



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
“ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ & ΥΠΗΡΕΣΙΕΣ”

Μπεϋζιανή Βελτιστοποίηση για τη Ρύθμιση Υπερπαραμέτρων σε
Αλγορίθμους Μηχανικής Μάθησης

Από Γεώργιος Μαλισόβας

Υποβάλλεται για την εκπλήρωση των προϋποθέσεων λήψης
Μεταπτυχιακού Διπλώματος στην ειδίκευση «ΜΔΑ»
του ΠΜΣ “Πληροφορικά Συστήματα & Υπηρεσίες”
στο ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
Ιούλιος 2026

Επιβλέπων: Χ. Δουλκερίδης
Ακαδημαϊκή Θέση: Καθηγητής

Πανεπιστήμιο Πειραιώς. Κάτοχος όλων των δικαιωμάτων
University of Piraeus. All rights reserved.

Συγγραφέας: Γεώργιος Μαλισόβας

Ευχαριστίες

Με την ολοκλήρωση της παρούσας μεταπτυχιακής διπλωματικής εργασίας, αισθάνομαι την ανάγκη να ευχαριστήσω θερμά τον καθηγητή μου και επιβλέποντα στην διπλωματική εργασία, κ. Χρήστο Δουλκερίδη, για την πολύτιμη καθοδήγηση και την εμπιστοσύνη που μου έδειξε καθ' όλη τη διάρκεια εκπόνησης της έρευνας. Η επιστημονική του συγκρότηση και οι καίριες υποδείξεις του υπήρξαν καθοριστικές για την επιτυχή διεκπεραίωση της εργασίας αυτής.

Επιπλέον, θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες προς τους υποψήφιους διδάκτορες, Δημήτρη Πετράτο και Ιωάννη Πουλάκη, για την ουσιαστική βοήθεια που μου προσέφεραν σε διάφορα στάδια της πειραματικής και θεωρητικής διαδικασίας.

Πίνακας Περιεχομένων

Ευχαριστίες	3
Πίνακας Περιεχομένων	4
Κατάλογος Εικόνων	5
Κατάλογος Πινάκων	5
Περίληψη	6
Abstract	7
1. Εισαγωγή	8
1.1 Η Ανάγκη για Μπεϋζιανή Βελτιστοποίηση	8
1.2 Στόχος της Εργασίας	9
2. Η Μπεϋζιανή Βελτιστοποίηση	11
2.1 Η Συνάρτηση Απόκτησης (Acquisition Function)	12
2.2 Βήματα της Μπεϋζιανής Βελτιστοποίησης	13
3. Σχετική Έρευνα	16
4. Μεθοδολογία	19
4.1 Δομή και Συνιστώσες του Πλαισίου Μπεϋζιανής Βελτιστοποίησης	19
4.2 Οι περιπτώσεις των κατηγορικών μεταβλητών	20
5. Πειραματική Αξιολόγηση και Αποτελέσματα	22
5.1 Πειραματικό Πλαίσιο και Μεθοδολογία Αξιολόγησης	22
5.2 Αξιολόγηση σε Μονοδιάστατες (1D) Συναρτήσεις	23
5.3 Multidimensional Συναρτήσεις	28
5.3.1 Ackley	28
5.3.2 Booth	34
5.3.3 Colville	36
5.3.4 Three-hump Camel	39
5.3.5 Styblinski-Tang	41
5.4 Optuna & HyperOpt vs Multidimensional Συναρτήσεις	47
5.4.1 Ackley	48
5.4.2 Booth-Colville-Three hump Camel	51
5.4.3 Styblinski-Tang	54
5.5 Agglomerative Clustering Αλγόριθμος	57
5.6 Σχολιασμός των Πειραματικών Αποτελεσμάτων	58
6. Συμπεράσματα	61
Βιβλιογραφία	62

Κατάλογος Εικόνων

Εικόνα 1. $f_1(x)$ – Matern	25
Εικόνα 2. $f_2(x)$ – RBF.....	25
Εικόνα 3. $f_3(x)$ – Rational Quadratic.....	26
Εικόνα 4. $f_4(x)$ – Matern	26
Εικόνα 5. Ackley (2D απεικόνιση).....	28
Εικόνα 6. Ackley (2D).....	30
Εικόνα 7. Booth (2D απεικόνιση)	34
Εικόνα 8. Booth	36
Εικόνα 10. Three-hump Camel (2D απεικόνιση)	39
Εικόνα 11. Three-hump Camel.....	40
Εικόνα 12. Styblinski-Tang (2D απεικόνιση).....	41
Εικόνα 13. Styblinski-Tang (2D)	42
Εικόνα 14. Ackley (2D)-EI vs Optuna/HyperOpt.....	49
Εικόνα 15. Ackley (5D)-PI vs Optuna/HyperOpt.....	49
Εικόνα 16. Ackley (10D)-PI vs Optuna/HyperOpt.....	50
Εικόνα 17. Booth-EI vs Optuna/HyperOpt	52
Εικόνα 18. Colville-EI vs Optuna/HyperOpt	52
Εικόνα 19. Three-hump Camel-Optuna/HyperOpt	53
Εικόνα 20. Styblinski-Tang (2D)-PI vs Optuna/HyperOpt.....	55
Εικόνα 21. Styblinski-Tang (5D)-Optuna/HyperOpt.....	55
Εικόνα 22. Styblinski-Tang (10D)-PI vs Optuna/HyperOpt.....	56

Κατάλογος Πινάκων

Πίνακας 1. 1D Συναρτήσεις – χωρίς θόρυβο	24
Πίνακας 2. 1D Συναρτήσεις – Με θόρυβο	27
Πίνακας 3. Ackley	29
Πίνακας 4. Ackley (5D) – Χωρίς θόρυβο.....	30
Πίνακας 5. Ackley (5D) – Με θόρυβο	31
Πίνακας 6. Ackley (10D) – Χωρίς θόρυβο	32
Πίνακας 7. Ackley (10D) – Με θόρυβο	33
Πίνακας 8. Booth	35
Πίνακας 9. Colville (χωρίς θόρυβο)	37
Πίνακας 10. Colville (Με θόρυβο)	37
Πίνακας 11. Three-hump Camel.....	39
Πίνακας 12. Styblinski-Tang (2D).....	41
Πίνακας 13. Styblinski-Tang (5D) - Χωρίς θόρυβο.....	43
Πίνακας 14. Styblinski-Tang (5D) – Με θόρυβο	44
Πίνακας 15. Styblinski-Tang (10D) - Χωρίς θόρυβο.....	45
Πίνακας 16. Styblinski-Tang (10D) - Με θόρυβο	46
Πίνακας 17. Σύγκριση Ackley με Optuna-HyperOpt (χωρίς θόρυβο).....	48
Πίνακας 18. Σύγκριση Booth-Colville-Three hump Camel με Optuna-HyperOpt (χωρίς θόρυβο)	51
Πίνακας 19. Σύγκριση Styblinski-Tang με Optuna-HyperOpt (χωρίς θόρυβο).....	54
Πίνακας 20. Σύγκριση Custom GP's με Optuna-HyperOpt (Agglomerative Clustering)	57

Περίληψη

Η παρούσα μεταπτυχιακή διπλωματική εργασία εξετάζει τη μεθοδολογία της Μπεϋζιανής Βελτιστοποίησης (Bayesian Optimization) για τη διαχείριση υπολογιστικά δαπανηρών συναρτήσεων «μαύρου κουτιού» (black-box). Το κεντρικό ερευνητικό ερώτημα εστιάζει στη συγκριτική αξιολόγηση εξατομικευμένων υλοποιήσεων —βασισμένων σε Gaussian Processes με διαφορετικούς πυρήνες και συναρτήσεις απόκτησης— έναντι καθιερωμένων βιβλιοθηκών αυτοματοποιημένης μηχανικής μάθησης, όπως το Optuna και το HyperOpt.

Η κύρια επιστημονική συνεισφορά της έρευνας έγκειται στην ανάπτυξη ενός παραμετροποιήσιμου πλαισίου βελτιστοποίησης και στην εκτενή ανάλυση της συμπεριφοράς του, τόσο σε μαθηματικές συναρτήσεις αναφοράς ποικίλων διαστάσεων υπό την παρουσία θορύβου, όσο και σε ένα πραγματικό πρόβλημα ρύθμισης υπερπαραμέτρων για τον αλγόριθμο ιεραρχικής συσταδοποίησης (Agglomerative Clustering).

Τα αποτελέσματα αναδεικνύουν μια σαφή σχέση συμβιβασμού μεταξύ ακρίβειας και ταχύτητας. Οι ιδιοκατασκευασμένες προσεγγίσεις υπερέχουν σταθερά σε ακρίβεια, προσεγγίζοντας με εξαιρετική πιστότητα το ολικό βέλτιστο. Αντίθετα, οι έτοιμες βιβλιοθήκες προσφέρουν σημαντικά μεγαλύτερη υπολογιστική αποδοτικότητα και ταχύτητα εκτέλεσης, ειδικότερα σε προβλήματα υψηλής διαστασιμότητας.

Οι τομείς εφαρμογής των ερευνητικών αποτελεσμάτων εκτείνονται στην Επιστήμη των Δεδομένων για την αυτόματη ρύθμιση σύνθετων μοντέλων, καθώς και σε βιομηχανικά προβλήματα όπου η αξιολόγηση των πραγματικών συναρτήσεων είναι οικονομικά ή χρονικά ασύμφορη.

Abstract

This master's thesis investigates Bayesian Optimization for handling computationally expensive black-box functions. The core research question focuses on evaluating custom implementations —based on Gaussian Processes with various kernels and acquisition functions— against established automated machine learning libraries, specifically Optuna and HyperOpt.

The main scientific contribution lies in the development of a fully customizable optimization framework and its extensive performance analysis. The methodology was evaluated on mathematical benchmark functions of varying dimensions under noisy conditions, as well as on a real-world hyperparameter tuning task for the Agglomerative Clustering algorithm.

The results reveal a distinct trade-off between accuracy and computational efficiency. The custom implementations consistently outperform standard libraries in terms of solution precision, locating values remarkably close to the global optimum. Conversely, standard libraries offer vastly superior execution speed and scalability as the problem dimensionality increases.

The application domains of these outcomes span across Data Science for automated model tuning, as well as industrial optimization problems where evaluating the objective function is financially or computationally prohibitive.

1. Εισαγωγή

Η βελτιστοποίηση αποτελεί έναν θεμελιώδη κλάδο των μαθηματικών και της επιστήμης των υπολογιστών, που ασχολείται με την εύρεση της καλύτερης δυνατής λύσης σε ένα πρόβλημα, σύμφωνα με κάποιο αντικειμενικό κριτήριο. Τυπικά, ένα πρόβλημα βελτιστοποίησης διατυπώνεται ως:

$$\min_{x \in X} f(x)$$

όπου $f(x)$ είναι η αντικειμενική συνάρτηση που θέλουμε να ελαχιστοποιήσουμε (ή να μεγιστοποιήσουμε), και X είναι ο χώρος των επιτρεπτών λύσεων.

Προβλήματα βελτιστοποίησης συναντώνται σχεδόν σε κάθε επιστημονικό και τεχνολογικό πεδίο: από τη μείωση του κόστους παραγωγής στη βιομηχανία και τη βελτίωση της απόδοσης αλγορίθμων μηχανικής μάθησης, έως τη βελτιστοποίηση της κατανάλωσης ενέργειας σε έξυπνα δίκτυα^{[1][2]}. Για τον λόγο αυτό, η ανάπτυξη αποδοτικών και ευέλικτων μεθόδων βελτιστοποίησης αποτελεί διαρκή ερευνητική πρόκληση. Οι μέθοδοι βελτιστοποίησης μπορούν να ταξινομηθούν ανάλογα με τη φύση της αντικειμενικής συνάρτησης (γραμμική ή μη γραμμική, κυρτή ή μη κυρτή), τη διαθεσιμότητα παραγώγων (με ή χωρίς χρήση gradient), τον ντετερμινιστικό ή στοχαστικό χαρακτήρα τους, καθώς και τη διάσταση του χώρου αναζήτησης (μονοδιάστατες ή πολυδιάστατες συναρτήσεις).

Οι κλασικές μέθοδοι, όπως η κατιούσα κλίση (Gradient Descent) και ο αλγόριθμος Newton-Raphson, επιτυγχάνουν καλά αποτελέσματα όταν η συνάρτηση $f(x)$ είναι κυρτή και διαφορίσιμη. Ωστόσο, αποτυγχάνουν ή καθίστανται αναποτελεσματικές όταν η $f(x)$ είναι συνάρτηση μαύρου κουτιού (black-box), δηλαδή όταν δεν υπάρχει αναλυτική έκφρασή της ή δεν είναι δυνατόν να υπολογιστούν οι παράγωγοί της^{[3][4]}. Παρόμοια προβλήματα προκύπτουν και όταν η αντικειμενική συνάρτηση είναι μη παραμετρική, κοστοβόρα στην αξιολόγηση, ή θορυβώδης (περιέχει στοχαστικά στοιχεία ή αβεβαιότητα στις μετρήσεις).

Σε τέτοιες περιπτώσεις, καθίσταται αναγκαία η χρήση πιο ευέλικτων και στοχαστικών μεθόδων, οι οποίες βασίζονται είτε σε τυχαία αναζήτηση (όπως το Random Search), είτε σε model-based προσεγγίσεις, όπως η Μπεϋζιανή Βελτιστοποίηση (Bayesian Optimization).

1.1 Η Ανάγκη για Μπεϋζιανή Βελτιστοποίηση

Σε πολλές σύγχρονες εφαρμογές, η αντικειμενική συνάρτηση που πρέπει να βελτιστοποιηθεί δεν διαθέτει αναλυτική μορφή – για παράδειγμα, πρόκειται για ένα πολύπλοκο μοντέλο προσομοίωσης ή έναν αλγόριθμο μηχανικής μάθησης του οποίου η εσωτερική λειτουργία δεν είναι πλήρως γνωστή. Επιπλέον, η αξιολόγηση αυτής της συνάρτησης μπορεί να είναι ιδιαίτερα κοστοβόρα, αφού ενδέχεται να απαιτεί την εκπαίδευση ενός ολόκληρου μοντέλου ή την εκτέλεση μιας χρονοβόρας

προσομοίωσης. Σε άλλες περιπτώσεις, η συνάρτηση περιλαμβάνει θόρυβο ή αβεβαιότητα, καθιστώντας τη διαδικασία βελτιστοποίησης ακόμα πιο δύσκολη.

