = [\nabla^2 f(x_k)]^{-1}$
  - Στη μέθοδο Levenberg-Marquardt, $\Delta_k = [\nabla^2 f(x_k) \vdash \mu_k I]^{-1}$
- **γk** είναι το βήμα μετακίνησης το οποίο θέλει ιδιαίτερη προσοχή στον καθορισμό του. Μικρές τιμές του βήματος μπορεί να μας οδηγήσουν σε αργή σύγκλιση, ενώ μεγάλες στο «να αστοχήσουμε» το ελάχιστο (ξεκινάμε να αποκλίνουμε) ή να αποσταθεροποιήσουμε τον αλγόριθμο.

Η διαδικασία τερματίζεται όταν το μέτρο της κλίσης γίνει μικρότερο από την προκαθορισμένη ανοχή ε (προσεγγίζει στάσιμο σημείο). Δηλαδή όταν $\|\nabla f(x_k)\| \le \varepsilon$, και άρα κοντά σε ένα τοπικό ή ολικό ακρότατο.

Για κάθε αλγόριθμο, εξετάστηκαν τρεις τρόποι για τον υπολογισμό του γ_k:

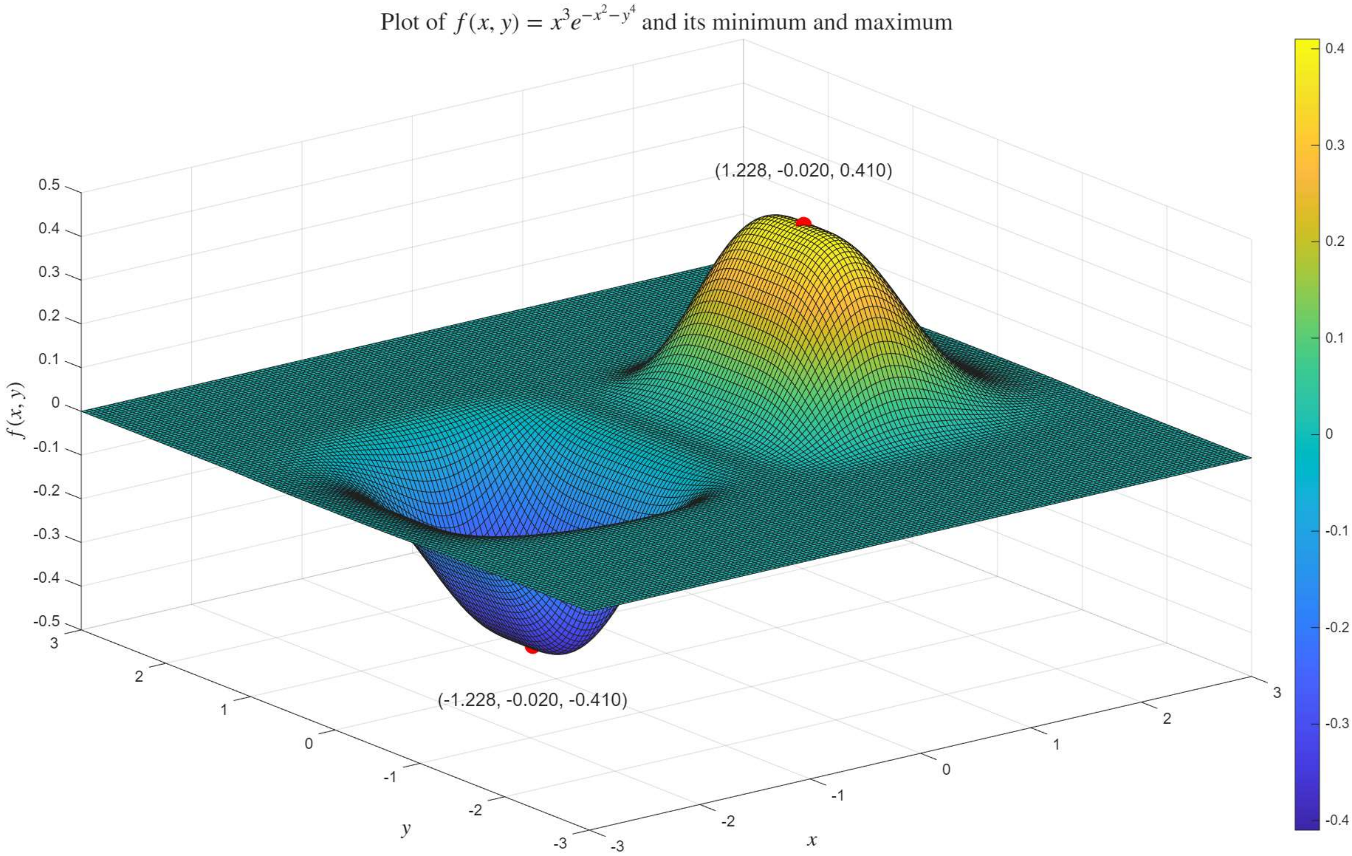
1. **Σταθερό Βήμα**, το γk είναι προκαθορισμένο.
2. **Τέτοιο ώστε να ελαχιστοποιεί την $f(x_k + \gamma_k d_k)$,** χρήση μέθοδο διχοτόμου με παραγώγους.
3. **Κανόνας Armijo**, μία μέθοδος οπισθοδρόμησης που διασφαλιστεί ότι το μέγεθος του βήματος είναι αρκετά μεγάλο ώστε να εγγυάται, επαρκή μείωση της συνάρτησης σε κάθε επανάληψη.

Μελετάμε τα αρχικά σημεία x0: [0, 0], [-1, -1], [1, 1].

II. ΘΕΜΑ 1: ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΑΚΡΟΤΑΤΑ

Σχεδιάζοντας την συνάρτηση παρατηρούμε, ολικό **μέγιστο** στο σημείο (1.228, -0.02) σε ύψος 0.410 και ολικό **ελάχιστο** στο (-1.228, -0.02) σε ύψος -0.410. Τέλος, στο (0,0) έχουμε σαγματικό σημείο (η κλίση είναι μηδέν) που σημαίνει ότι οι αλγόριθμοι μας θα τερματίζουν κατευθείαν.

Αξίζει να σημειωθεί επιπλέον πως η συνάρτηση δημιουργεί εκτεταμένες επίπεδες περιοχές όπου η κλίση είναι σχεδόν μηδενική.



III. ΘΕΜΑ 2: ΜΕΘΟΔΟΣ ΜΕΓΙΣΤΗΣ ΚΑΘΟΔΟΥ

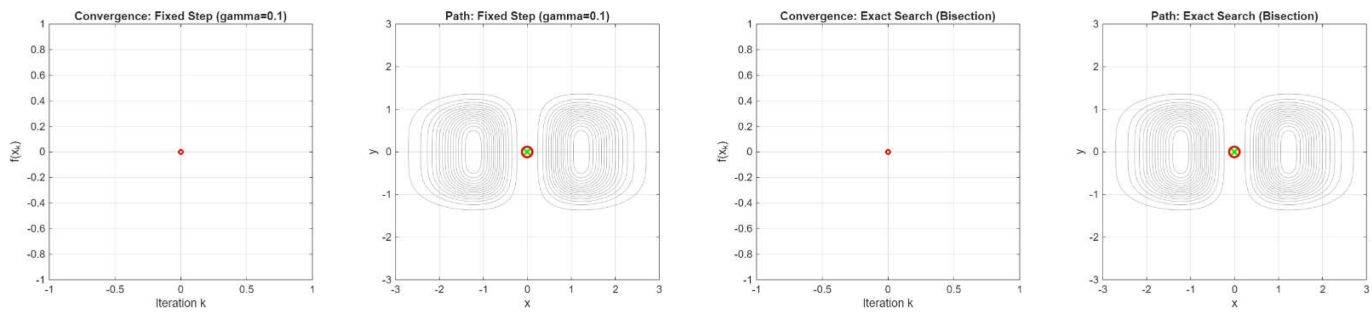
Η μέθοδος αυτή χρησιμοποιεί ως κατεύθυνση έρευνας την αρνητική κλίση της συνάρτησης ($\Delta_k = I$):

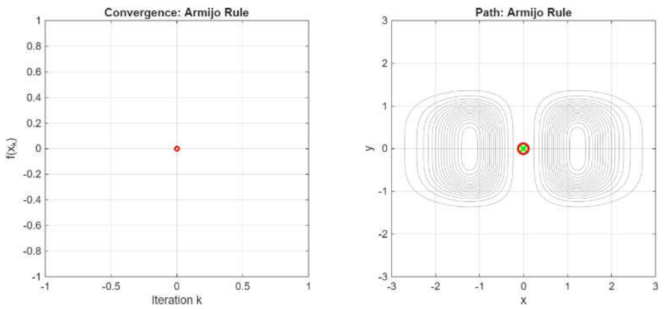
$$d_k = -\nabla f(\chi_k)$$

Ακολουθούν τα πειράματα:

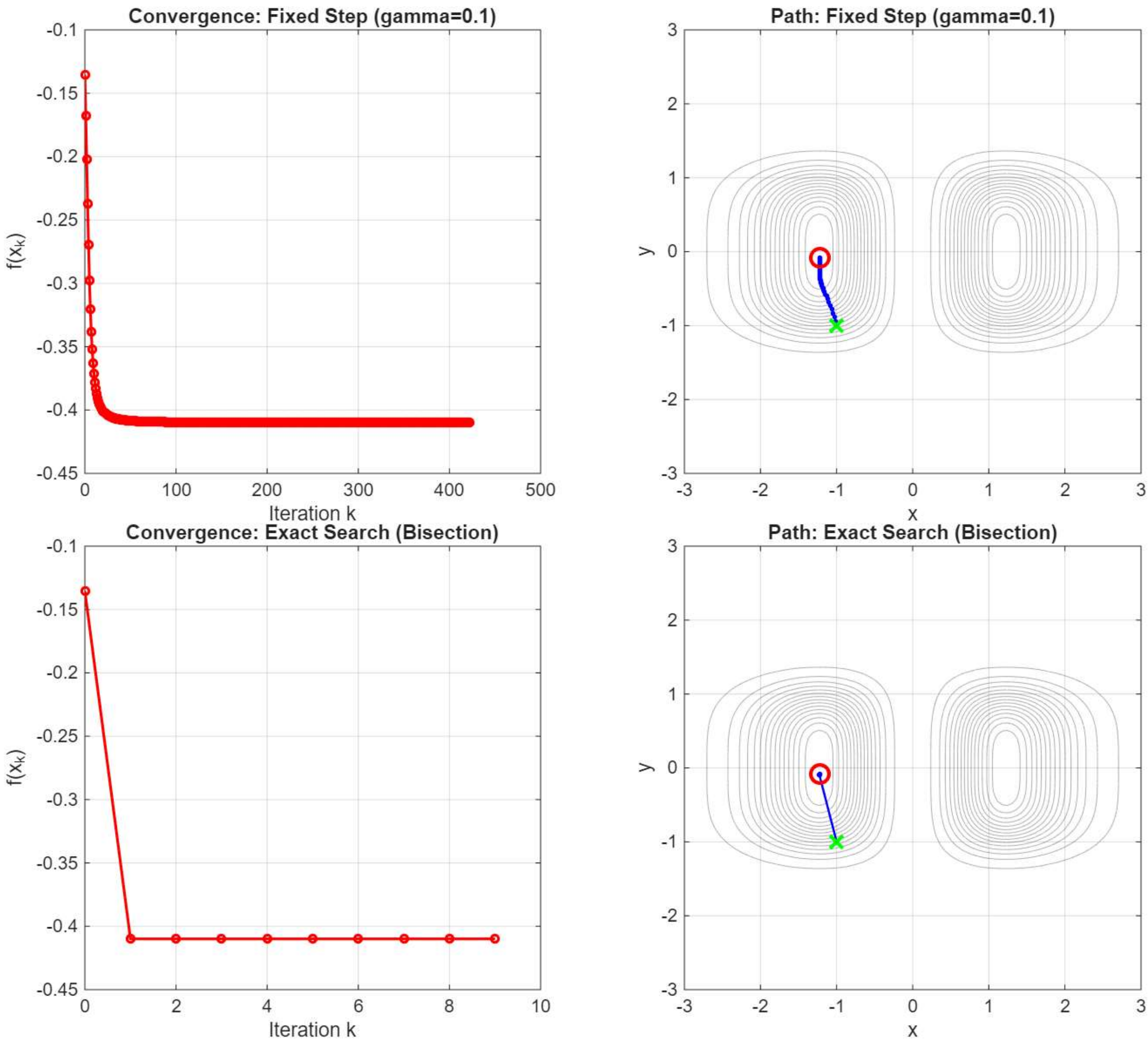
1) (0,0)

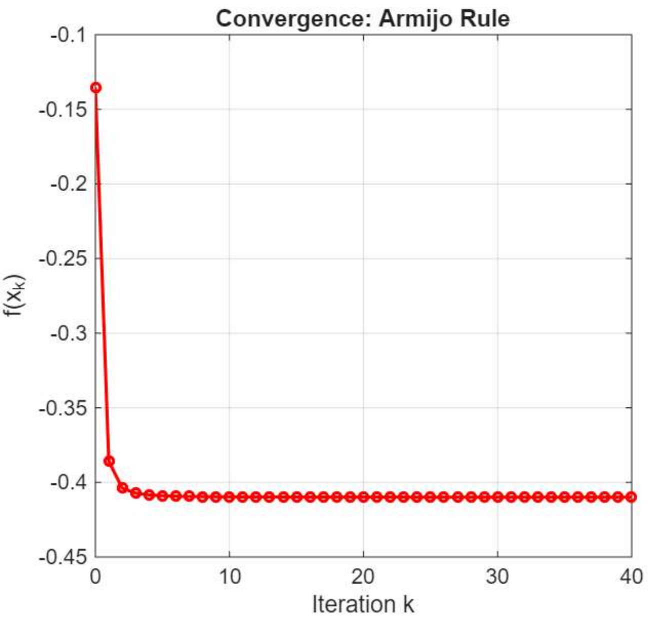
Όπως περιμέναμε, ο αλγόριθμος τερματίζει αμέσως (μηδέν επαναλήψεις), καθώς στο σημείο (0,0) υπολογίζει μηδενική κλίση. Η μέθοδος επιλογής του βήματος μετακίνησης δεν επηρεάζει το αποτέλεσμα όπως φαίνεται και στα παρακάτω γραφήματα

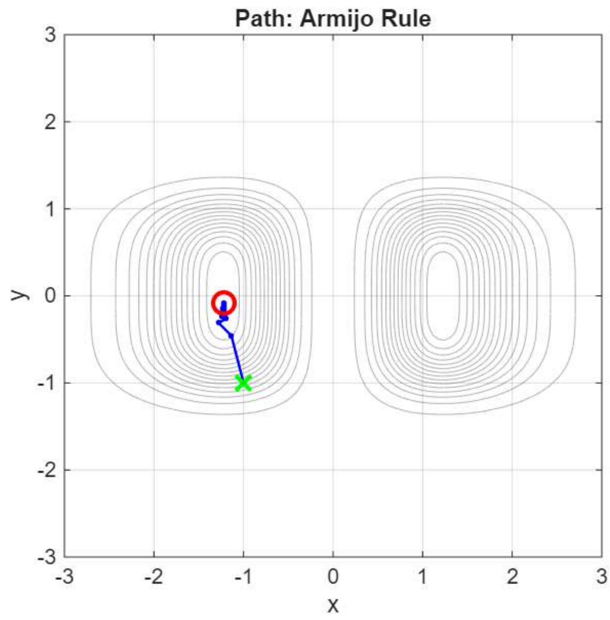



Convergence: Armijo Rule
f(x_k)
Iteration k
Path: Armijo Rule
y
x

2) (-1,-1)


Convergence: Fixed Step (gamma=0.1)
f(x_k)
Iteration k
Path: Fixed Step (gamma=0.1)
y
x
Convergence: Exact Search (Bisection)
f(x_k)
Iteration k
Path: Exact Search (Bisection)
y
x