Οι παραδοσιακές τεχνικές, όπως το Grid Search, το Random Search και οι μέθοδοι κατιούσας κλίσης (Gradient Descent), αποδεικνύονται συχνά ανεπαρκείς ή εξαιρετικά αργές σε τέτοιου είδους προβλήματα. Αυτό οδήγησε στην ανάπτυξη της Μπεϋζιανής Βελτιστοποίησης (Bayesian Optimization – BO), μιας πιθανοτικής μεθόδου που έχει σχεδιαστεί ειδικά για τη βελτιστοποίηση «black-box συναρτήσεων» με υψηλό υπολογιστικό κόστος.

Η βασική ιδέα της Μπεϋζιανής Βελτιστοποίησης είναι να κατασκευαστεί ένα πιθανοτικό μοντέλο (συνήθως ένας Gaussian Process, αν και υπάρχουν και εναλλακτικές όπως ο Tree-structured Parzen Estimator – TPE) που προσεγγίζει τη συμπεριφορά της αντικειμενικής συνάρτησης. Στη συνέχεια, το μοντέλο αυτό χρησιμοποιείται για να καθοδηγήσει την επιλογή των σημείων όπου θα γίνει αξιολόγηση της $f(x)$, επιτυγχάνοντας μια ισορροπία μεταξύ εκμετάλλευσης (δοκιμή σημείων που φαίνεται να είναι υποσχόμενα) και εξερεύνησης (δοκιμή νέων περιοχών του χώρου αναζήτησης).

Η επιλογή του επόμενου σημείου γίνεται μέσω μιας acquisition function, όπως το *Expected Improvement (EI)*, το *Probability of Improvement (PI)* ή το *Upper Confidence Bound (UCB)*. Με τον τρόπο αυτό, η Μπεϋζιανή Βελτιστοποίηση καταφέρνει να βρει το ολικό βέλτιστο σημείο με πολύ λιγότερες αξιολογήσεις της πραγματικής συνάρτησης σε σύγκριση με άλλες μεθόδους, γεγονός που την καθιστά ιδανική για προβλήματα όπου κάθε αξιολόγηση είναι δαπανηρή ή χρονοβόρα^[7].

Η Μπεϋζιανή Βελτιστοποίηση έχει βρει εκτεταμένη εφαρμογή στη βελτιστοποίηση υπερπαραμέτρων αλγορίθμων μηχανικής μάθησης^[7], στη ρύθμιση προσομοιώσεων σε φυσική, βιολογία και ρομποτική, στον σχεδιασμό πειραμάτων στη βιοτεχνολογία, καθώς και στην αυτόματη ρύθμιση (tuning) αλγορίθμων σε συστήματα AutoML όπως τα Optuna, HyperOpt και SMAC^[8].

Τέλος, σκοπός της παρούσας εργασίας είναι να μελετήσει την εφαρμογή της Μπεϋζιανής Βελτιστοποίησης σε διαφορετικά σενάρια, διερευνώντας τη συμπεριφορά της υπό διαφορετικές συναρτήσεις, kernels, και acquisition functions, καθώς και να συγκρίνει τη μεθοδολογία της με αυτήν των σύγχρονων βιβλιοθηκών βελτιστοποίησης.

1.2 Στόχος της Εργασίας

Αντικείμενο και κύριος στόχος της παρούσας μεταπτυχιακής διπλωματικής εργασίας είναι η σχεδίαση, η ανάπτυξη και η συστηματική συγκριτική αξιολόγηση ενός παραμετροποιήσιμου πλαισίου (framework) Μπεϋζιανής Βελτιστοποίησης για την αυτοματοποιημένη ρύθμιση υπερπαραμέτρων σε αλγορίθμους Μηχανικής Μάθησης. Ιδιαίτερη έμφαση δίνεται στη μελέτη και τη σύγκριση διαφορετικών στρατηγικών εξισορρόπησης μεταξύ της εξερεύνησης άγνωστων περιοχών

(exploration) και της εκμετάλλευσης γνωστών βέλτιστων περιοχών (exploitation) του χώρου αναζήτησης.

Για την επίτευξη του στόχου αυτού, η μεθοδολογία που ακολουθήθηκε βασίστηκε αρχικά στην υλοποίηση ενός custom πλαισίου Μπεϋζιανής Βελτιστοποίησης που στηρίζεται σε Γκαουσιανές Διαδικασίες (Gaussian Processes) και αναπτύχθηκε εξ ολοκλήρου από το μηδέν σε περιβάλλον Python, ενσωματώνοντας εναλλακτικές συναρτήσεις απόκτησης (acquisition functions), όπως τις Expected Improvement, Probability of Improvement, Upper Confidence Bound και Thompson Sampling. Παράλληλα, αναπτύχθηκαν και ενσωματώθηκαν εξειδικευμένες τεχνικές κωδικοποίησης, όπως το One-Hot, το Hash και το Target Encoding, με σκοπό τη δυναμική διαχείριση κατηγορικών μεταβλητών εντός του αλγορίθμου.

Η προτεινόμενη μεθοδολογία υποβλήθηκε σε εκτενή πειραματική αξιολόγηση, η οποία πραγματοποιήθηκε τόσο σε πολυδιάστατες μαθηματικές συναρτήσεις αναφοράς (benchmarks) όσο και σε ένα πραγματικό πρόβλημα ιεραρχικής συσταδοποίησης (Agglomerative Clustering). Τέλος, η ερευνητική πορεία ολοκληρώθηκε με την άμεση αντιπαραβολή και σύγκριση των αποτελεσμάτων του custom πλαισίου με καταξιωμένες βιομηχανικές βιβλιοθήκες υπερπαραμετροποίησης της αγοράς, όπως το Optuna και το HyperOpt, αναδεικνύοντας τα πλεονεκτήματα και τους συμβιβασμούς (trade-offs) κάθε προσέγγισης.

2. Η Μπεϋζιανή Βελτιστοποίηση

Η Μπεϋζιανή Βελτιστοποίηση (Bayesian Optimization, BO) αποτελεί μία στοχαστική, model-based προσέγγιση για την εύρεση του ολικού ακροτάτου μιας black-box συνάρτησης $f(x)$, η οποία είναι δαπανηρή στην αξιολόγηση και στερείται αναλυτικής μορφής^[9]. Η βασική ιδέα πίσω από τη μέθοδο έγκειται στη δημιουργία ενός πιθανοτικού μοντέλου που προσεγγίζει τη συνάρτηση στόχου και στην αξιοποίησή του για την καθοδήγηση της επιλογής των επόμενων σημείων αξιολόγησης. Η διαδικασία αυτή συνδυάζει αποτελεσματικά τη γνώση που αποκτάται από προηγούμενες παρατηρήσεις με την εξερεύνηση περιοχών του χώρου που παραμένουν αβέβαιες, επιτυγχάνοντας έναν ισορροπημένο συνδυασμό εκμετάλλευσης και εξερεύνησης^[10].

Στην πράξη, το πιο διαδεδομένο πιθανοτικό μοντέλο που χρησιμοποιείται στη Μπεϋζιανή Βελτιστοποίηση είναι η Gaussian Process (GP)^[12]. Μια Gaussian Process ορίζει μια κατανομή πιθανότητας πάνω σε συναρτήσεις και καθορίζεται πλήρως από δύο συναρτήσεις: τη μέση συνάρτηση $m(x)$ και τη συνάρτηση συσχέτισης ή kernel $k(x, x')$. Συνήθως υποθέτουμε ότι η μέση συνάρτηση είναι μηδενική χωρίς απώλεια γενικότητας, ενώ ο kernel εκφράζει τη συσχέτιση μεταξύ των τιμών της συνάρτησης σε διαφορετικά σημεία του χώρου εισόδου.

Συχνά χρησιμοποιούνται συγκεκριμένες μορφές του kernel, όπως ο *Squared Exponential (RBF)*, ο οποίος επιβάλλει ομαλότητα στη συνάρτηση και δίνεται από τη σχέση:

$$k(x, x') = \sigma^2 \exp\left(-\frac{|x - x'|^2}{2l^2}\right)$$

ή ο *Matern kernel*, που επιτρέπει λιγότερο ομαλές συναρτήσεις και ορίζεται ως:

$$k(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}|x - x'|}{l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}|x - x'|}{l}\right)$$

Εναλλακτικά, μπορεί να χρησιμοποιηθεί και ο *Rational Quadratic kernel*, ο οποίος αντιστοιχεί σε μείγμα πολλών RBF kernels διαφορετικών κλιμάκων. Η επιλογή του kernel επηρεάζει σημαντικά τη συμπεριφορά της βελτιστοποίησης, καθώς καθορίζει τις υποθέσεις που κάνουμε σχετικά με την ομαλότητα και τη δομή συσχέτισης της συνάρτησης $f(x)$.

Αφού συλλεχθούν δεδομένα $D = \{(x_i, y_i)\}_{i=1}^n$, η Gaussian Process επιτρέπει τον υπολογισμό της εκ των υστέρων (posterior) κατανομής της συνάρτησης στο κάθε νέο σημείο x_* :

$$f(x_*)|D \sim N(\mu(x_*), \sigma^2(x_*))$$

όπου η μέση τιμή και η διακύμανση δίνονται από τις σχέσεις:

$$\mu(x_*) = k_*^T (K + \sigma_n^2 I)^{-1} y, \sigma^2(x_*) = k(x_*, x_*) - k_*^T (K + \sigma_n^2 I)^{-1} k_*$$

Ο πίνακας K περιέχει τις συσχετίσεις μεταξύ των παρατηρούμενων σημείων, το διάνυσμα k_* εκφράζει τη συσχέτιση του νέου σημείου με τα δεδομένα, ενώ το σ_n^2 αντιπροσωπεύει τη διακύμανση του θορύβου. Με τον τρόπο αυτό, το μοντέλο μάς παρέχει ταυτόχρονα μια εκτίμηση της αναμενόμενης τιμής της συνάρτησης και του βαθμού αβεβαιότητας που τη συνοδεύει — ένα στοιχείο κρίσιμο για την επόμενη φάση της διαδικασίας βελτιστοποίησης.

2.1 Η Συνάρτηση Απόκτησης (Acquisition Function)

Η συνάρτηση απόκτησης (acquisition function) αποτελεί το δεύτερο θεμελιώδες στοιχείο της Μπεϋζιανής Βελτιστοποίησης, καθώς καθορίζει ποιο σημείο θα αξιολογηθεί στη συνέχεια. Η επιλογή αυτή βασίζεται στην εκ των υστέρων κατανομή του μοντέλου και επιχειρεί να εξισορροπήσει την εξερεύνηση περιοχών με υψηλή αβεβαιότητα με την εκμετάλλευση σημείων που φαίνονται υποσχόμενα.

Μια από τις πιο δημοφιλείς επιλογές είναι η *Expected Improvement (EI)*^[11], η οποία μετρά την αναμενόμενη βελτίωση σε σχέση με την καλύτερη τιμή που έχει βρεθεί μέχρι στιγμής. Ορίζεται μαθηματικά ως:

$$EI(x) = (f(x^+) - \mu(x))\Phi(z) + \sigma(x)\varphi(z)$$

όπου $z = \frac{f(x^+) - \mu(x)}{\sigma(x)}$, και Φ και φ αντιστοιχούν στην σωρευτική και στην πυκνότητα της κανονικής κατανομής αντίστοιχα.

Άλλη γνωστή επιλογή είναι η *Probability of Improvement (PI)*, η οποία υπολογίζει την πιθανότητα το νέο σημείο x να βελτιώσει την καλύτερη μέχρι στιγμής τιμή της αντικειμενικής συνάρτησης. Πιο συγκεκριμένα, εκφράζει την πιθανότητα η τιμή της συνάρτησης στο σημείο αυτό να είναι μικρότερη από την τρέχουσα βέλτιστη τιμή $f(x^+)$. Η μαθηματική της διατύπωση δίνεται από τη σχέση:

$$PI(x) = \Phi\left(\frac{\mu(x) - f(x^+) - \xi}{\sigma(x)}\right)$$

όπου $\mu(x)$ και $\sigma(x)$ είναι η μέση τιμή και η τυπική απόκλιση της πρόβλεψης του surrogate model, Φ αθροιστική συνάρτηση κατανομής της κανονικής κατανομής και ξ μία παράμετρος που ελέγχει το βαθμό εξερεύνησης.

Μια ακόμη ευρέως χρησιμοποιούμενη acquisition function είναι η *Upper Confidence Bound (UCB)*, η οποία επιλέγει το επόμενο σημείο λαμβάνοντας υπόψη τόσο την εκτιμώμενη τιμή της συνάρτησης όσο και την αβεβαιότητα της πρόβλεψης. Συγκεκριμένα, ευνοεί σημεία στα οποία η αναμενόμενη τιμή είναι χαμηλή ή/και η αβεβαιότητα υψηλή. Η μαθηματική της μορφή δίνεται από:

$$UCB(x) = \mu(x) + \kappa\sigma(x)$$

όπου $\mu(x)$ είναι η μέση τιμή της πρόβλεψης, $\sigma(x)$ η τυπική απόκλιση και κ μια παράμετρος που ρυθμίζει την ισορροπία μεταξύ εξερεύνησης και εκμετάλλευσης.

Πιο πρόσφατα, έχει προταθεί η *Weighted Expected Improvement (WEI)*^[17], η οποία εισάγει έναν δυναμικό συντελεστή στάθμισης που προσαρμόζει αυτόματα τη βαρύτητα της εξερεύνησης ανάλογα με το στάδιο της διαδικασίας.

2.2 Βήματα της Μπεϋζιανής Βελτιστοποίησης

Η Μπεϋζιανή Βελτιστοποίηση εκτελείται επαναληπτικά, ξεκινώντας από ένα αρχικό σύνολο σημείων που επιλέγονται τυχαία ή μέσω πλέγματος. Στη συνέχεια, εκπαιδεύεται η Gaussian Process πάνω στα δεδομένα αυτά, και η acquisition function βελτιστοποιείται προκειμένου να επιλεγεί το επόμενο σημείο αξιολόγησης. Η συνάρτηση στόχου υπολογίζεται στο νέο σημείο, το μοντέλο ενημερώνεται, και η διαδικασία επαναλαμβάνεται έως ότου επιτευχθεί κάποιο κριτήριο τερματισμού, όπως ο μέγιστος αριθμός επαναλήψεων ή η μηδενική βελτίωση στο αποτέλεσμα. Η παραπάνω επαναληπτική διαδικασία συνοψίζεται στον ψευδοκώδικα του Αλγορίθμου 1.

Παρόλο που οι Gaussian Processes αποτελούν την πιο διαδεδομένη επιλογή για συνεχείς και σχετικά χαμηλής διάστασης συναρτήσεις, υπάρχουν και άλλες model-based προσεγγίσεις που χρησιμοποιούνται ευρέως στη βιβλιογραφία. Χαρακτηριστικά παραδείγματα είναι οι βιβλιοθήκες *Optuna* και *HyperOpt*^[14], οι οποίες βασίζονται στον αλγόριθμο *Tree-structured Parzen Estimator (TPE)*^[8]. Στην περίπτωση αυτή, δεν ορίζεται ρητά μια acquisition function, αλλά εκτιμώνται δύο πιθανότητες: η μία για τις "καλές" λύσεις $p(x|y < y^*)$ και η άλλη για τις "λιγότερο καλές" $p(x|y > y^*)$. Το επόμενο σημείο επιλέγεται έτσι ώστε να μεγιστοποιεί τον λόγο αυτών των πιθανοτήτων, προτιμώντας τις περιοχές του χώρου που έχουν υψηλή πιθανότητα να αποδώσουν βελτίωση. Αυτή η

προσέγγιση είναι ιδιαίτερα αποτελεσματική όταν οι μεταβλητές είναι κατηγορικές ή όταν ο χώρος αναζήτησης είναι μη συνεχής, όπου οι Gaussian Processes δυσκολεύονται να εφαρμοστούν.

Μια άλλη γνωστή προσέγγιση είναι το πλαίσιο *SMAC (Sequential Model-based Algorithm Configuration)*^[16], το οποίο χρησιμοποιεί τον αλγόριθμο Random Forests ως surrogate model, προσφέροντας μια πιο ευέλικτη αλλά λιγότερο θεωρητικά τεκμηριωμένη εναλλακτική.

Συνολικά, η επιλογή του κατάλληλου μοντέλου εξαρτάται σε μεγάλο βαθμό από τη φύση του προβλήματος. Οι Gaussian Processes υπερέχουν σε προβλήματα με συνεχείς, ομαλές και χαμηλής διάστασης συναρτήσεις, ενώ οι μέθοδοι TPE και SMAC προσφέρουν καλύτερη απόδοση σε περιπτώσεις με κατηγορικές ή υψηλής διάστασης μεταβλητές.

Αλγόριθμος 1 Μπεϋζιανή Βελτιστοποίηση

1. **BO** (f, A, X, M, ϑ)
/* f : αντικειμενική συνάρτηση, A : acquisition function,
 X : χώρος αναζήτησης, ϑ : παράμετροι της A , M : surrogate model
*/
2. Έστω ότι έχουμε το αρχικό σύνολο $X = \{x_1, x_2, \dots, x_n\} \subset X$;
3. Υπολογίζουμε την $Y \leftarrow \{f(x_1), f(x_2), \dots, f(x_n)\}$;
4. Εκπαιδεύουμε ένα surrogate model M στα (X, Y) ;
5. **while** τα κριτήρια τερματισμού δεν ικανοποιούνται **do**
 $x' \leftarrow \arg \max_{x \in X} A(x; M, \vartheta)$;
 $y' \leftarrow f(x')$;
 $X \leftarrow X \cup \{x'\}$;
 $Y \leftarrow Y \cup \{y'\}$;
επανεκπαιδεύουμε το surrogate model M στα (X, Y)
end while
6. **return** $x_{best} = \arg \min_{x \in X} f(x)$

Παρακάτω παρατίθεται και ο ψευδοκώδικας της μεθόδου TPE (Tree-structured Parzen Estimator):

Αλγόριθμος 2 Tree-structured Parzen Estimator

1. TPE (f, X, D, γ, M, N)

/* f : αντικειμενική συνάρτηση, X : χώρος αναζήτησης,

D : Αρχικό σύνολο παρατηρήσεων $\{(x_i, y_i)\}_{i=1}^n$, γ : Παράμετρος ποσοστιαίου

ορίου $\gamma \in (0,1)$, M : αριθμός υποψήφιων δειγμάτων, N : μέγιστος αριθμός επαναλήψεων */

2. Αν D είναι κενό, δημιουργούμε ένα μικρό αρχικό design (π.χ. τυχαία δείγματα) και υπολογίζουμε την f για αυτά ώστε να έχουμε αρχικά δεδομένα.

3. Επανάληψη έως ότου εξαντληθεί ο μέγιστος αριθμός επαναλήψεων N :

- Ταξινομούμε τις υπάρχουσες τιμές $\{y_i\}$ και ορίζουμε το όριο γ^* ως το γ -quantile των $\{y_i\}$
- Ορίζουμε δύο σύνολα: το «καλό» $D_l = \{x_i : y_i < \gamma^*\}$ και το «κακό» $D_g = \{x_i : y_i \geq \gamma^*\}$
- Χρησιμοποιούμε Parzen estimator για να προσεγγίσουμε: $l(x) \approx p(x | y < \gamma^*)$, $g(x) \approx p(x | y \geq \gamma^*)$
- Για M επαναλήψεις: δειγματοληπτούμε $x^{(j)} \sim l(x)$ για $j = 1, \dots, M$
- $r^{(j)} = \frac{p(x^{(j)} | y < \gamma^*)}{p(x^{(j)} | y \geq \gamma^*)} \approx \frac{l(x^{(j)})}{g(x^{(j)})}$ για κάθε $x^{(j)}$
- $x_{next} = \arg \max_j r^j$
- $y_{next} = f(x_{next})$
- $D \leftarrow D \cup \{(x_{next}, y_{next})\}$

4. Επιστροφή του καλύτερου σημείου $x^* = \arg \min_{x \in D_x} f(x)$

3. Σχετική Έρευνα

Η Μπεϋζιανή Βελτιστοποίηση έχει μελετηθεί εκτενώς τις τελευταίες δύο δεκαετίες ως μία από τις πλέον αποδοτικές μεθόδους για τη βελτιστοποίηση ακριβών, black-box συναρτήσεων. Οι πρώτες θεμελιώδεις εργασίες στον χώρο, όπως αυτές των Jones et al.^[11] και Mockus^[13], έθεσαν τη βάση της model-based βελτιστοποίησης μέσω πιθανοτικών μοντέλων και acquisition functions, εισάγοντας την ιδέα της βελτιστοποίησης υπό αβεβαιότητα με ελάχιστο αριθμό αξιολογήσεων. Στη συνέχεια, η χρήση των Gaussian Processes κυριάρχησε, λόγω της ευελιξίας τους, της αναλυτικής μορφής της posterior κατανομής και της δυνατότητας ποσοτικοποίησης της αβεβαιότητας.

Σημαντική αναφορά αποτελεί η εργασία των Rasmussen^[12] και Williams^[12], οι οποίοι παρείχαν το θεωρητικό υπόβαθρο των Gaussian Processes για μηχανική μάθηση, καθιστώντας εφικτή τη συστηματική ενσωμάτωσή τους σε αλγορίθμους Bayesian Optimization. Σε επόμενες μελέτες, όπως αυτές των Brochu et al.^[9] και Shahriari et al.^[5], η Bayesian Optimization παρουσιάστηκε ως γενικό πλαίσιο για προβλήματα μηχανικής μάθησης, ρομποτικής και πειραματικού σχεδιασμού, με ιδιαίτερη έμφαση στην εξισορρόπηση εξερεύνησης και εκμετάλλευσης.

Ιδιαίτερη ώθηση στη χρήση της Bayesian Optimization για hyperparameter tuning δόθηκε από τις εργασίες των Snoek, Larochelle και Adams^[7], οι οποίοι έδειξαν τη σημαντική υπεροχή της έναντι του Grid Search και του Random Search σε προβλήματα εκπαίδευσης βαθιών νευρωνικών δικτύων. Τα αποτελέσματά τους κατέδειξαν ότι η χρήση Gaussian Processes σε συνδυασμό με acquisition function τύπου Expected Improvement μειώνει θεαματικά τον αριθμό των απαιτούμενων εκπαιδύσεων.

Παράλληλα με τις Gaussian Processes, αναπτύχθηκαν και εναλλακτικές προσεγγίσεις model-based. Κεντρική θέση σε αυτές κατέχει ο αλγόριθμος Tree-structured Parzen Estimator (TPE), που παρουσιάστηκε από τους Bergstra et al.^[8] στο πλαίσιο της HyperOpt. Ο TPE διαφοροποιείται από την κλασική Bayesian προσέγγιση, καθώς προσεγγίζει απευθείας τις κατανομές $p(x|y)$ αντί για το $p(y|x)$, επιτρέποντας αποτελεσματική βελτιστοποίηση σε χώρους με κατηγορικές, διακριτές ή εξαρτημένες μεταβλητές. Η αποτελεσματικότητά του επιβεβαιώθηκε σε πλήθος πειραματικών αξιολογήσεων, ειδικά σε προβλήματα υψηλής διάστασης.

Σημειώνεται πως η ραγδαία εξάπλωση της Μπεϋζιανής Βελτιστοποίησης και της αυτοματοποίησης της αναζήτησης υπερπαραμέτρων έχει οδηγήσει στην ανάπτυξη εξειδικευμένων βιβλιοθηκών AutoML, με πιο διαδεδομένες τις HyperOpt, Optuna^[18] και SMAC^[16]. Οι βιβλιοθήκες αυτές υλοποιούν διαφορετικές εκδοχές model-based βελτιστοποίησης, με στόχο την αποτελεσματική εξερεύνηση πολύπλοκων και υψηλής διάστασης χώρων υπερπαραμέτρων.

Η HyperOpt αποτελεί μία από τις πρώτες και πιο καθιερωμένες βιβλιοθήκες για Bayesian Optimization σε προβλήματα υπερπαραμέτρων και βασίζεται στον αλγόριθμο Tree-structured Parzen Estimator (TPE). Σε αντίθεση με τις Gaussian Processes, όπου εκτιμάται η κατανομή $p(y|x)$, η

HyperOpt μοντελοποιεί άμεσα τις κατανομές $p(x|y)$, διαχωρίζοντας τον χώρο των παρατηρήσεων σε «καλές» και «κακές» περιοχές ως προς μια κατώφλια τιμή y^* , η οποία ορίζεται ως σταθερό ποσοστιαίο σημείο των καλύτερων τιμών της αντικειμενικής συνάρτησης.

Στην πράξη, η HyperOpt εκτιμά τις πυκνότητες $p(x|y < y^*)$ και $p(x|y \geq y^*)$ με τη χρήση ανεξάρτητων Parzen estimators ανά διάσταση του χώρου. Για συνεχείς μεταβλητές χρησιμοποιούνται Gaussian kernels, ενώ για κατηγορικές μεταβλητές χρησιμοποιούνται κατανομές συχνοτήτων. Το επόμενο σημείο επιλέγεται από δειγματοληψία στην κατανομή των «καλών» σημείων και μεγιστοποίηση του λόγου $l(x)/g(x)$. Η HyperOpt υποστηρίζει ρητά tree-structured χώρους αναζήτησης, όπου η ενεργοποίηση κάποιων υπερπαραμέτρων εξαρτάται από την τιμή άλλων, γεγονός που την καθιστά ιδιαίτερα κατάλληλη για σύνθετους αλγορίθμους μηχανικής μάθησης.

Η Optuna^[18] αποτελεί πιο σύγχρονη βιβλιοθήκη, η οποία επίσης υποστηρίζει TPE ως βασικό sampler, αλλά με σημαντικές επεκτάσεις στο πλαίσιο της αποδοτικής υλοποίησης και της ευελιξίας. Σε επίπεδο αλγορίθμου, η Optuna χρησιμοποιεί παραλλαγή του TPE, στην οποία η εκτίμηση των πιθανοτήτων $p(x|y < y^*)$ και $p(x|y \geq y^*)$ γίνεται με πιο δυναμικό τρόπο, επιτρέποντας πολλαπλές ενημερώσεις, adaptive επιλογή του quantile γ και καλύτερη διαχείριση μικρών συνόλων δεδομένων στα αρχικά στάδια της αναζήτησης.

Ένα από τα βασικά πλεονεκτήματα της Optuna είναι ο μηχανισμός pruning, ο οποίος επιτρέπει τον πρόωρο τερματισμό μη αποδοτικών δοκιμών, μειώνοντας δραστικά το υπολογιστικό κόστος. Η Optuna επιπλέον υποστηρίζει τόσο TPE όσο και τυχαία δειγματοληψία, grid search και custom samplers, επιτρέποντας υβριδικές στρατηγικές. Σε αντίθεση με την HyperOpt, η Optuna διαχειρίζεται τον χώρο αναζήτησης μέσω δυναμικών Python αντικειμένων και όχι μέσω στατικών δηλώσεων χώρων, γεγονός που αυξάνει την πρακτική ευχρηστία της αλλά μειώνει ελαφρώς τον μαθηματικό φορμαλισμό της περιγραφής.

Το πλαίσιο SMAC (Sequential Model-based Algorithm Configuration)^[19] υιοθετεί διαφορετική προσέγγιση ως προς το surrogate model, χρησιμοποιώντας τυχαία δάση (Random Forests)^[20] αντί για Gaussian Processes ή Parzen estimators. Με τον τρόπο αυτό, το SMAC μπορεί να διαχειριστεί αποτελεσματικά μικτές μεταβλητές (συνεχείς, διακριτές και κατηγορικές), καθώς και πολύ υψηλής διάστασης χώρους αναζήτησης. Η acquisition function βασίζεται συνήθως στο Expected Improvement, ωστόσο η εκτίμηση της αβεβαιότητας προκύπτει από τη διασπορά των προβλέψεων των επιμέρους δέντρων του δάσους.

Σε αντίθεση με τις καθαρά Bayesian μεθόδους, το SMAC δεν στηρίζεται σε αυστηρό πιθανοτικό μοντέλο, γεγονός που του χαρίζει υψηλή πρακτική αποδοτικότητα, εις βάρος όμως της θεωρητικής ερμηνευσιμότητας. Παρόλα αυτά, έχει επιδείξει εξαιρετικές επιδόσεις σε προβλήματα algorithm configuration και AutoML, ιδίως σε περιπτώσεις με πολύπλοκους περιορισμούς παραμέτρων.