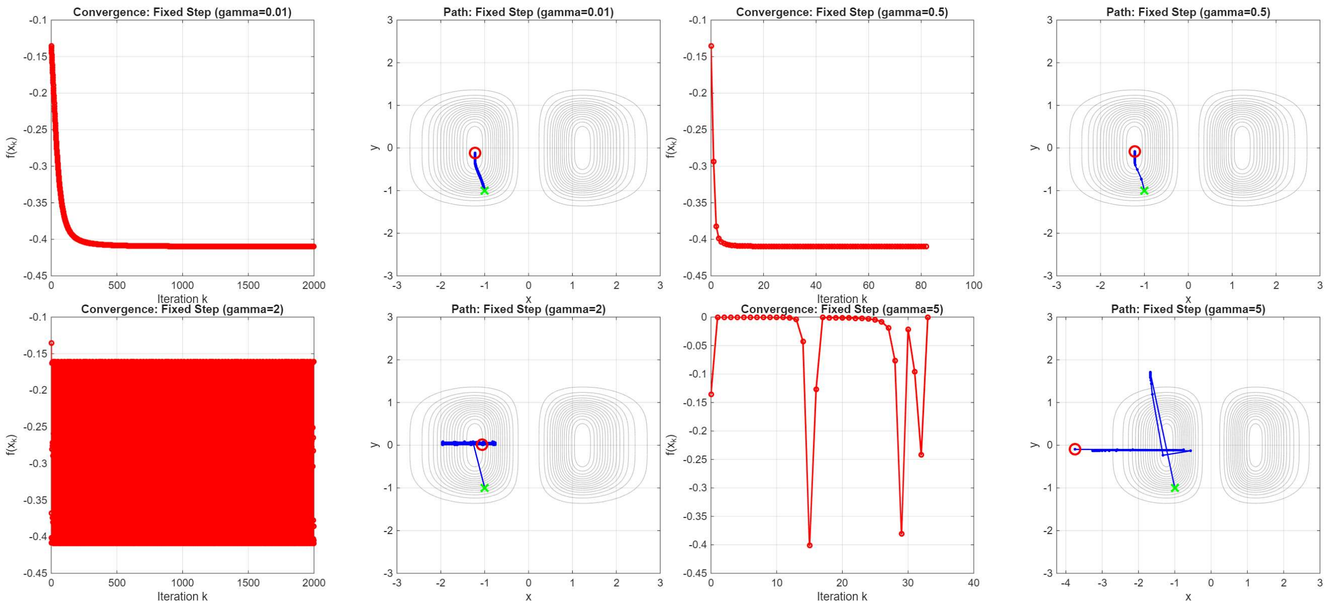




Στο σημείο (-1,-1) όλες οι τεχνικές επιλογής βήματος φαίνεται να συγκλίνουν (Η χειροκίνητη επιλογή γk στο fixed step θα εξεταστεί παρακάτω). Αυτό είναι και αναμενόμενο, καθώς η κλίση από το σημείο (-1,-1) και σε μια περιοχή γύρο από το ελάχιστο έχει αρνητική κλίση. Το πρόβλημα προκύπτει όταν ξεκινήσουμε από ένα σημείο εκτός της "κοιλάδας" αρνητικής κλίσης.

- Η **"Exact Search"** είναι η ταχύτερη καθώς χρειάστηκε τον λιγότερο αριθμό επαναλήψεων (~9) για να φτάσει στο τοπικό ελάχιστο με προδιαγεγραμμένη ακρίβεια ε (= 0.001).
- Η **Armijo** συγκλίνει επίσης γρήγορα (~40), προσαρμόζοντας δυναμικά το βήμα σε κάθε επανάληψη.
- Η **Fixed Step** (γ = 0.1) παρουσιάζει εξαιρετικά αργή σύγκλιση (>400). Καθώς ο αλγόριθμος πλησιάζει το ελάχιστο, η κλίση μειώνεται, και το σταθερό βήμα πολλαπλασιαζόμενο με τη μικρή κλίση οδηγεί σε αμελητέες μετακινήσεις.

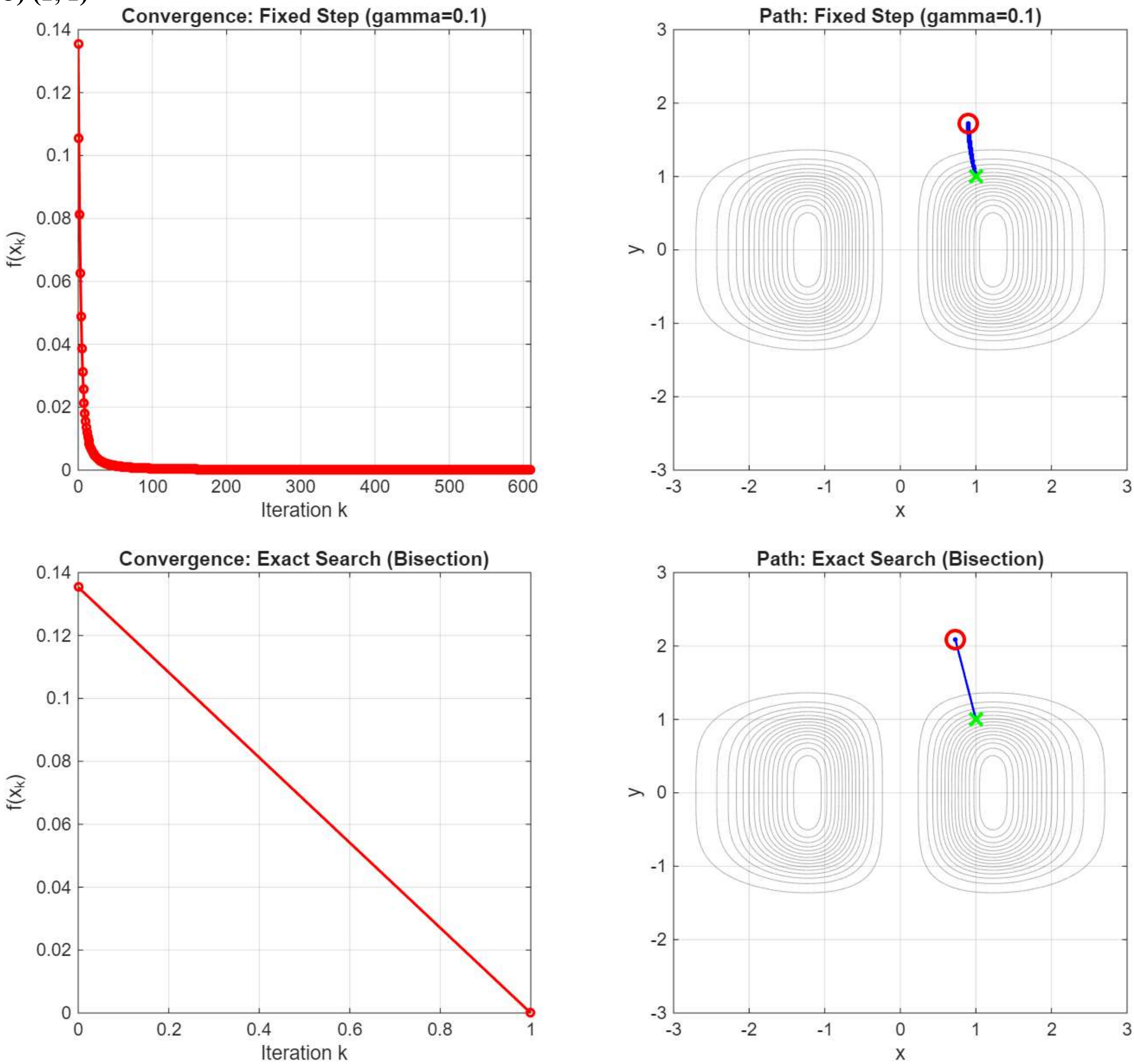
Θα εξετάσουμε και άλλες τέσσερις τιμές για το σταθερό γ (0.01, 0.5, 2, 5):