Από την ανάλυση των αλγορίθμων που αναφέρθηκαν προηγουμένως, παρατηρούμε ότι η HyperOpt και η Optuna βασίζονται στον ίδιο θεμελιώδη αλγόριθμο TPE, διαφοροποιούνται όμως ως προς τη στρατηγική δειγματοληψίας, την προσαρμοστικότητα του quantile γ και τη διαχείριση των conditional χώρων. Η Optuna υπερέχει σε επίπεδο μηχανισμών ελέγχου και επιτάχυνσης (pruning, parallel trials, dynamic sampling), ενώ η HyperOpt διατηρεί πιο καθαρό probabilistic σχεδιασμό. Το SMAC, από την άλλη πλευρά, διαφοροποιείται ριζικά ως προς το είδος του surrogate model, καθώς δεν χρησιμοποιεί Parzen estimators ούτε Gaussian Processes αλλά Random Forests. Ως αποτέλεσμα, προσαρμόζεται ευκολότερα σε προβλήματα πολύ υψηλής διάστασης, ενώ οι TPE-based μέθοδοι παρουσιάζουν πλεονέκτημα σε κατηγορικές και tree-structured μεταβλητές.

Σε σύγκριση με τις Gaussian Process-based μεθόδους, όλες οι παραπάνω προσεγγίσεις θυσιάζουν μέρος της θεωρητικής ακρίβειας για χάρη της υπολογιστικής αποδοτικότητας και της επεκτασιμότητας. Οι Gaussian Processes παραμένουν ανώτερες σε χαμηλής διάστασης, συνεχή και λεία προβλήματα, ενώ οι TPE και SMAC υπερτερούν σε μικτούς, υψηλής διάστασης και μη συνεχείς χώρους.

4. Μεθοδολογία

Στο παρόν κεφάλαιο παρουσιάζεται η αρχιτεκτονική της προτεινόμενης υλοποίησης Μπεϋζιανής Βελτιστοποίησης. Αρχικά περιγράφονται οι βασικοί μηχανισμοί του αλγορίθμου, οι χρησιμοποιούμενες συναρτήσεις απόκτησης και ο τρόπος πειραματικής αξιολόγησης. Ακολούθως, αναλύονται οι τεχνικές επέκτασης της μεθοδολογίας για τη διαχείριση κατηγορικών μεταβλητών.

4.1 Δομή και Συνιστώσες του Πλαισίου Μπεϋζιανής Βελτιστοποίησης

Για τις ανάγκες της παρούσας διπλωματικής αναπτύχθηκε μία πλήρως παραμετροποιήσιμη υλοποίηση Μπεϋζιανής Βελτιστοποίησης σε περιβάλλον Python, με κύριο στόχο τη συστηματική σύγκριση διαφορετικών στρατηγικών εξερεύνησης–εκμετάλλευσης. Ο σχεδιασμός της υλοποίησης εστιάζει στην υποστήριξη τόσο συνεχών όσο και κατηγορικών μεταβλητών, εξασφαλίζοντας παράλληλα τη στατιστική αξιοπιστία των αποτελεσμάτων μέσω επαναλαμβανόμενων πειραματικών εκτελέσεων.

Η διαδικασία βελτιστοποίησης βασίζεται στη χρήση Gaussian Process ως surrogate model, το οποίο εκπαιδεύεται επαναληπτικά στα μέχρι στιγμής παρατηρηθέντα δεδομένα. Σε κάθε επανάληψη, η Gaussian Process παρέχει εκτιμήσεις της μέσης τιμής και της διακύμανσης της αντικειμενικής συνάρτησης για κάθε υποψήφιο σημείο του χώρου αναζήτησης. Οι εκτιμήσεις αυτές αξιοποιούνται από τις acquisition functions για την επιλογή του επόμενου σημείου αξιολόγησης, επιτυγχάνοντας τη γνωστή ισορροπία μεταξύ εξερεύνησης και εκμετάλλευσης.

Στο πλαίσιο της υλοποίησης ενσωματώθηκαν και μελετήθηκαν τέσσερις βασικές acquisition functions: Expected Improvement, Probability of Improvement, Upper Confidence Bound και Thompson Sampling. Η Expected Improvement υπολογίζει την αναμενόμενη βελτίωση ως προς την καλύτερη μέχρι στιγμής παρατηρηθείσα τιμή και ευνοεί σημεία που συνδυάζουν υψηλή αναμενόμενη απόδοση και υψηλή αβεβαιότητα. Η Probability of Improvement εστιάζει στην πιθανότητα ένα νέο σημείο να υπερέχει του τρέχοντος βέλτιστου, χωρίς να λαμβάνει υπόψη το μέγεθος της βελτίωσης. Η Upper Confidence Bound επιλέγει σημεία μεγιστοποιώντας ένα άνω όριο εμπιστοσύνης που ορίζεται ως γραμμικός συνδυασμός της μέσης τιμής και της τυπικής απόκλισης της Gaussian Process, ελέγχοντας άμεσα το trade-off εξερεύνησης–εκμετάλλευσης μέσω παραμέτρου. Τέλος, το Thompson Sampling υλοποιήθηκε μέσω δειγματοληψίας συναρτήσεων από το posterior της Gaussian Process και επιλογής του σημείου που μεγιστοποιεί τη κάθε συνάρτηση.

Η υλοποίηση υποστηρίζει τόσο καθαρά θορυβώδεις όσο και μη θορυβώδεις αντικειμενικές συναρτήσεις, επιτρέποντας τη μελέτη της σύγκλισης των αλγορίθμων σε διαφορετικά είδη προβλημάτων. Παράλληλα, ενσωματώθηκε μηχανισμός καταγραφής της εξέλιξης της ελάχιστης τιμής της αντικειμενικής συνάρτησης σε κάθε επανάληψη, ώστε να είναι δυνατός ο υπολογισμός του δείκτη

convergence@iteration, ο οποίος αποτυπώνει την ταχύτητα σύγκλισης της εκάστοτε μεθόδου. Ο δείκτης σύγκλισης (*convergence@iteration*) ορίζεται ως η πρώτη επανάληψη στην οποία η τιμή της αντικειμενικής συνάρτησης προσεγγίζει επαρκώς το ολικό ελάχιστο. Πιο συγκεκριμένα, έστω f^* η βέλτιστη (ελάχιστη) τιμή της αντικειμενικής συνάρτησης και y_t η καλύτερη τιμή που έχει παρατηρηθεί μέχρι την επανάληψη t . Ο δείκτης *convergence@iteration* ορίζεται ως:

$$t^* = \min\{t \in \mathbb{N} : |y_t - f^*| \leq \varepsilon\}$$

όπου $\varepsilon > 0$ είναι ένα μικρό κατώφλι ανοχής που καθορίζει το πότε θεωρείται ότι έχει επιτευχθεί σύγκλιση.

Για τη στατιστικά αξιόπιστη αξιολόγηση της απόδοσης των μεθόδων, κάθε πείραμα επαναλήφθηκε δέκα φορές με διαφορετικές τυχαίες αρχικές τιμές. Για κάθε σύνολο επαναλήψεων καταγράφηκαν η μέση τιμή και η τυπική απόκλιση της βέλτιστης τιμής της αντικειμενικής συνάρτησης, καθώς και η μέση επανάληψη στην οποία επιτυγχανόταν σύγκλιση. Με τον τρόπο αυτό κατέστη δυνατή η ποσοτική σύγκριση τόσο μεταξύ των διαφορετικών *acquisition functions* όσο και μεταξύ των διαφορετικών στρατηγικών κωδικοποίησης κατηγορικών μεταβλητών.

Η υλοποίηση σχεδιάστηκε με τρόπο ώστε να είναι πλήρως συμβατή με τις πειραματικές συγκρίσεις με βιβλιοθήκες αυτόματης βελτιστοποίησης υπερπαραμέτρων, όπως οι Optuna και HyperOpt. Αυτό επέτρεψε τη σύγκριση των αποτελεσμάτων της Gaussian Process–based προσέγγισης με εκείνα του Tree-structured Parzen Estimator σε ελεγχόμενα πειραματικά σενάρια, χρησιμοποιώντας τα ίδια σύνολα παραμέτρων, ίδιους περιορισμούς χώρου αναζήτησης και ίδιες αντικειμενικές συναρτήσεις.

4.2 Οι περιπτώσεις των κατηγορικών μεταβλητών

Στην πράξη, πολλά προβλήματα βελτιστοποίησης υπερπαραμέτρων περιλαμβάνουν όχι μόνο συνεχείς αλλά και κατηγορικές μεταβλητές. Για την αντιμετώπιση αυτού του ζητήματος, η προτεινόμενη μεθοδολογία επεκτάθηκε ώστε να υποστηρίζει και κατηγορικά χαρακτηριστικά μέσω κατάλληλων τεχνικών κωδικοποίησης. Συγκεκριμένα, οι κατηγορικές μεταβλητές μετασχηματίζονται σε αριθμητική μορφή πριν εισαχθούν στο Gaussian Process surrogate model, επιτρέποντας την εφαρμογή της Μπεϋζιανής Βελτιστοποίησης σε μικτού τύπου χώρους αναζήτησης. Στο πλαίσιο αυτό εξετάστηκαν τρεις διαφορετικές τεχνικές κωδικοποίησης:

- *One-Hot Encoding*, όπου κάθε κατηγορική τιμή αναπαρίσταται ως δυαδικό διάνυσμα αυξάνοντας τη διάσταση του χώρου,
- *Hash Encoding*, όπου οι κατηγορικές τιμές προβάλλονται σε χαμηλότερης διάστασης χώρο μέσω συναρτήσεων κατακερματισμού,
- *Target Encoding*, όπου κάθε κατηγορία αντιστοιχίζεται σε μία τιμή που εξαρτάται από την παρατηρούμενη τιμή της αντικειμενικής συνάρτησης.

Η διαδικασία βελτιστοποίησης παραμένει η ίδια, με τη διαφορά ότι το διάνυσμα εισόδου προς το Gaussian Process περιλαμβάνει πλέον τόσο κανονικοποιημένες αριθμητικές μεταβλητές όσο και τις κωδικοποιημένες κατηγορικές μεταβλητές. Κατά την αξιολόγηση της αντικειμενικής συνάρτησης, πραγματοποιείται η αντίστροφη αντιστοίχιση (decoding) από το διανυσματικό χώρο στις αρχικές κατηγορικές τιμές. Ιδιαίτερη προσοχή δόθηκε στην περίπτωση του Target Encoding, όπου οι τιμές των κατηγοριών ενημερώνονται δυναμικά κατά τη διάρκεια της βελτιστοποίησης, αξιοποιώντας τις μέχρι στιγμής παρατηρήσεις. Με τον τρόπο αυτό, η πληροφορία της αντικειμενικής συνάρτησης ενσωματώνεται σταδιακά στη διαδικασία κωδικοποίησης. Η επέκταση αυτή καθιστά δυνατή την εφαρμογή της Μπεϋζιανής Βελτιστοποίησης σε προβλήματα μικτού τύπου (continuous–categorical), τα οποία εμφανίζονται συχνά σε εφαρμογές μηχανικής μάθησης, όπως η βελτιστοποίηση υπερπαραμέτρων αλγορίθμων. Η προσέγγιση αυτή εφαρμόστηκε πειραματικά σε πρόβλημα βελτιστοποίησης υπερπαραμέτρων αλγορίθμου συσταδοποίησης, όπως περιγράφεται στην επόμενη ενότητα.

5. Πειραματική Αξιολόγηση και Αποτελέσματα

Στο παρόν κεφάλαιο παρουσιάζονται, αναλύονται και συγκρίνονται τα αποτελέσματα της πειραματικής αξιολόγησης της Μπεϋζιανής Βελτιστοποίησης, τόσο για την custom υλοποίηση που αναπτύχθηκε στο πλαίσιο της παρούσας εργασίας, όσο και για τις καθιερωμένες βιομηχανικές βιβλιοθήκες Optuna και HyperOpt. Η αξιολόγηση επικεντρώθηκε στη συγκριτική μελέτη διαφορετικών συναρτήσεων απόκτησης (acquisition functions), διαφορετικών surrogate μοντέλων και στρατηγικών δειγματοληψίας, με κύριο στόχο τη διερεύνηση της ταχύτητας σύγκλισης, της ποιότητας των τελικών λύσεων και της σταθερότητας των αποτελεσμάτων. Η παρουσίαση των αποτελεσμάτων οργανώνεται συστηματικά ανά κατηγορία προβλήματος και ανά χρησιμοποιούμενη μέθοδο βελτιστοποίησης.

5.1 Πειραματικό Πλαίσιο και Μεθοδολογία Αξιολόγησης

Η πειραματική πορεία πραγματοποιήθηκε αρχικά σε ένα σύνολο καθιερωμένων συναρτήσεων αναφοράς (Ackley, Booth, Colville, Styblinski–Tang), οι οποίες καλύπτουν ένα ευρύ φάσμα δυσκολίας, μη κυρτότητας, πολυτροπικότητας και έντονης μη γραμμικότητας. Οι δοκιμές εκτελέστηκαν σε μονοδιάστατες και πολυδιάστατες περιπτώσεις (2–4 μεταβλητές), ενώ για τις δισδιάστατες πραγματοποιήθηκε και οπτικοποίηση του τοπίου της αντικειμενικής συνάρτησης καθώς και της πορείας σύγκλισης, για ποιοτική αξιολόγηση. Η αρχικοποίηση έγινε με τυχαία επιλογή λίγων αρχικών σημείων ώστε να εκπαιδευτεί το Gaussian Process surrogate model. Κάθε πείραμα εκτελέστηκε σε 200 επαναλήψεις για την παρατήρηση της σύγκλισης, χρησιμοποιώντας διαφορετικά kernels (Squared Exponential, Matern) και λαμβάνοντας υπόψη θορυβώδεις και μη θορυβώδεις εκδόσεις των συναρτήσεων.

Για κάθε πειραματική ρύθμιση και για κάθε acquisition function πραγματοποιήθηκαν 10 ανεξάρτητες εκτελέσεις με διαφορετικές τυχαίες αρχικές τιμές, ώστε να εξαχθούν στατιστικά αξιόπιστα συμπεράσματα. Από τις εκτελέσεις αυτές υπολογίστηκαν η μέση τιμή και η τυπική απόκλιση της καλύτερης παρατηρηθείσας τιμής της αντικειμενικής συνάρτησης, καθώς και ο μέσος αριθμός επαναλήψεων μέχρι τη σύγκλιση (converge@iter). Παράλληλα με την custom υλοποίηση, εκτελέστηκαν πειράματα με τις βιβλιοθήκες Optuna και HyperOpt (TPE), χρησιμοποιώντας κοινά όρια χώρου αναζήτησης, ίδιο αριθμό επαναλήψεων και ταυτόσημη αντικειμενική συνάρτηση για τη διασφάλιση μιας δίκαιης σύγκρισης. Εξετάστηκαν επίσης διαφορετικοί samplers της Optuna (TPE, Random, Grid) για τη διερεύνηση της επίδρασής τους στη σύγκλιση.

Στην περίπτωση κατηγορικών μεταβλητών, χρησιμοποιήθηκε ως παράδειγμα το πρόβλημα βελτιστοποίησης υπερπαραμέτρων του αλγορίθμου Agglomerative Clustering στο σύνολο δεδομένων Iris. Οι υπερπαραμέτροι περιλάμβαναν τον αριθμό ομάδων (n_clusters), το μέτρο απόστασης (metric)

και τη μέθοδο σύνδεσης (linkage). Οι κατηγορικές μεταβλητές ενσωματώθηκαν στο Gaussian Process μέσω one-hot, hashing ή target encoding, ενώ στις TPE-based υλοποιήσεις η διαχείριση των κατηγορικών μεταβλητών γινόταν εγγενώς μέσω δειγματοληψίας.

Ο Agglomerative Clustering, αποτελεί μία ιεραρχική μέθοδο συσταδοποίησης που ξεκινά με κάθε δείγμα ως ξεχωριστό cluster και συγχωνεύει σταδιακά τα πλησιέστερα clusters έως ότου επιτευχθεί ο επιθυμητός αριθμός clusters. Η απόσταση μεταξύ δύο clusters C_i και C_j υπολογίζεται με διάφορες στρατηγικές σύνδεσης (linkage), όπως η *single linkage* ($\min_{x \in C_i, y \in C_j} d(x, y)$) η *complete linkage* ($\max_{x \in C_i, y \in C_j} d(x, y)$) η *average linkage* ($\frac{1}{|C_i||C_j|} \sum_{x \in C_i} \sum_{y \in C_j} d(x, y)$) ακαι η *Ward linkage*, η οποία επιλέγει τα clusters που ελαχιστοποιούν την αύξηση της ενδο-ομαδικής διασποράς των σημείων:

$$\Delta E = \frac{|C_i||C_j|}{|C_i| + |C_j|} \|\mu_i - \mu_j\|^2$$

όπου μ_i και μ_j είναι οι μέσοι όροι των clusters C_i και C_j . Το αποτέλεσμα της διαδικασίας είναι ένα δέντρο ιεραρχίας (dendrogram), που απεικονίζει τον τρόπο συγχώνευσης των clusters σε κάθε βήμα. Η αντικειμενική συνάρτηση ήταν ο δείκτης silhouette score, που μετρά την ομοιότητα κάθε δείγματος με τη συστάδα του σε σχέση με τις υπόλοιπες, και χρησιμοποιήθηκε αρνητικός για ελαχιστοποίηση. Με βάση αυτές τις μεθόδους, αξιολογήθηκαν η ελάχιστη τιμή της αντικειμενικής συνάρτησης, η ταχύτητα σύγκλισης και η σταθερότητα των αποτελεσμάτων ως προς διαφορετικές αρχικοποιήσεις, επιτρέποντας μια ολοκληρωμένη συγκριτική ανάλυση μεταξύ των συναρτήσεων απόκτησης (acquisition functions), καθώς και μεταξύ των προσεγγίσεων που βασίζονται σε Γκαουσιανές Διαδικασίες (Gaussian Process-based) και σε TPE (TPE-based) σε προβλήματα μικτού τύπου.

5.2 Αξιολόγηση σε Μονοδιάστατες (1D) Συναρτήσεις

Τα αποτελέσματα και οι καμπύλες σύγκλισης που παρουσιάζονται στην παρούσα ενότητα προκύπτουν αποκλειστικά από την εκτέλεση του custom πλαισίου Μπεϋζιανής Βελτιστοποίησης που αναπτύχθηκε στο πλαίσιο της παρούσας εργασίας. Στο πρώτο κομμάτι των πειραμάτων, πραγματοποιήθηκε Μπεϋζιανή Βελτιστοποίηση σε 4 μονοδιάστατες συναρτήσεις, με τη βοήθεια 3 kernels: των RBF, Rational Quadratic και Matern kernels. Οι μονοδιάστατες συναρτήσεις που εισαγάγαμε στην υλοποίηση ήταν οι εξής:

- $f_1(x) = x^2 + 2x \sin(x), x \in [-10, 10]$
- $f_2(x) = x^5 - 4x^3 + 2x^2 + x - 3, x \in [-2, 2]$
- $f_3(x) = \sin(x) \cos(x) (x^5 - 4x^3 + 2x^2 + x - 3), x \in [-7, 7]$
- $f_4(x) = \sin(1/x), x \in [-0.1, 0.1]$

Ως acquisition function ορίστηκε η Expected Improvement, και τα αποτελέσματα της πειραματικής εκτέλεσης απεικονίζονται παρακάτω:

Πίνακας 1. 1D Συναρτήσεις – χωρίς θόρυβο

Συνάρτηση (Obj. Func)	Kernel	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
f1	RBF	± 1.307	-0.020	-0.001	6
f1	Matern	± 1.307	-0.020	-0.001	5
f1	RatQuadr	± 1.307	-0.020	-0.001	47
f2	RBF	+1.288	-0.405	-2.823	200
f2	Matern	+1.288	+1.287	-3.396	26
f2	RatQuadr	+1.288	+1.297	-3.396	2
f3	RBF	+5.713	+5.850	-2332.954	1
f3	Matern	+5.713	+5.678	-2449.918	1
f3	RatQuadr	+5.713	+5.955	-2050.577	1
f4	RBF	Inf	-0.071	-0.999	74
f4	Matern	Inf	-0.071	-0.999	1
f4	RatQuadr	Inf	-0.071	-0.999	55

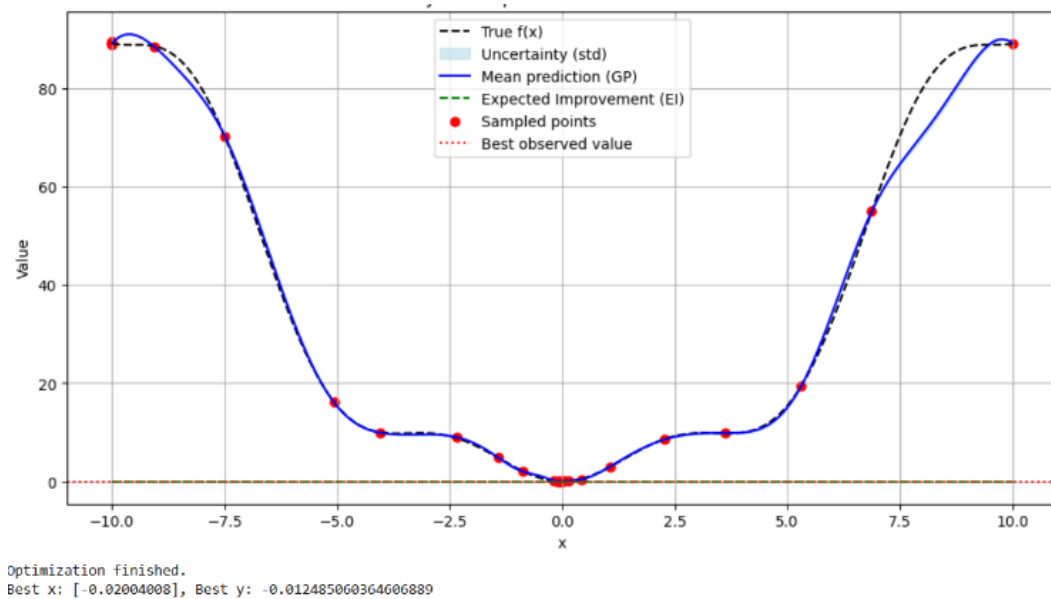
Με βάση τα αποτελέσματα του Πίνακα 1, παρατηρείται ότι η Μπεϋζιανή Βελτιστοποίηση παρουσιάζει ιδιαίτερα καλή συμπεριφορά στις μονοδιάστατες συναρτήσεις όταν η αντικειμενική συνάρτηση είναι απαλλαγμένη από θόρυβο. Για τη συνάρτηση f1, η οποία είναι ομαλή και σχετικά απλή, και οι τρεις πυρήνες συγκλίνουν στο ίδιο ελάχιστο, με τον Matern kernel να επιτυγχάνει τη σύγκλιση με τον μικρότερο αριθμό επαναλήψεων. Αντίθετα, ο Rational Quadratic kernel εμφανίζει σημαντικά μεγαλύτερο αριθμό επαναλήψεων, γεγονός που υποδηλώνει αυξημένη εξερευνητική συμπεριφορά σε σχέση με τους άλλους πυρήνες.

Στη συνάρτηση f2, η οποία παρουσιάζει πολυωνυμική συμπεριφορά με πολλαπλά τοπικά ακρότατα, παρατηρείται σαφής διαφοροποίηση μεταξύ των kernels. Ο Matern και ο Rational Quadratic kernel επιτυγχάνουν σύγκλιση σε πολύ μικρό αριθμό επαναλήψεων, ενώ ο RBF kernel αποτυγχάνει να συγκλίνει εντός του προκαθορισμένου αριθμού επαναλήψεων. Το αποτέλεσμα αυτό καταδεικνύει ότι ο RBF kernel, λόγω της ισχυρής υπόθεσης ομαλότητας που ενσωματώνει, δυσκολεύεται να προσαρμοστεί σε πιο σύνθετες συναρτήσεις.

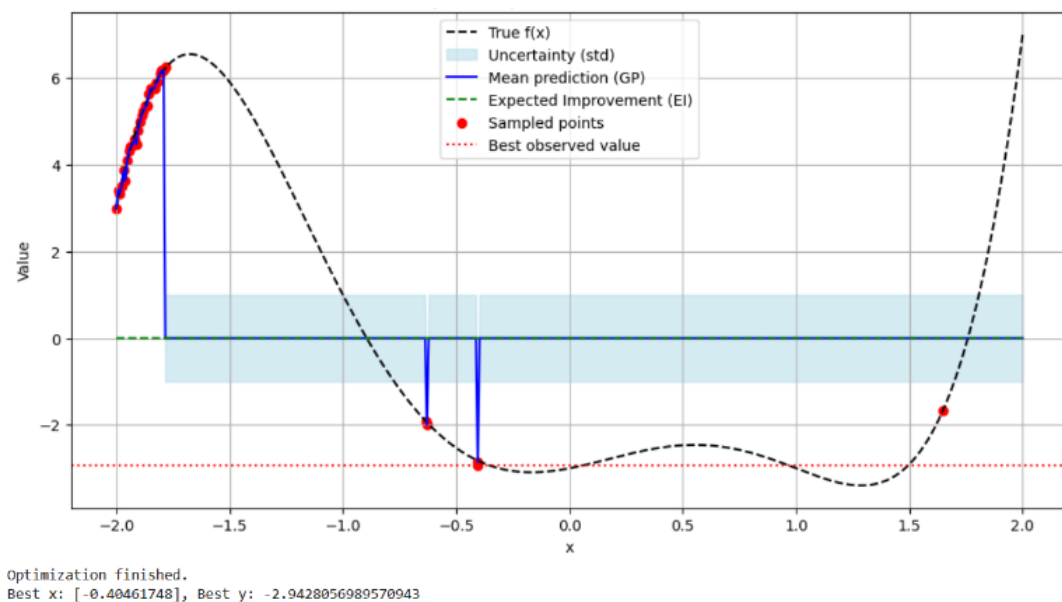
Για τη συνάρτηση f3, η οποία συνδυάζει ταλαντωτική συμπεριφορά με υψηλής τάξης πολυωνυμικό όρο, όλοι οι πυρήνες εντοπίζουν γρήγορα περιοχές πολύ χαμηλών τιμών της αντικειμενικής συνάρτησης. Η άμεση σύγκλιση σε μία επανάληψη υποδηλώνει ότι η αρχική δειγματοληψία ήταν ιδιαίτερα ευνοϊκή, ωστόσο η σημαντική διακύμανση στις τελικές τιμές μεταξύ των kernels δείχνει την ευαισθησία της διαδικασίας στην επιλογή του πυρήνα.

Τέλος, στη συνάρτηση f4, η οποία χαρακτηρίζεται από έντονη μη ομαλότητα κοντά στο μηδέν, παρατηρείται ότι ο Matern kernel υπερέχει σαφώς, επιτυγχάνοντας σύγκλιση σε μία μόλις

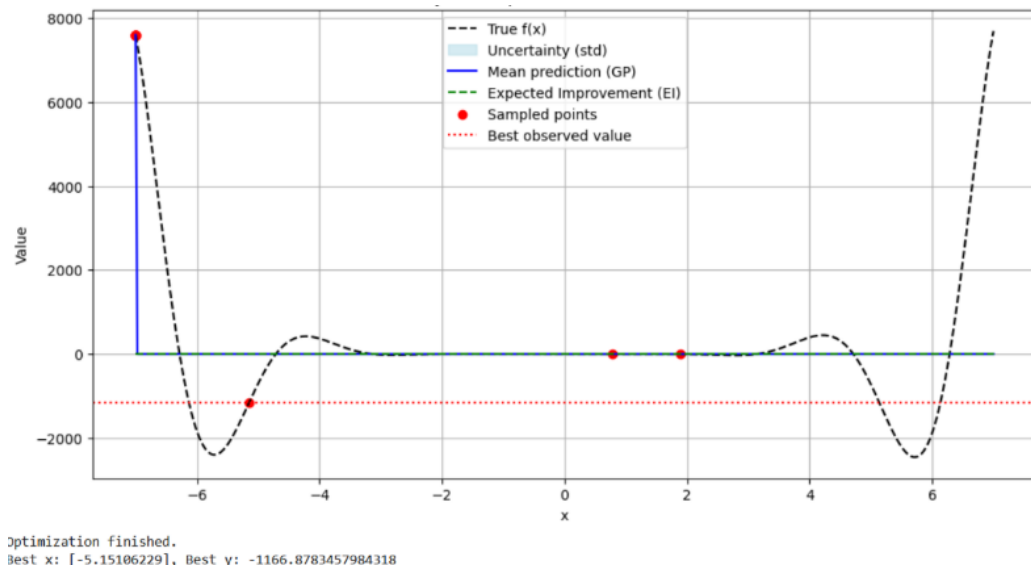
επανάληψη. Αντίθετα, οι RBF και Rational Quadratic kernels απαιτούν σημαντικά περισσότερες επαναλήψεις, γεγονός που επιβεβαιώνει ότι οι πυρήνες με χαλαρότερες υποθέσεις ομαλότητας είναι καταλληλότεροι για συναρτήσεις με έντονες ασυνέχειες.