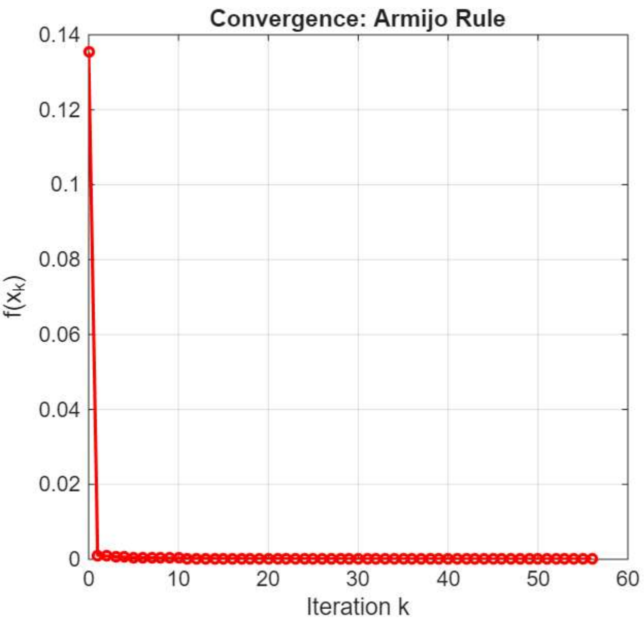


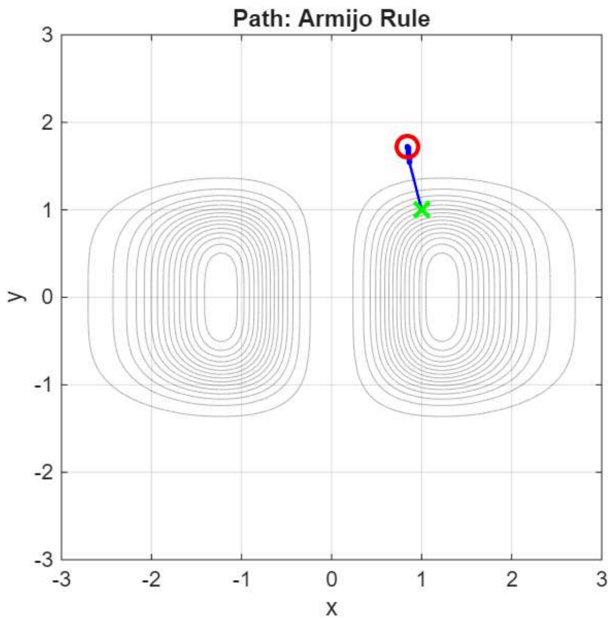
- **Το 0.01** μας δείχνει το πρόβλημα πολύ μικρού βήματος, η σύγκλιση είναι πολύ αργή. Η μέθοδος δεν τερμάτισε πριν το όριο των επαναλήψεων (2000) αλλά βλέπουμε ότι θα φτάναμε στο ολικό ελάχιστο αν συνέχιζε.
- **Το 0.5** συγκλίνει στις 80 περίπου επαναλήψεις. Είναι μία λογική τιμή για το step size και μάλιστα γρηγορότερη του 0.1.
- **Το 2** μας δείχνει το πρόβλημα πολύ μεγάλου βήματος, ο αλγόριθμος ταλαντώνει χωρίς να τερματίσει διότι σε κάθε update προσπερνάει το ολικό ελάχιστο.

- **Τέλος το 5**, μας δείχνει τι γίνεται όταν η επιλογή βήματος είναι κοντά στο μέγεθος της “κοιλάδας”. Προσπερνάμε το ολικό ελάχιστο σε βαθμό που μας κάνει να ξεφύγουμε σε περιοχή που το gradient είναι κοντά στο μηδέν.

3) (1, 1)







Όπως φαίνεται στα διαγράμματα, τα αποτελέσματα δεν είναι ακριβός ιδεατά. Ο αλγόριθμός μας τερματίζει σε σημείο που η κλίση είναι κοντά στο μηδέν και όχι στο ολικό ελάχιστο. Οι τεχνικές επιλογής βήματος δεν διορθώνουν το πρόβλημα, ανεξάρτητα από το αρχικό γ.

Το φαινόμενο αυτό συμβαίνει διότι, ο εκθετικός όρος της συνάρτησης τείνει γρήγορα στο μηδέν, όσο τα x και y απομακρύνονται από την αρχή των αξόνων. Αυτό δημιουργεί επίπεδες περιοχές με κλίση σχεδόν μηδενική. Η κλίση είναι αρνητική προς την κατεύθυνση αυτής της επίπεδης περιοχής και έτσι ο αλγόριθμος “παγιδεύεται” στο πρώτο σημείο που θα συναντήσει και έχει κλίση ~ε (περίπου στην ανοχή).

Όσο αφορά την ταχύτητα σύγκλισης, τα συμπεράσματα είναι παρόμοια με το (-1,-1).

IV. ΘΕΜΑ 3: ΜΕΘΟΔΟΣ NEWTON

Η μέθοδος Newton χρησιμοποιεί πληροφορία δεύτερης παραγώγου (καμπυλότητα) μέσω του Εσσιανού Πίνακα $H(x_k) = \nabla^2 f(x_k)$. Η κατεύθυνση δίνεται από ($\Delta_k = [H(x_k)]^{-1}$):

$$d_k = -[H(x_k)]^{-1} \nabla f(x_k)$$

Όμως, έπειτα από υπολογισμό των μερικών παραγώγων της συνάρτησης f, της ορίζουσας του εσσιανού της πίνακα και των κύριων ελασσόνων, αποδεικνύεται ότι ο $\nabla^2 f(x_k)$ δεν είναι θετικά ορισμένος. Τα αποτελέσματα αυτού του φαινομένου απεικονίζονται στις γραφικές παραστάσεις που ακολουθούν.

Ακολουθούν τα πειράματα:

1) (0,0)

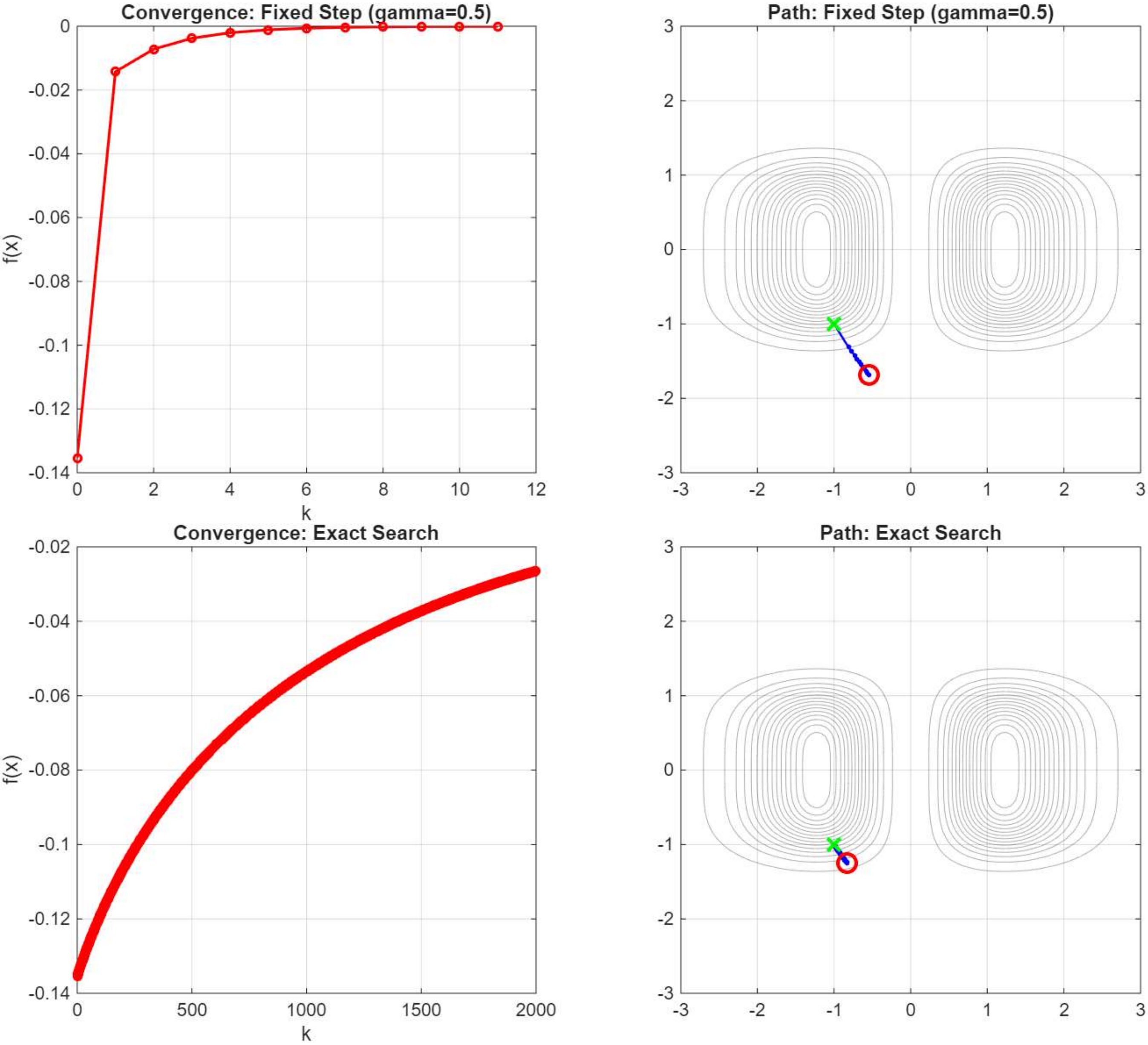
Προφανώς δεν έχουμε διαφορετικά αποτελέσματα από το θέμα 2 καθώς ο αλγόριθμος τερματίζει άμεσα.