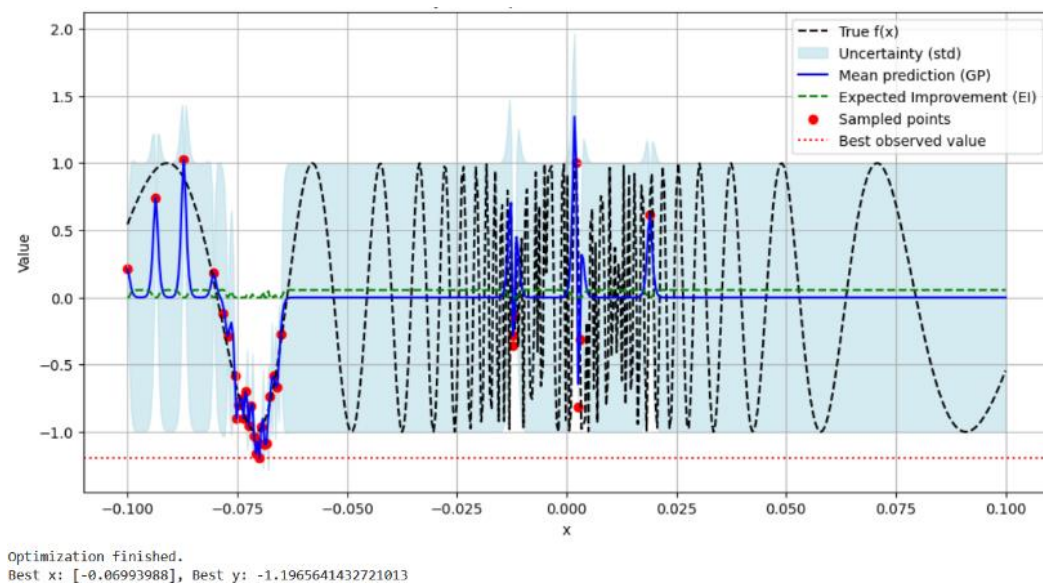
Εικόνα 1. $f_1(x)$ – Matern



Εικόνα 2. $f_2(x)$ – RBF



Εικόνα 3. $f_3(x)$ – Rational Quadratic



Εικόνα 4. $f_4(x)$ – Matern

Πίνακας 2. 1D Συναρτήσεις – Με θόρυβο

Συνάρτηση (Obj. Func)	Kernel	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
f1	RBF	± 1.307	+0.060	-0.019	5
f1	Matern	± 1.307	+0.060	-0.132	11
f1	RatQuadr	± 1.307	-0.020	-0.194	6
f2	RBF	+1.288	-0.405	-2.794	200
f2	Matern	+1.288	-0.230	-2.990	1
f2	RatQuadr	+1.288	+1.278	-3.542	107
f3	RBF	+5.713	-4.979	-644.465	5
f3	Matern	+5.713	-5.216	-1371.530	1
f3	RatQuadr	+5.713	+5.809	-2397.984	1
f4	RBF	Inf	-0.030	-1.119	175
f4	Matern	Inf	-0.048	-1.152	131
f4	RatQuadr	Inf	+0.093	-1.222	13

Η εισαγωγή θορύβου στην αντικειμενική συνάρτηση επηρεάζει αισθητά τη συμπεριφορά της Μπεϋζιανής Βελτιστοποίησης, όπως φαίνεται από τα αποτελέσματα του Πίνακα 2. Στη συνάρτηση f1, όλοι οι πυρήνες εξακολουθούν να συγκλίνουν σε περιοχές κοντά στο πραγματικό ελάχιστο, ωστόσο παρατηρείται αύξηση στη διασπορά των τελικών τιμών και μικρές μετατοπίσεις στο σημείο σύγκλισης. Ο Rational Quadratic kernel εμφανίζει τη μεγαλύτερη ανθεκτικότητα στον θόρυβο, επιτυγχάνοντας τη χαμηλότερη τιμή της αντικειμενικής συνάρτησης. Στη συνάρτηση f2, ο θόρυβος επιδεινώνει περαιτέρω τη συμπεριφορά του RBF kernel, ο οποίος και πάλι δεν συγκλίνει εντός του διαθέσιμου αριθμού επαναλήψεων. Αντίθετα, ο Matern kernel παρουσιάζει εξαιρετικά γρήγορη σύγκλιση, ενώ ο Rational Quadratic kernel επιτυγχάνει την καλύτερη τελική τιμή, αν και με μεγαλύτερο αριθμό επαναλήψεων. Το φαινόμενο αυτό αναδεικνύει τη διαφορά μεταξύ ταχύτητας σύγκλισης και ποιότητας τελικής λύσης.

Για τη συνάρτηση f3, παρατηρείται ότι ο θόρυβος οδηγεί τον RBF kernel σε περιοχές μακριά από το πραγματικό ολικό ελάχιστο, ενώ οι Matern και Rational Quadratic kernels διατηρούν υψηλή απόδοση, επιτυγχάνοντας πολύ χαμηλές τιμές της αντικειμενικής συνάρτησης σε ελάχιστες επαναλήψεις. Το αποτέλεσμα αυτό επιβεβαιώνει την αυξημένη ανθεκτικότητα των kernels αυτών σε στοχαστικά περιβάλλοντα. Στη συνάρτηση f4, η παρουσία θορύβου αυξάνει δραματικά τον αριθμό επαναλήψεων που απαιτούνται για τη σύγκλιση, ιδιαίτερα για τους RBF και Matern kernels. Ο Rational Quadratic kernel εμφανίζει σημαντικά καλύτερη συμπεριφορά, συγκλίνοντας ταχύτερα και επιτυγχάνοντας χαμηλότερη τελική τιμή, γεγονός που υποδηλώνει καλύτερη προσαρμογή σε μη ομαλές και θορυβώδεις συναρτήσεις.

5.3 Multidimensional Συναρτήσεις

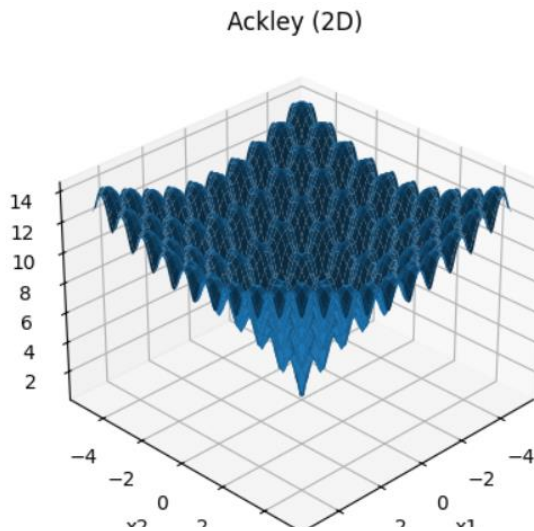
Όπως και στην προηγούμενη ενότητα, τα αποτελέσματα που απεικονίζονται παρακάτω προκύπτουν αποκλειστικά από τη δική μας custom υλοποίηση Μπεϋζιανής Βελτιστοποίησης. Σε αυτή τη φάση, το πλαίσιο αξιολογείται σε ένα σύνολο καθιερωμένων πολυδιάστατων μαθηματικών συναρτήσεων με σκοπό τη συγκριτική μελέτη των τεσσάρων συναρτήσεων απόκτησης (EI, PI, UCB, Thompson Sampling) ως προς την ακρίβεια και την ταχύτητα σύγκλισης.

5.3.1 Ackley

Η συνάρτηση *Ackley* σε διάσταση d ορίζεται ως:

$$f(x) = -20 \exp \left(-0.2 \sqrt{\frac{1}{d} \sum_{i=1}^d x_i^2} \right) - \exp \left(\frac{1}{d} \sum_{i=1}^d \cos(2\pi x_i) \right) + 20 + e$$

με $x \in [-5,5]^d$. Το ολικό ελάχιστο επιτυγχάνεται στο $x^* = (0, \dots, 0)$, όπου $f(x^*) = 0$.



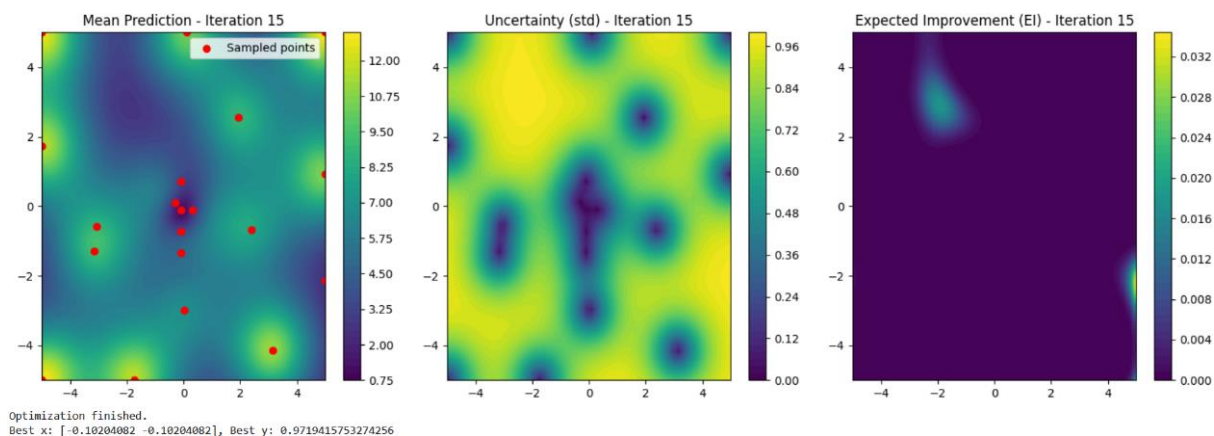
Εικόνα 5. Ackley (2D απεικόνιση)

Πίνακας 3. Ackley

Θόρυβος	Acquisition Functions	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
Όχι	EI	0	-0.102	+0.894	5
Ναι	EI	0	-0.102	+0.613	192
Όχι	PI	0	-0.102	+0.894	53
Ναι	PI	0	-0.102	+0.828	25
Όχι	UCB	0	-0.305	+2.383	0
Ναι	UCB	0	-1.820	+5.928	0
Όχι	Thompson Sampling	0	-1.168	+3.897	0
Ναι	Thompson Sampling	0	-1.735	+5.653	1

Στην περίπτωση της δισδιάστατης συνάρτησης Ackley, τα αποτελέσματα του Πίνακα 3 δείχνουν ότι η Μπεϋζιανή Βελτιστοποίηση επιτυγχάνει ταχεία σύγκλιση για τις περισσότερες acquisition functions, ιδιαίτερα όταν δεν υπάρχει θόρυβος στην αντικειμενική συνάρτηση. Η Expected Improvement και η Probability of Improvement συγκλίνουν σε χαμηλές τιμές της συνάρτησης με μικρό αριθμό επαναλήψεων, με την EI να εμφανίζει ελαφρώς ταχύτερη σύγκλιση. Η εισαγωγή θορύβου επηρεάζει σημαντικά τον αριθμό επαναλήψεων που απαιτούνται για τη σύγκλιση, ιδιαίτερα στην περίπτωση της EI, όπου παρατηρείται δραματική αύξηση του χρόνου σύγκλισης. Αντίθετα, η PI εμφανίζει μεγαλύτερη σταθερότητα υπό θόρυβο, αν και συγκλίνει σε ελαφρώς υψηλότερες τιμές της αντικειμενικής συνάρτησης.

Οι acquisition functions UCB και Thompson Sampling παρουσιάζουν έντονα εξερευνητική συμπεριφορά, οδηγώντας σε μεγαλύτερες τελικές τιμές της συνάρτησης και σε ταχεία, αλλά συχνά πρόωρη, σύγκλιση. Το αποτέλεσμα αυτό υποδεικνύει ότι σε χαμηλής διάστασης πολυτροπικές συναρτήσεις, η υπερβολική εξερεύνηση δεν είναι πάντοτε ευεργετική.



Εικόνα 6. Ackley (2D)

Πίνακας 4. Ackley (5D) – Χωρίς θόρυβο

Acquisition Functions	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
EI	0	-0.135 +0.059 +0.058 +0.070 +0.119	+0.792	182
PI	0	-0.070 +0.099 +0.177 +0.219 -0.326	+2.043	76
UCB	0	+2.846 -1.908 -3.071 -4.064 +0.623	+9.548	0
Thompson Sampling	0	+0.363 -0.909 +0.675 +0.981 -0.624	+4.479	0

Πίνακας 5. Ackley (5D) – Με θόρυβο

Acquisition Functions	x_default_ best	x_test	min(y)	επανάληψη (σύγκλισης)
EI	0	-0.056 -0.255 +0.029 +0.787 -0.055	2.374	183
PI	0	-0.015 -0.055 -0.334 -0.144 +0.002	1.656	112
UCB	0	+1.902 +2.762 +0.823 +3.009 +3.864	9.109	57
Thompson Sampling	0	-2.232 +0.659 +2.341 -1.876 -0.515	7.857	4

Στη 5D εκδοχή της συνάρτησης Ackley, η αύξηση της διάστασης καθιστά το πρόβλημα σημαντικά δυσκολότερο για τη Gaussian Process–based Μπεϋζιανή Βελτιστοποίηση. Παρατηρείται ότι η Expected Improvement απαιτεί μεγάλο αριθμό επαναλήψεων για να προσεγγίσει χαμηλές τιμές της αντικειμενικής συνάρτησης, τόσο με όσο και χωρίς θόρυβο.

Η Probability of Improvement παρουσιάζει ταχύτερη σύγκλιση σε σχέση με την EI, ωστόσο συγκλίνει σε υψηλότερες τελικές τιμές, γεγονός που υποδηλώνει τάση υπερεκμετάλλευσης τοπικών περιοχών. Οι acquisition functions UCB και Thompson Sampling εμφανίζουν αυξημένη εξερευνητική συμπεριφορά, οδηγώντας σε περιορισμένη απόδοση και μεγαλύτερη μεταβλητότητα, ιδιαίτερα στην παρουσία θορύβου. Τα αποτελέσματα αυτά αναδεικνύουν τη δυσκολία εφαρμογής των Gaussian Processes σε μεσαίας διάστασης προβλήματα, καθώς η αβεβαιότητα του μοντέλου αυξάνεται σημαντικά και η ποιότητα της εκτίμησης της συνάρτησης υποβαθμίζεται.