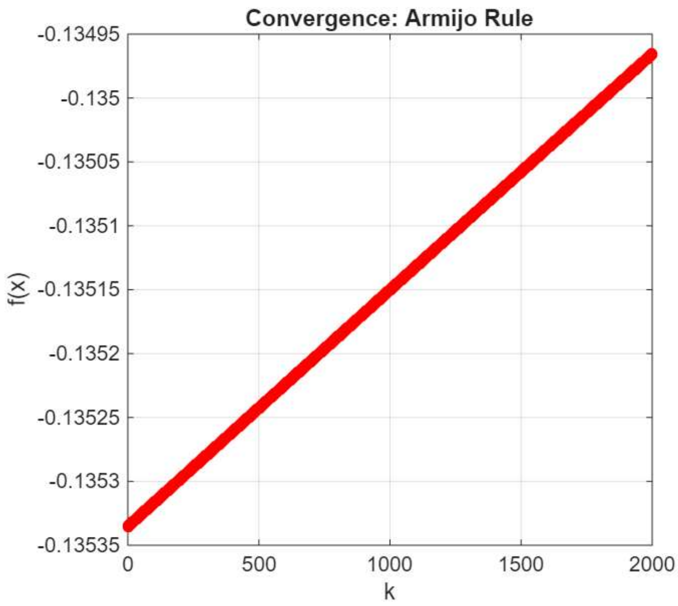
2) (-1,-1)

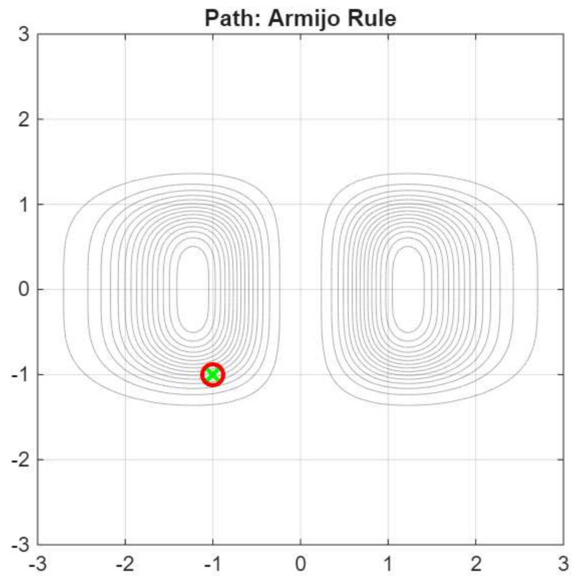
Για αρχικό σημείο (-1, -1) η μέθοδος newton συγκλίνει προς σημείο διαφορετικού του ολικού ελαχίστου, και μάλιστα δεν τερματίζει καν με τις τεχνικές “Exact Search” και Armijo. Ανεξαρτήτως της επιλογής αρχικού βήματος, και στις δύο τεχνικές, ο αλγόριθμος φαίνεται να προσεγγίζει σημείο της επιφάνειας “μηδενικής” κλίσης. Από τις τρείς τεχνικές, μονάχα η Fixed Step τερματίζει, η Exact Search κινείται προς την ίδια κατεύθυνση και η Armijo παραμένει στο ίδιο σχεδόν σημείο, που είναι όμως, σε μικρότερο ύψος από τις άλλες (και άρα κοντινότερο στο ολικό ελάχιστο).

Το φαινόμενο αυτό μάλλον οφείλεται στο γεγονός ότι η συνάρτηση παρουσιάζει αρνητική καμπυλότητα. Η “κατεύθυνση

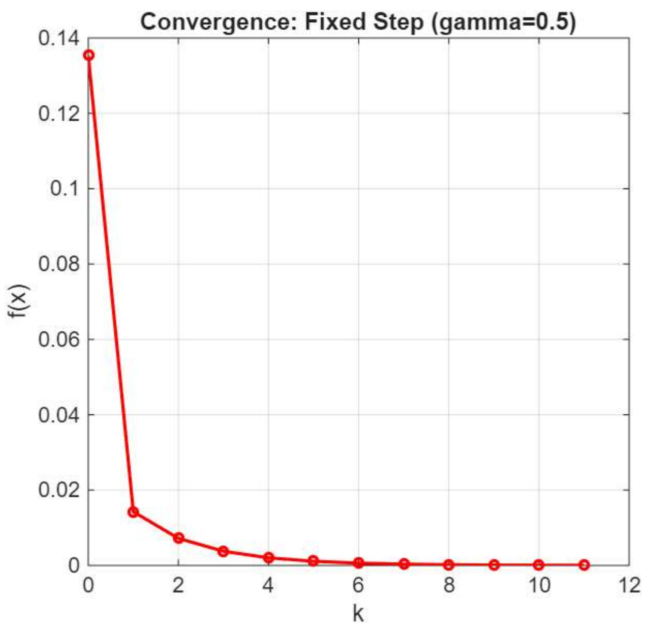
Newton" είναι λανθασμένη, καθώς ο εσσιανός είναι μη θετικά ορισμένος, οδηγώντας τον αλγόριθμο στο επίπεδο μηδενικής κλίσης.

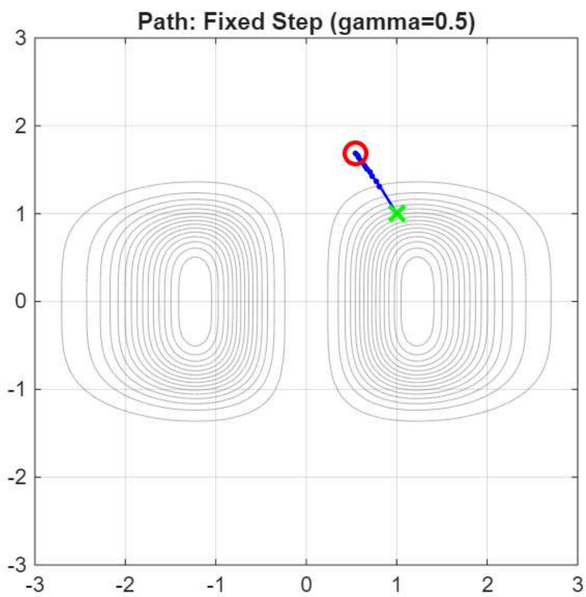


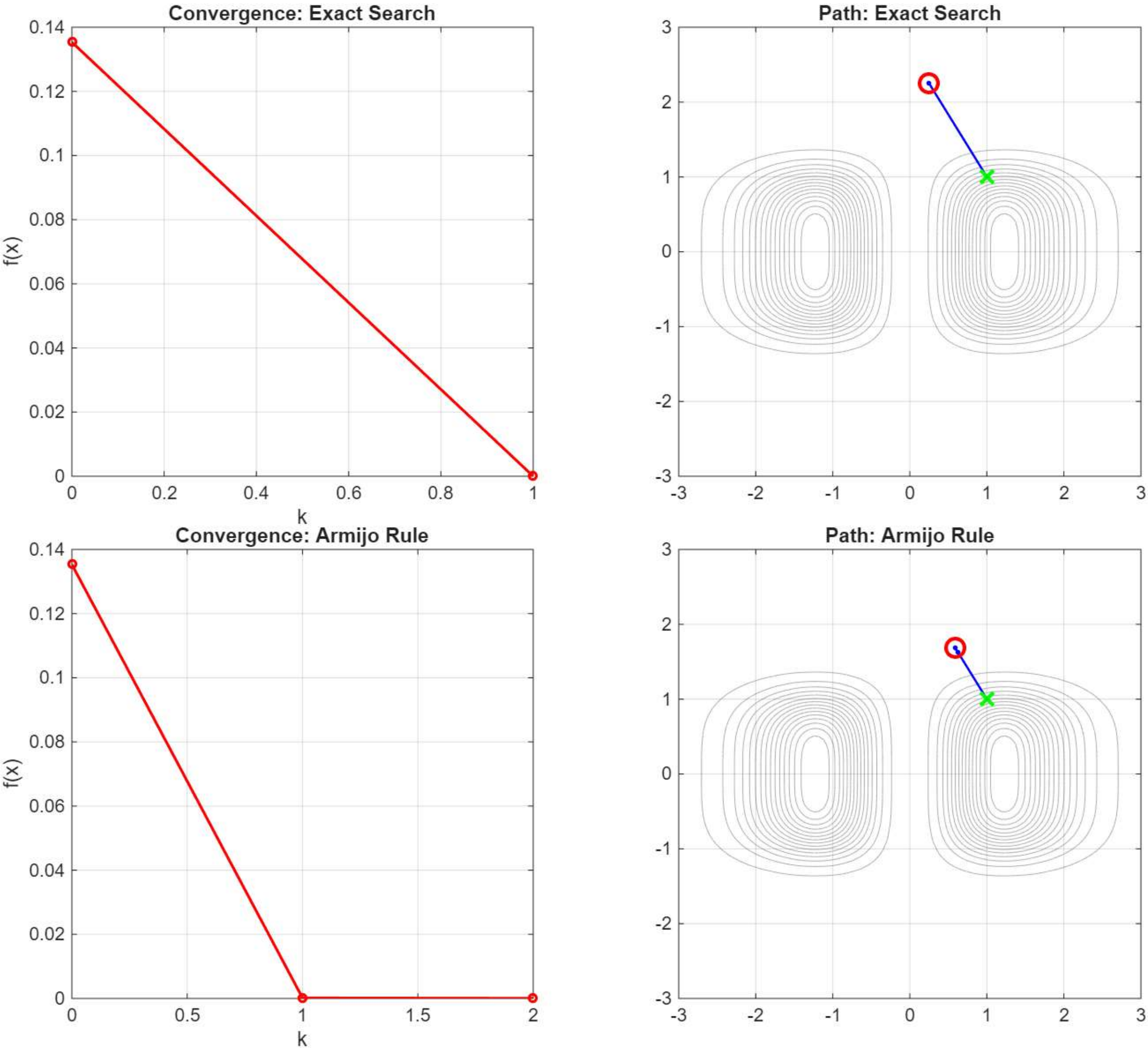

Convergence: Armijo Rule
f(x)
k
-0.13495
-0.135
-0.13505
-0.1351
-0.13515
-0.1352
-0.13525
-0.1353
-0.13535
0
500
1000
1500
2000


Path: Armijo Rule

3) (1, 1)


Convergence: Fixed Step (gamma=0.5)
f(x)
k
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
2
4
6
8
10
12


Path: Fixed Step (gamma=0.5)



Τα αποτελέσματα, με αρχικό σημείο (1, 1), είναι παρόμοια με την μέθοδο Μέγιστης Καθόδου. Με αξιοσημείωτη διαφορά, την βελτίωση του ρυθμό σύγκλισης προς το τοπικό "ελάχιστο", των Armijo και Fixed Step τεχνικών. Ωστόσο το πρόβλημα του εγκλωβισμού στην κοιλάδα μηδενικής κλίσης, παραμένει.

Αξίζει να σημειωθεί ότι παρόλο που η μέθοδος Newton απέτυχε να συγκλίνει στην συγκεκριμένη συνάρτηση (για τα εξεταζόμενα σημεία εκκίνησης), θεωρητικά προσφέρει γρηγορότερη σύγκλιση. Η αποτυχία οφείλεται στην απαίτηση για θετικά ορισμένο Εσσιανό πίνακα, καθώς και στο υψηλό υπολογιστικό κόστος ανά επανάληψη, που την καθιστά λιγότερο αποδοτική σε πολύπλοκα προβλήματα. Αντίθετα, η μέθοδος της μέγιστης καθόδου, αν και πιο αργή λόγω της γραμμικής σύγκλισης και των "ζιγκ-ζαγκ" κινήσεων, είναι γενικά πιο σταθερή (ως προς την σύγκλιση). Συμπερασματικά, κάθε μέθοδος πρέπει να εφαρμόζεται πάντα ανάλογα με τη φύση του προβλήματος που καλούμαστε να λύσουμε.

V. ΘΕΜΑ 4: ΜΕΘΟΔΟΣ LEVENBERG MARQUARDT

Η μέθοδος αυτή αποτελεί παραλλαγή της Newton, χρησιμοποιώντας έναν τροποποιημένο Εσσιανό Πίνακα:

$$H_{mod} = H(x_k) + \mu_k I$$

Όπου μk > 0 μία παράμετρος που επιλέγεται ώστε ο πίνακας να είναι θετικά ορισμένος ($\mu_k > |\lambda_{min}|, εδώ\ |\lambda min| + 0.1$), και άρα το dk ($\Delta_k = [\nabla^2 f(x_k) + \mu_k I]^{-1}$):

$$d_k = -[\nabla^2 f(x_k) + \mu_k I]^{-1} \nabla f(x_n)$$

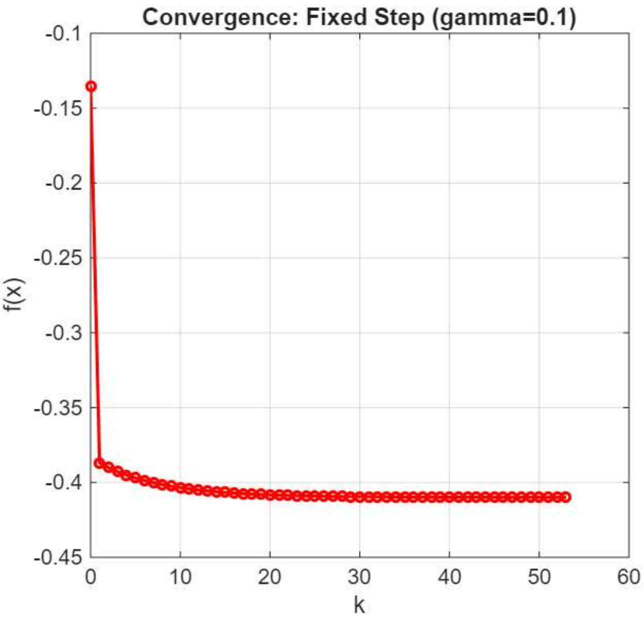
Ανάλογα με την επιλογή του μk, η μέθοδος Levenberg-Marquardt συμπεριφέρεται είτε παρόμοια με την μέθοδο Newton είτε παρόμοια με την μέθοδο Μέγιστης Καθόδου, καθώς ο όρος μk δίνει βάρος στο $\nabla^2 f(x_k)$ ή στο $\mu_k I$ αντίστοιχα.

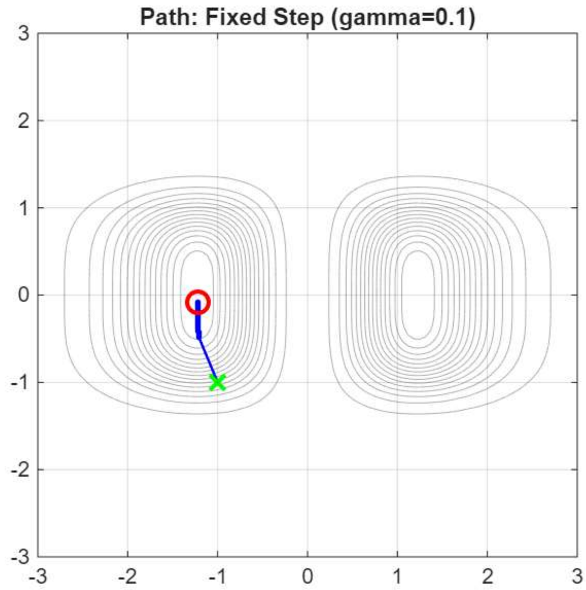
Ακολουθούν τα πειράματα:

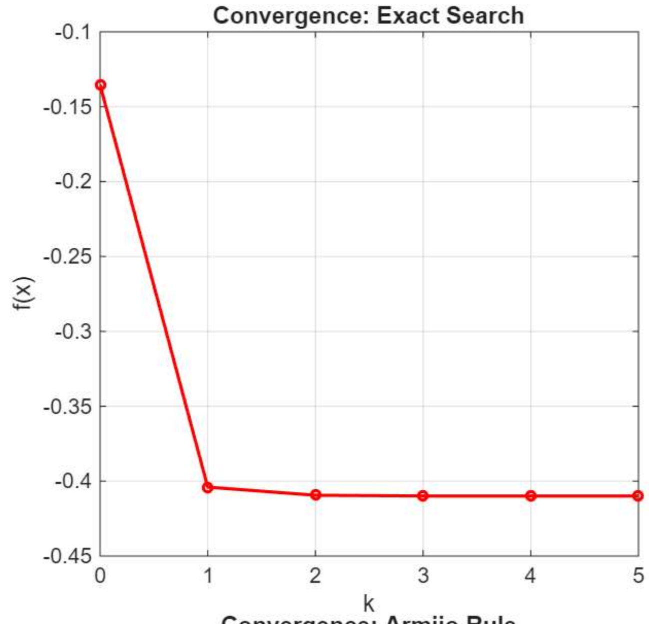
1) (0,0)

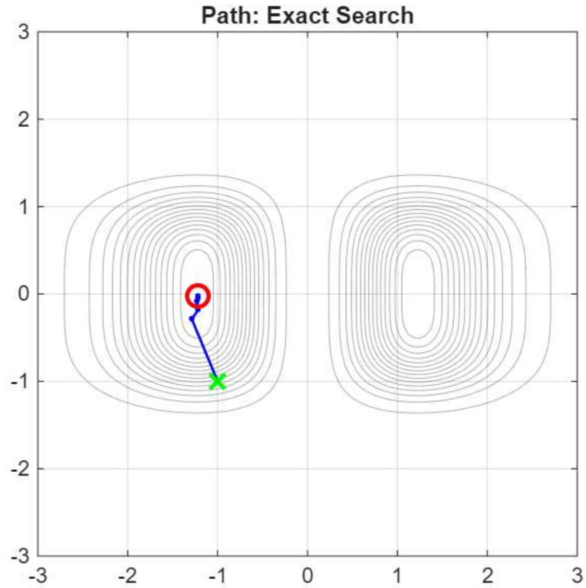
Προφανώς δεν έχουμε διαφορετικά αποτελέσματα από το θέμα 2 και 3, καθώς ο αλγόριθμος τερματίζει άμεσα.