Πίνακας 6. Ackley (10D) – Χωρίς θόρυβο

Acquisition Functions	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
EI	0	-0.765, -0.235 -0.587, -1.958 +0.587, +0.964 +0.489, +1.028 -0.156, -0.754	4.887	155
PI	0	-0.169, +1.274 -0.137, -0.928 +0.975, +0.431 -1.352, +0.281 -0.374, +1.176	4.723	151
UCB	0	+1.099, +2.000 +0.197, +0.769 +0.912, -2.173 -3.115, -2.388 -0.170, -2.612	7.472	13
Thompson Sampling	0	+1.570, +1.494 +1.795, -0.583 -3.110, +0.990 -2.748, +0.627 +2.146, +0.556	7.877	16

Πίνακας 7. Ackley (10D) – Με θόρυβο

Acquisition Functions	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
EI	0	-0.638, +0.822 -0.929, +0.648 +0.464, -0.062 -0.326, -0.753 +0.714, -0.044	+3.928	154
PI	0	-1.021, +0.789 -2.076, +0.070 -1.175, -0.277 -1.466, -1.082 -0.209, -0.980	+5.079	196
UCB	0	+1.984, +0.742 +2.946, +2.746 -2.093, -4.035 -2.997, -1.337 -2.449, -0.757	+8.845	119
Thompson Sampling	0	+0.752, -3.043 +4.467, -0.228 +2.107, -0.036 +0.465, -2.116 -0.661, -1.358	+8.392	18

Στη 10D συνάρτηση Ackley, η απόδοση της Gaussian Process–based Μπεϋζιανής Βελτιστοποίησης μειώνεται αισθητά. Όλες οι acquisition functions συγκλίνουν σε σχετικά υψηλές τιμές της αντικειμενικής συνάρτησης, ενώ απαιτείται μεγάλος αριθμός επαναλήψεων ακόμη και στην απουσία θορύβου. Η εισαγωγή θορύβου επιδεινώνει περαιτέρω την κατάσταση, οδηγώντας σε αυξημένη μεταβλητότητα και χαμηλή σταθερότητα των αποτελεσμάτων. Οι UCB και Thompson Sampling εμφανίζουν ταχύτερη σύγκλιση, ωστόσο αυτή συνοδεύεται από σημαντική απώλεια ακρίβειας, γεγονός που υποδηλώνει πρόωρη σύγκλιση σε μη βέλτιστες περιοχές του χώρου. Τα αποτελέσματα αυτά καταδεικνύουν τους περιορισμούς των Gaussian Processes σε υψηλής διάστασης προβλήματα και αιτιολογούν τη χρήση εναλλακτικών model-based προσεγγίσεων, όπως ο Tree-structured Parzen Estimator, οι οποίες εξετάζονται στη συνέχεια της εργασίας.

Συνολικά, τα πειράματα στη συνάρτηση Ackley αναδεικνύουν τον καθοριστικό ρόλο της διάστασης και του θορύβου στη συμπεριφορά της Μπεϋζιανής Βελτιστοποίησης. Σε χαμηλές διαστάσεις, οι Gaussian Processes σε συνδυασμό με κατάλληλες acquisition functions επιτυγχάνουν ταχεία και

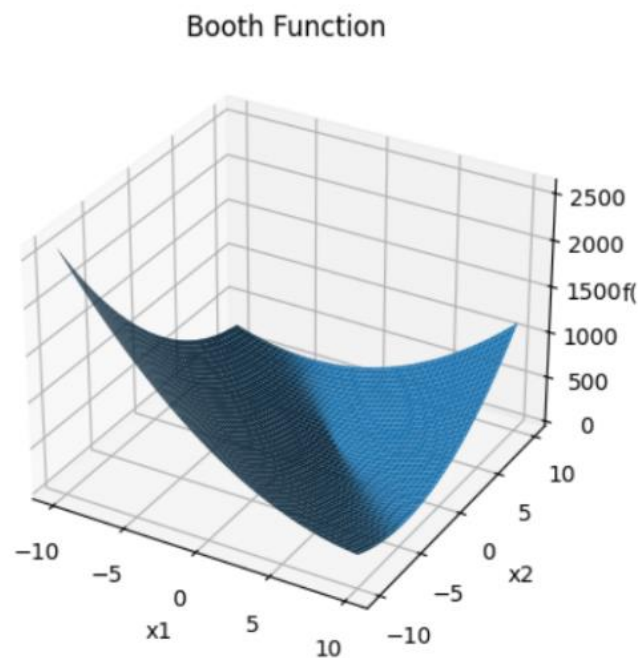
αξιόπιστη σύγκλιση. Ωστόσο, καθώς η διάσταση αυξάνεται, η απόδοση μειώνεται σημαντικά, καθιστώντας αναγκαία τη χρήση εναλλακτικών μεθόδων όπως ο TPE.

5.3.2 Booth

Η συνάρτηση *Booth*, δύο διαστάσεων, ορίζεται ως:

$$f(x_1, x_2) = (x_1 + 2x_2 - 7)^2 + (2x_1 + x_2 - 5)^2$$

με $(x_1, x_2) \in [-10, 10]^2$. Το ολικό ελάχιστο επιτυγχάνεται στο $(1, 3)$ όπου $f(x^*) = 0$.



Εικόνα 7. Booth (2D απεικόνιση)

Πίνακας 8. Booth

Θόρυβος	Acquisition Functions	x_default_ best	x_test	min(y)	επανάληψη (σύγκλισης)
Όχι	EI	+1 +3	+1.020 +3.061	+0.031	30
Ναι	EI	+1 +3	+0.204 +3.469	+1.075	193
Όχι	PI	+1 +3	+1.020 +3.061	+0.031	6
Ναι	PI	+1 +3	+1.020 +3.061	+0.033	19
Όχι	UCB	+1 +3	-2.200 +5.339	+18.680	0
Ναι	UCB	+1 +3	-1.862 +6.648	+23.948	0
Όχι	Thompson Sampling	+1 +3	+2.922 -0.171	+19.988	0
Ναι	Thompson Sampling	+1 +3	+0.138 +7.572	+76.659	0

Η συνάρτηση Booth αποτελεί ένα κλασικό παράδειγμα μονοτροπικής, κυρτής συνάρτησης δύο διαστάσεων, με μοναδικό ολικό ελάχιστο στο σημείο $(x_1, x_2) = (1, 3)$. Η απουσία τοπικών ελαχίστων την καθιστά κατάλληλη για την αξιολόγηση της ακρίβειας σύγκλισης και της σταθερότητας των acquisition functions, χωρίς την επιπλοκή πολυτροπικότητας.

Χωρίς θόρυβο

Στην περίπτωση απουσίας θορύβου, οι acquisition functions Expected Improvement (EI) και Probability of Improvement (PI) παρουσίασαν εξαιρετική συμπεριφορά. Και οι δύο μέθοδοι συγκλίνουν πολύ κοντά στο θεωρητικό ελάχιστο, με τελική τιμή της αντικειμενικής συνάρτησης $\min(y) \approx 0.03$, επιβεβαιώνοντας την ικανότητά τους να εκμεταλλεύονται αποτελεσματικά τη λεία και κυρτή επιφάνεια της Booth.

Αξιοσημείωτο είναι ότι η PI επιτυγχάνει σύγκλιση σημαντικά ταχύτερα (6 επαναλήψεις) σε σύγκριση με την EI (30 επαναλήψεις), γεγονός που υποδηλώνει ότι σε απλά και καλά ορισμένα προβλήματα η PI μπορεί να λειτουργήσει πιο επιθετικά προς την εκμετάλλευση του τρέχοντος βέλτιστου.

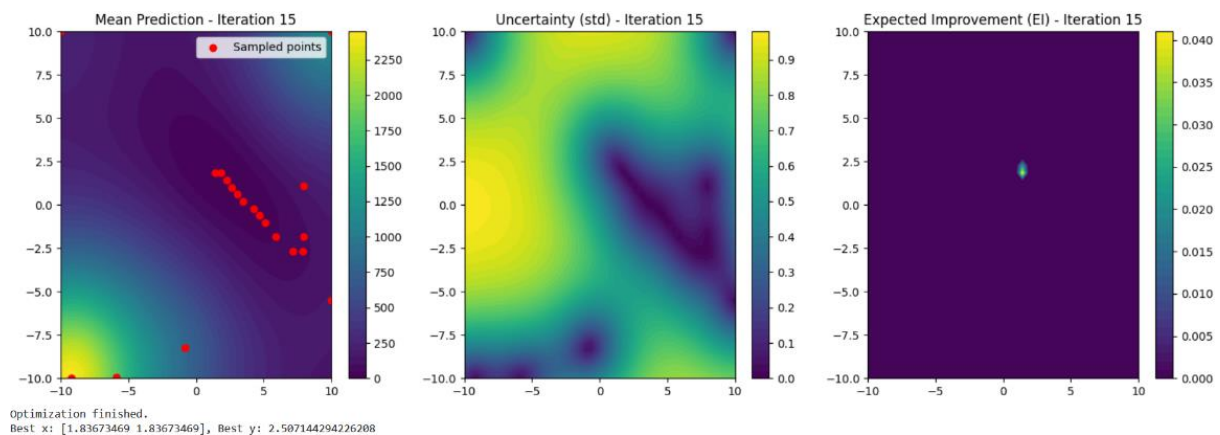
Αντίθετα, οι μέθοδοι UCB και Thompson Sampling αποτυγχάνουν να προσεγγίσουν το ολικό ελάχιστο, καταλήγοντας σε σημεία σημαντικά απομακρυσμένα από τη βέλτιστη λύση και με πολύ υψηλές τιμές της αντικειμενικής συνάρτησης. Το αποτέλεσμα αυτό είναι αναμενόμενο, καθώς και οι δύο μέθοδοι δίνουν έμφαση στην εξερεύνηση, η οποία σε μονοτροπικά προβλήματα οδηγεί σε περιττή δειγματοληψία μακριά από την περιοχή του ελαχίστου.

Με θόρυβο

Με την εισαγωγή θορύβου, παρατηρείται επιβράδυνση της σύγκλισης για όλες τις μεθόδους. Η EI εξακολουθεί να συγκλίνει κοντά στο βέλτιστο, αλλά απαιτεί σημαντικά περισσότερες επαναλήψεις (193), ενώ η τελική τιμή της αντικειμενικής συνάρτησης αυξάνεται, γεγονός που αντικατοπτρίζει την επίδραση του θορύβου στο surrogate model.

Η PI εμφανίζεται ιδιαίτερα ανθεκτική στον θόρυβο, διατηρώντας χαμηλό αριθμό επαναλήψεων (19) και πολύ μικρή απόκλιση από το ολικό ελάχιστο. Αυτό υποδηλώνει ότι η PI μπορεί να προσφέρει καλή ισορροπία μεταξύ ταχύτητας και ακρίβειας σε απλά, χαμηλής διάστασης προβλήματα, ακόμη και υπό στοχαστικές συνθήκες.

Αντιθέτως, οι UCB και Thompson Sampling παρουσιάζουν έντονη αστάθεια υπό θόρυβο, με ιδιαίτερα μεγάλες τιμές της αντικειμενικής συνάρτησης, επιβεβαιώνοντας ότι οι στρατηγικές έντονης εξερεύνησης δεν είναι κατάλληλες για κυρτά προβλήματα χαμηλής πολυπλοκότητας.



Εικόνα 8. Booth

5.3.3 Colville

Η συνάρτηση *Colville*, τεσσάρων διαστάσεων, ορίζεται ως:

$$f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 + (x_3 - 1)^2 + 90(x_3^2 - x_4)^2 + 10.1[(x_2 - 1)^2 + (x_4 - 1)^2] + 19.8(x_2 - 1)(x_4 - 1)$$

με $x \in [-10, 10]^4$. Το ολικό ελάχιστο επιτυγχάνεται στο $x^* = (1, 1, 1, 1)$, όπου $f(x^*) = 0$.

Πίνακας 9. Colville (χωρίς θόρυβο)

Acquisition Functions	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
EI	+1	+1.550	+15.874	111
	+1	+2.640		
	+1	-0.494		
	+1	+0.137		
PI	+1	+0.159	+25.526	179
	+1	+0.425		
	+1	-0.975		
	+1	+1.131		
UCB	+1	-2.990	+3280.388	0
	+1	+7.216		
	+1	+2.518		
	+1	+9.368		
Thompson Sampling	+1	-2.725	+10739.555	0
	+1	+1.877		
	+1	-4.175		
	+1	+8.691		

Πίνακας 10. Colville (Με θόρυβο)

Acquisition Functions	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
EI	+1	+0.561	+73.585	181
	+1	+0.299		
	+1	-2.069		
	+1	+3.853		
PI	+1	-1.389	+11.5869	119
	+1	+2.033		
	+1	-0.542		
	+1	+0.370		
UCB	+1	+2.018	+13739.946	0
	+1	-5.565		
	+1	-2.149		
	+1	-1.694		
Thompson Sampling	+1	-5.118	+45278.064	0
	+1	+9.013		
	+1	+3.060		
	+1	-3.803		

Χωρίς θόρυβο

Στην απουσία θορύβου, οι acquisition functions Expected Improvement (EI) και Probability of Improvement (PI) εμφανίζουν συγκλίνουσα συμπεριφορά, ωστόσο δεν καταφέρνουν να προσεγγίσουν το θεωρητικό ολικό ελάχιστο. Η EI επιτυγχάνει χαμηλότερη τελική τιμή της αντικειμενικής συνάρτησης σε σχέση με την PI (15.87 έναντι 25.53), υποδηλώνοντας καλύτερη ικανότητα εκμετάλλευσης της πληροφορίας του surrogate model σε αυτό το πιο απαιτητικό πρόβλημα.

Ωστόσο, και στις δύο περιπτώσεις απαιτείται σχετικά μεγάλος αριθμός επαναλήψεων για τη σύγκλιση (111 για EI και 179 για PI), γεγονός που καταδεικνύει τη δυσκολία της συνάρτησης ακόμη και σε συνθήκες πλήρους πληροφορίας.

Αντίθετα, οι UCB και Thompson Sampling αποτυγχάνουν πλήρως να συγκλίνουν προς τη βέλτιστη περιοχή, καταλήγοντας σε σημεία πολύ απομακρυσμένα από το ολικό ελάχιστο και με εξαιρετικά μεγάλες τιμές της αντικειμενικής συνάρτησης. Το αποτέλεσμα αυτό υποδεικνύει ότι οι στρατηγικές έντονης εξερεύνησης δεν είναι επαρκείς για την αποτελεσματική αντιμετώπιση της Colville, καθώς η εξερεύνηση οδηγεί σε δειγματοληψία περιοχών υψηλής καμπυλότητας χωρίς ουσιαστική βελτίωση.