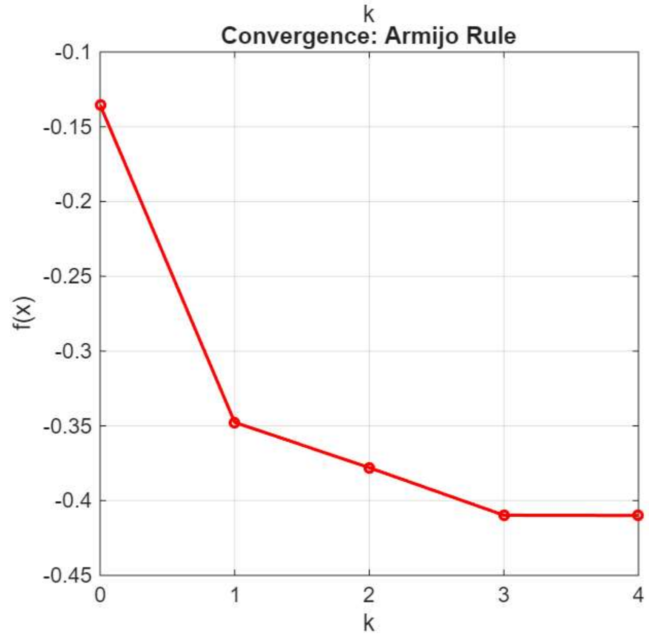
2) (-1,-1)

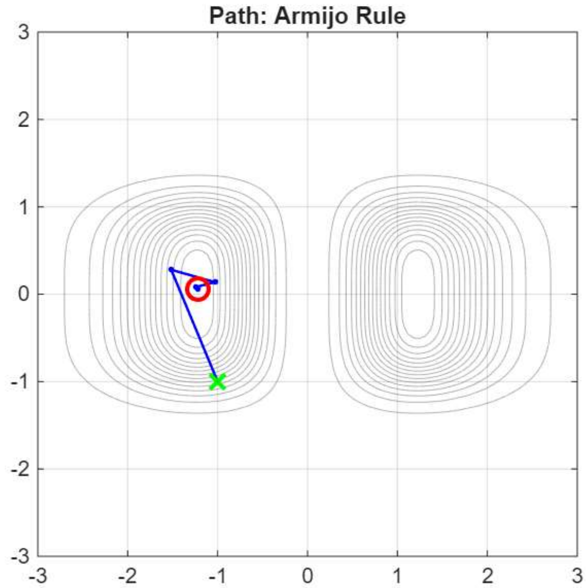




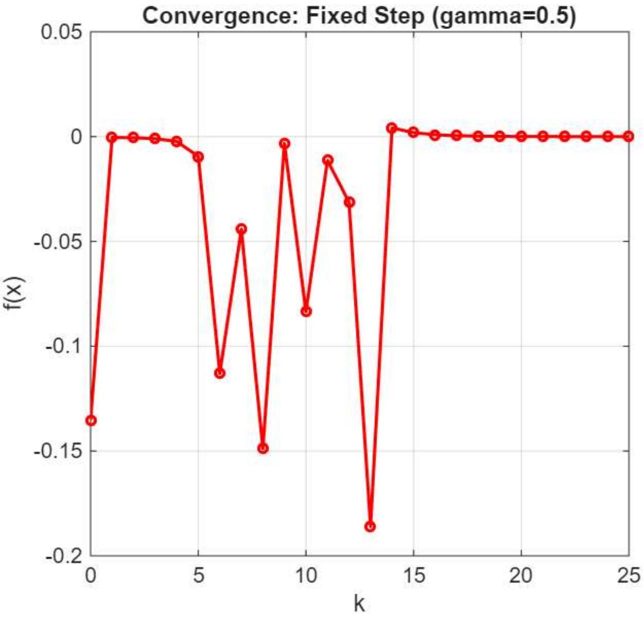


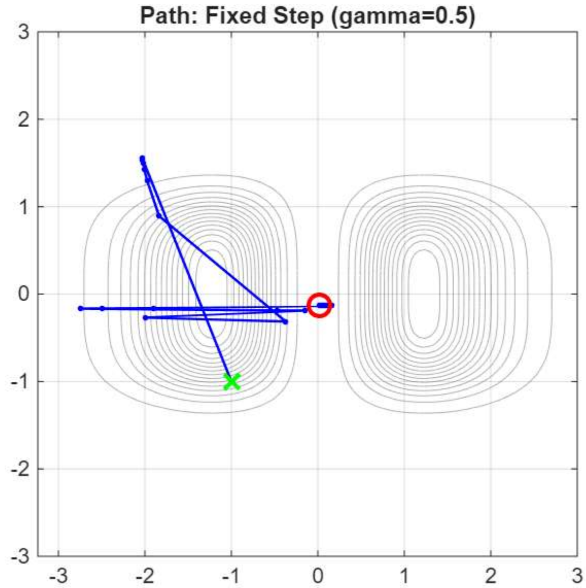






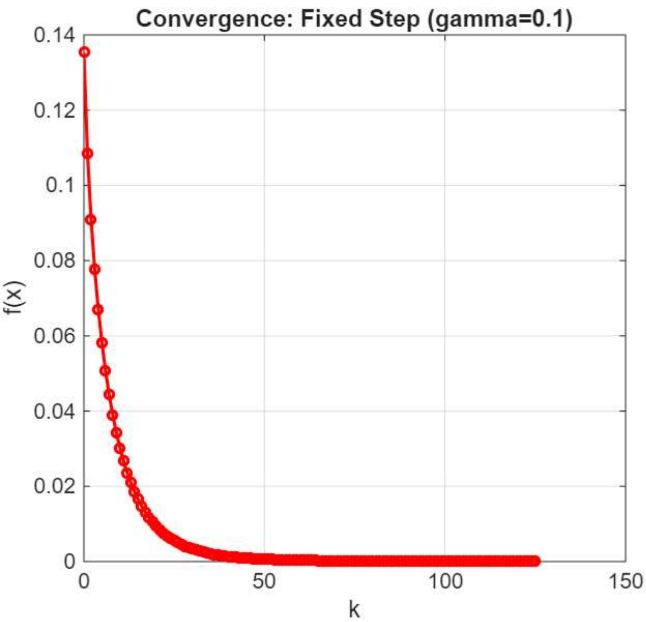
Τα συμπεράσματα που προκύπτουν από τις γραφικές παραστάσεις συμφωνούν με αυτά της μεθόδου της μέγιστης καθόδου. Και με τις τρείς τεχνικές ο αλγόριθμος συγκλίνει πολύ γρηγορότερα στο ολικό ελάχιστο, σε σχέση με την μέθοδο Μέγιστης Καθόδου. Ωστόσο από το παρακάτω πείραμα, με σταθερό βήμα 0.5, φαίνεται ότι είναι και πιο επιρρεπείς στην αλλαγή του βήματος.

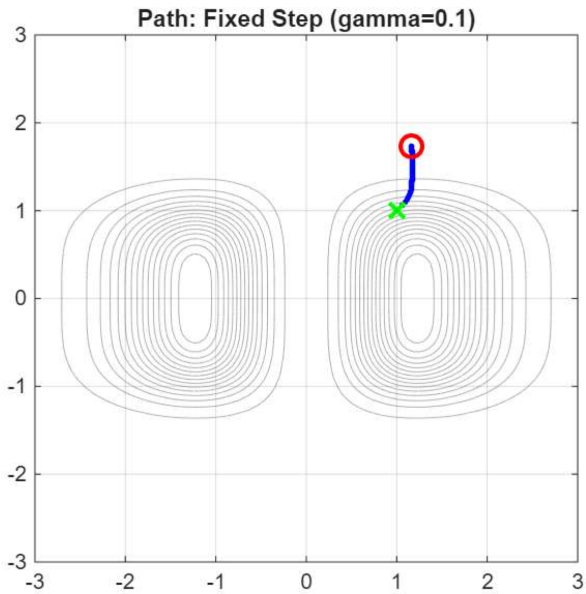


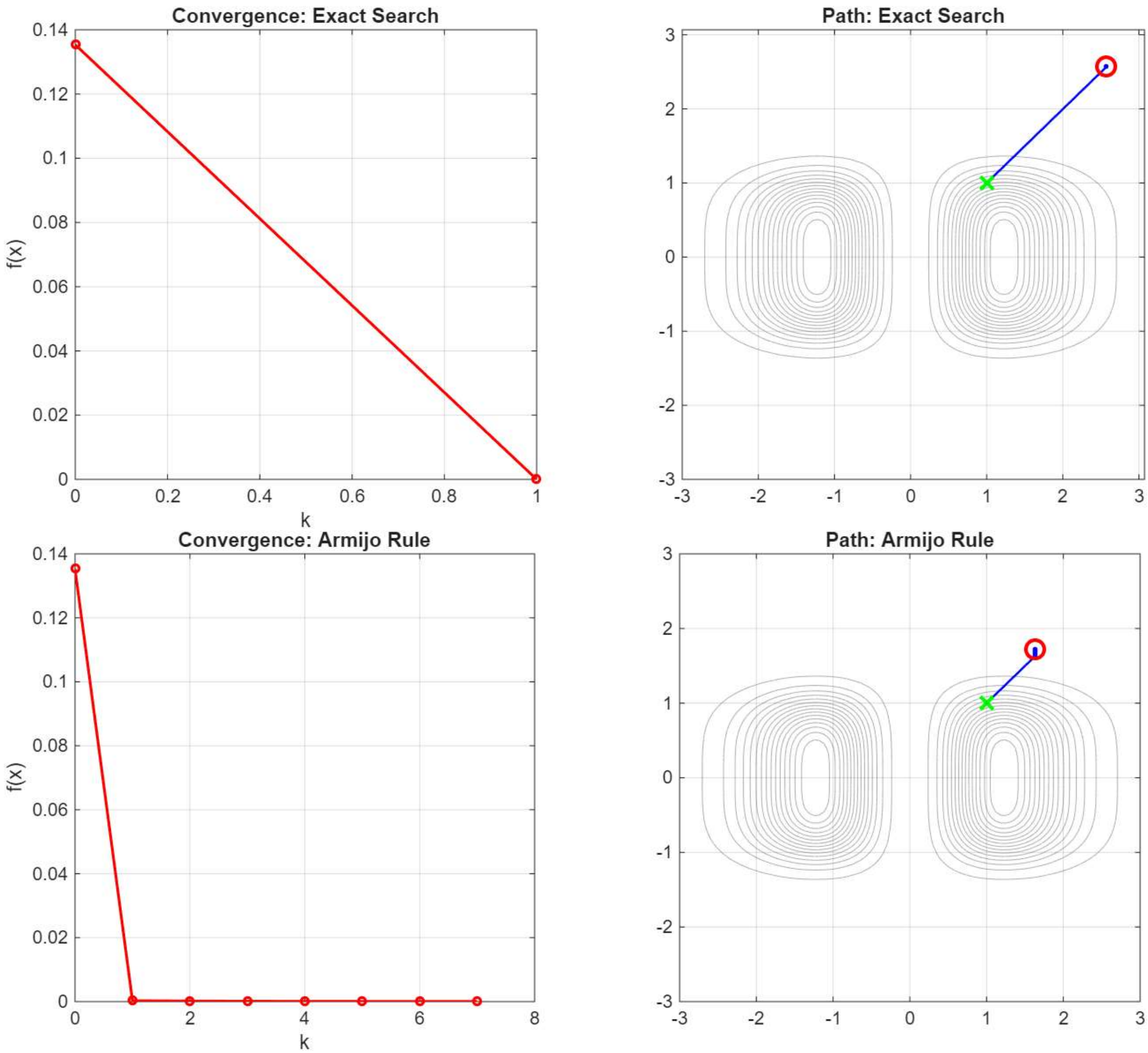


Σε αντίθεση με την μέγιστη κάθοδο, η μέθοδος Levenberg-Marquardt γίνεται ασταθής σε πολύ μικρότερο μέγεθος βήματος. Αλλαγές στο αρχικό γ της τεχνικής Armijo δεν προκάλεσαν προβλήματα στην σύγκλιση προς το ολικό ελάχιστο. Αντιθέτως για την "Exact Search" αρχική τιμή γ, πάνω από 2, οδηγεί τον αλγόριθμο στο να συγκλίνει στην "κοιλάδα" μηδενικής κλίσης.

3) (1, 1)







Επίσης όμοια με την Μέγιστη Κάθοδο, με την διαφορά την σύγκλισης σε λιγότερες επαναλήψεις. Μόνη εξαίρεση αποτελεί η "Exact Search" η οποία συγκλίνει σε μία επανάληψη και με τις δύο μεθόδους.

Δυστυχώς η μέθοδος Levenberg-Marquardt δεν βοήθησε στο να ξεπεράσουμε το πρόβλημα σύγκλισης προς την "κοιλάδα" μηδενικής κλίσης, ωστόσο διόρθωσε το πρόβλημα της Newton, στο αρχικό σημείο (-1,-1) χωρίς να θυσιάσει την ταχύτητα σύγκλισης.

VI. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΣΥΚΡΙΣΗ

Αρχικό Σημείο	Μέθοδος	Σταθερό Βήμα	Ακριβής Αναζήτηση	Κανόνας Armijo
(0, 0)	Steepest Descent	0	0	0
(0, 0)	Newton	0	0	0
(0, 0)	Levenberg-Marquardt	0	0	0
(-1, -1)	Steepest Descent	422	9	40
(-1, -1)	Newton	11 (Not Min)	2000 (Fail)	2000 (Fail)
(-1, -1)	Levenberg-Marquardt	53	5	4
(1, 1)	Steepest Descent	610 (Not Min)	1 (Not Min)	56 (Not Min)
(1, 1)	Newton	11 (Not Min)	1 (Not Min)	2 (Not Min)
(1, 1)	Levenberg-Marquardt	125 (Not Min)	1 (Not Min)	7 (Not Min)

Το **Not Min** υποδηλώνει ότι ο αλγόριθμος τερμάτισε σε σημείο διάφορο του ολικού ελαχίστου, και το **Fail** ότι τερμάτισε λόγο αριθμού επαναλήψεων.

Από τον παραπάνω πίνακα, την θεωρητική ανάλυση και τα αποτελέσματα των πειραμάτων μπορούμε να καταλήξουμε σε κάποια συμπεράσματα για τις συγκεκριμένες **τεχνικές επιλογής βήματος**:

- Η **Fixed Step** παρουσιάζει την βραδύτερη, με διαφορά, σύγκλιση και είναι η λιγότερο αξιόπιστη όταν έχουμε αλλαγές αρχικών συνθηκών (μέγεθος βήματος).
- Αντίθετα, η **"Exact Step"** (ελαχιστοποίηση $f(x_k + \gamma_k d_k)$) οδηγεί στην ταχύτερη σύγκλιση συγκρινόμενη με τις άλλες τεχνικές. Όμως αυξάνει το υπολογιστικό κόστος ανά επανάληψη.
- Τέλος, η **Armijo** απαιτεί λιγότερες επαναλήψεις από την Fixed αλλά περισσότερες από την Exact για να συγκλίνει, έχοντας όμως την μεγαλύτερη ανθεκτικότητα σε αλλαγές του αρχικού μεγέθους βήματος.

Και οι **μέθοδοι βελτιστοποίησης**:

- Η **Steepest Descent** εξασφαλίζει γενικά μεγάλη πιθανότητα σύγκλισης (αρκεί να υπάρχει κλίση), με χαμηλό υπολογιστικό κόστος ανά επανάληψη. Ωστόσο παρουσιάζει γραμμική σύγκλιση, πιο αργή από τις άλλες μεθόδους.
- Αντίθετα, η μέθοδος **Newton** προσφέρει τετραγωνική σύγκλιση αλλά απαιτεί θετικά ορισμένο εσσιανό, γεγονός που μας οδήγησε (στο συγκεκριμένο πρόβλημα) σε αποτυχία. Αξίζει επίσης να σημειωθεί ότι το υπολογιστικό κόστος ανά επανάληψη είναι υψηλότερο της steepest descent.
- Τέλος, η μέθοδος **Levenberg-Marquardt** συνδυάζει τα πλεονεκτήματα των δύο παραπάνω. Μακριά από τη λύση (ή σε μη κυρτές περιοχές) συμπεριφέρεται σαν steepest descent (ευστάθεια), ενώ κοντά στη λύση συμπεριφέρεται σαν Newton (ταχύτητα). Δεν απαιτεί θετικά ορισμένο Εσσιανό. Φαίνεται να είναι η πιο αποδοτική επιλογή για γενικά προβλήματα μη γραμμικής βελτιστοποίησης, καθώς εξισορροπεί τον αριθμό των επαναλήψεων με την ευστάθεια.

Η φύση του συγκεκριμένου προβλήματος, καθιστά τη μέθοδο Newton ακατάλληλη στην καθαρή της μορφή. Καταλήγουμε λοιπόν στην μέθοδο Levenberg-Marquardt ως την πιο αξιόπιστη λύση για την εύρεση του ολικού ελαχίστου, υπό την προϋπόθεση ότι αποφεύγονται οι περιοχές όπου η κλίση μηδενίζεται.