Με θόρυβο

Με την εισαγωγή θορύβου, η δυσκολία του προβλήματος αυξάνεται περαιτέρω. Η Expected Improvement παρουσιάζει σημαντική υποβάθμιση της απόδοσής της, τόσο ως προς την ποιότητα της τελικής λύσης ($\min(y) = 73.59$) όσο και ως προς τον αριθμό επαναλήψεων μέχρι τη σύγκλιση (181). Το αποτέλεσμα αυτό αναδεικνύει την ευαισθησία της EI σε στοχαστικές διακυμάνσεις σε προβλήματα αυξημένης διάστασης.

Αντίθετα, η Probability of Improvement εμφανίζεται πιο ανθεκτική στον θόρυβο, επιτυγχάνοντας αισθητά χαμηλότερη τελική τιμή της αντικειμενικής συνάρτησης ($\min(y) = 11.59$) και μικρότερο αριθμό επαναλήψεων (119). Η συμπεριφορά αυτή υποδηλώνει ότι η PI μπορεί να λειτουργήσει πιο σταθερά σε στοχαστικά περιβάλλοντα, έστω και εις βάρος της ακρίβειας σύγκλισης σε προβλήματα χωρίς θόρυβο.

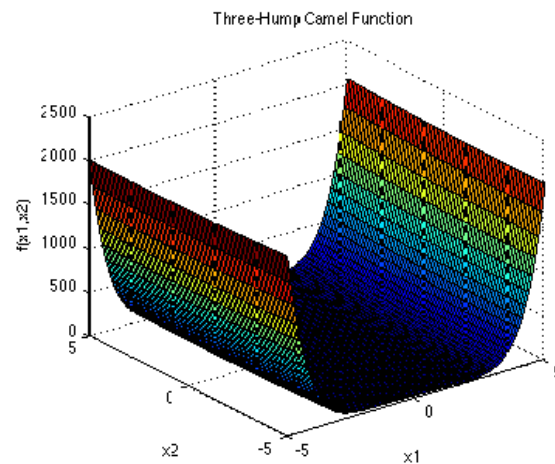
Όπως και στην περίπτωση χωρίς θόρυβο, οι UCB και Thompson Sampling παρουσιάζουν έντονη αστάθεια, με εξαιρετικά μεγάλες τιμές της αντικειμενικής συνάρτησης, επιβεβαιώνοντας ότι η υπερβολική εξερεύνηση σε συνδυασμό με θόρυβο οδηγεί σε κατάρρευση της απόδοσης του surrogate model.

5.3.4 Three-hump Camel

Η συνάρτηση *Three-hump Camel* ορίζεται ως:

$$f(x) = 2x_1^2 - 1.05x_1^4 + \frac{x_1^6}{6} + x_1x_2 + x_2^2$$

με $(x_1, x_2) \in [-5, 5]^2$. Το ολικό ελάχιστο επιτυγχάνεται στο $x^* = (0, 0)$ όπου $f(x^*) = 0$.



Εικόνα 9. *Three-hump Camel* (2D απεικόνιση)

Πίνακας 11. *Three-hump Camel*

Θόρυβος	Acquisition Functions	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
Όχι	EI	0	+1.734	+0.021	62
		0	-0.510		
Ναι	EI	0	+1.734	-0.167	170
		0	-0.510		
Όχι	PI	0	+1.734	+0.021	23
		0	-0.511		
Ναι	PI	0	+1.734	-0.072	28
		0	-0.511		
Όχι	UCB	0	+1.734	+2.503	0
		0	-0.512		
Ναι	UCB	0	+1.734	+29.311	0
		0	-0.513		
Όχι	Thompson Sampling	0	+1.734	+10.233	0
		0	-0.514		
Ναι	Thompson Sampling	0	+1.734	+6.148	0
		0	-0.515		

Χωρίς θόρυβο

Στην απουσία θορύβου, οι acquisition functions Expected Improvement (EI) και Probability of Improvement (PI) εμφανίζουν συγκλίνουσα συμπεριφορά, προσεγγίζοντας σχεδόν ακριβώς το θεωρητικό ολικό ελάχιστο. Η EI επιτυγχάνει χαμηλότερη τελική τιμή της αντικειμενικής συνάρτησης σε σχέση με την PI (+0.021 έναντι -0.167), υποδηλώνοντας καλύτερη ικανότητα εκμετάλλευσης της πληροφορίας του surrogate model σε αυτό το πιο απαιτητικό πρόβλημα.

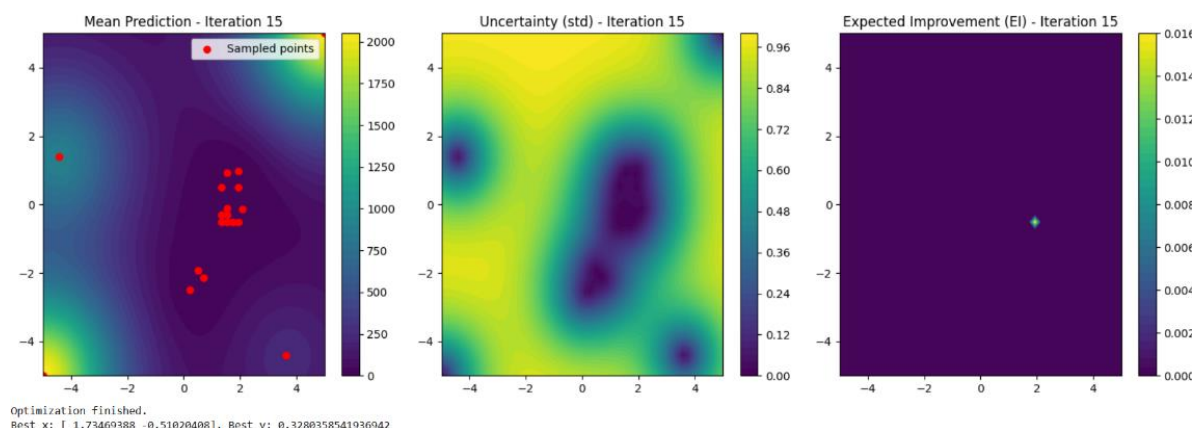
Ωστόσο, στην περίπτωση της PI απαιτείται σχετικά μεγάλος αριθμός επαναλήψεων για τη σύγκλιση (170 έναντι 62 για EI), γεγονός που προσθέτει ακόμα ένα πλεονέκτημα υπέρ της EI.

Αντίθετα, οι UCB και Thompson Sampling αποτυγχάνουν να συγκλίνουν προς τη βέλτιστη περιοχή, καταλήγοντας σε σημεία αρκετά απομακρυσμένα από το ολικό ελάχιστο και με εξαιρετικά μεγάλες τιμές της αντικειμενικής συνάρτησης. Το αποτέλεσμα αυτό υποδεικνύει ότι οι στρατηγικές έντονης εξερεύνησης δεν είναι επαρκείς για την αποτελεσματική αντιμετώπιση της Colville, καθώς η εξερεύνηση οδηγεί σε δειγματοληψία περιοχών υψηλής καμπυλότητας χωρίς ουσιαστική βελτίωση.

Με θόρυβο

Με την εισαγωγή θορύβου, η δυσκολία του προβλήματος αυξάνεται περαιτέρω. Ωστόσο, δεν επηρεάζεται η απόδοση της Expected Improvement, τόσο ως προς την ποιότητα της τελικής λύσης όσο και ως προς τον αριθμό επαναλήψεων μέχρι τη σύγκλιση (23). Το αποτέλεσμα αυτό αναδεικνύει την σταθερότητα της EI σε στοχαστικές διακυμάνσεις σε διδιάστατα προβλήματα.

Όπως και στην περίπτωση χωρίς θόρυβο, οι UCB και Thompson Sampling παρουσιάζουν έντονη αστάθεια, με εξαιρετικά μεγάλες τιμές της αντικειμενικής συνάρτησης, επιβεβαιώνοντας ότι η υπερβολική εξερεύνηση σε συνδυασμό με θόρυβο οδηγεί σε κατάρρευση της απόδοσης του surrogate model.



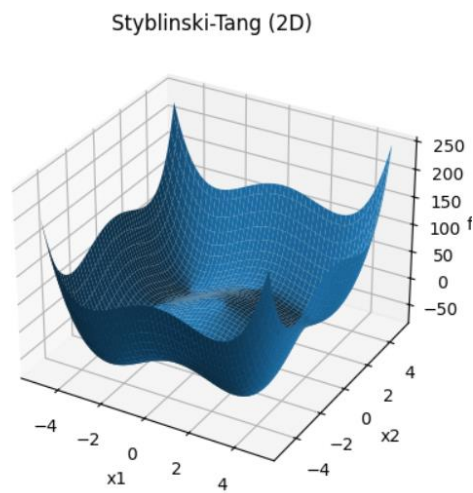
Εικόνα 10. Three-hump Camel

5.3.5 Styblinski-Tang

Η συνάρτηση *Styblinski-Tang* ορίζεται ως:

$$f(x) = \frac{1}{2} \sum_{i=1}^d (x_i^4 - 16x_i^2 + 5x_i)$$

με $x \in [-5, 5]^d$. Το ολικό ελάχιστο επιτυγχάνεται στο $x^* \approx (-2.9035 \dots, -2.9035)$,
με $f(x^*) \approx -39.166 \cdot d$.



Εικόνα 11. *Styblinski-Tang* (2D απεικόνιση)

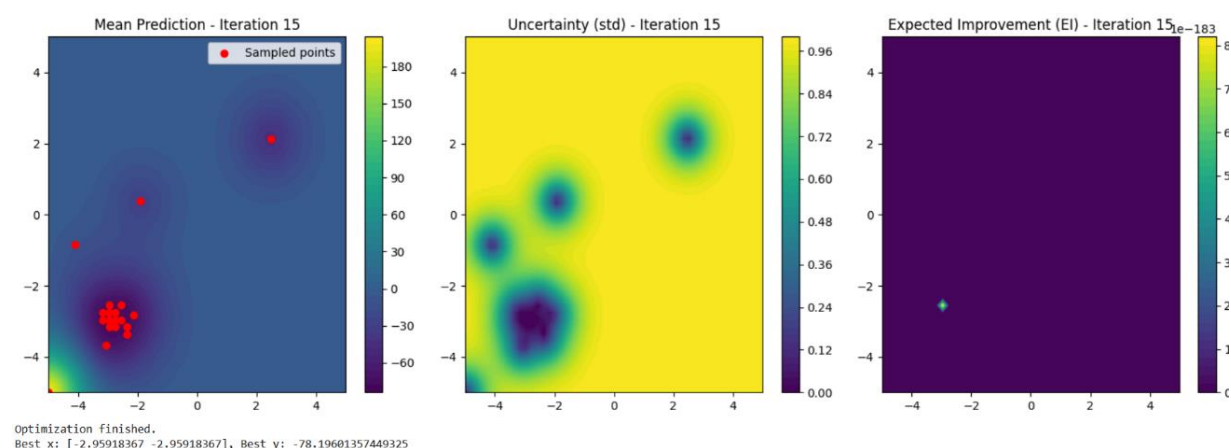
Πίνακας 12. *Styblinski-Tang* (2D)

Θόρυβος	Acquisition Functions	x _default_ best	x _test	min(y)	επανάληψη (σύγκλισης)
Όχι	EI	-2.903 -2.903	+3.458 -2.274	-49.209	1
Ναι	EI	-2.903 -2.903	+2.755 -2.959	-64.343	85
Όχι	PI	-2.903 -2.903	-2.959 -2.959	-78.223	8
Ναι	PI	-2.903 -2.903	-2.959 +2.755	-64.282	32
Όχι	UCB	-2.903 -2.903	-3.375 -2.721	-73.324	0
Ναι	UCB	-2.903 -2.903	-3.728 -3.218	-61.104	0

Στη δισδιάστατη εκδοχή της συνάρτησης, τόσο η Expected Improvement (EI) όσο και η Probability of Improvement (PI) επιτυγχάνουν πολύ γρήγορη και ακριβή σύγκλιση προς το ολικό ελάχιστο στην περίπτωση χωρίς θόρυβο. Ιδιαίτερα η EI εντοπίζει βέλτιστη λύση ήδη από την πρώτη επανάληψη, γεγονός που υποδηλώνει την ευνοϊκή γεωμετρία του προβλήματος σε χαμηλή διάσταση.

Με την εισαγωγή θορύβου, παρατηρείται αύξηση στον αριθμό επαναλήψεων, κυρίως για την EI, ωστόσο και οι δύο μέθοδοι συνεχίζουν να συγκλίνουν σε περιοχές πολύ κοντά στο θεωρητικό βέλτιστο. Η PI εμφανίζει μεγαλύτερη σταθερότητα ως προς τον αριθμό επαναλήψεων, επιβεβαιώνοντας για ακόμη μία φορά την αποτελεσματικότητά της σε προβλήματα χαμηλής διάστασης.

Οι μέθοδοι UCB παρουσιάζουν έντονη εξερευνητική συμπεριφορά, οδηγώντας σε λύσεις που αν και αρκετά καλές, δεν επιτυγχάνουν το βέλτιστο επίπεδο ακρίβειας, ενώ συγκλίνουν άμεσα χωρίς ουσιαστική εκμετάλλευση της πληροφορίας του surrogate model.



Εικόνα 12. Styblinski-Tang (2D)

Πίνακας 13. Styblinski-Tang (5D) - Χωρίς θόρυβο

Acquisition Functions	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
EI	-2.903	-2.980	-175.999	182
	-2.903	-3.307		
	-2.903	-1.792		
	-2.903	-3.102		
	-2.903	-2.578		
PI	-2.903	+2.837	-164.190	190
	-2.903	+2.974		
	-2.903	-2.555		
	-2.903	-2.788		
	-2.903	-3.040		
UCB	-2.903	-2.608	-143.020	27
	-2.903	-2.389		
	-2.903	-2.604		
	-2.903	+1.969		
	-2.903	+3.521		
Thompson Sampling	-2.903	-2.055	-126.112	11
	-2.903	+2.583		
	-2.903	+2.786		
	-2.903	-1.039		
	-2.903	-3.317		

Πίνακας 14. Styblinski-Tang (5D) – Με θόρυβο

Θόρυβος	Acquisition Functions	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
Ναι	EI	-2.903	-2.350	-166.320	141
		-2.903	-2.206		
		-2.903	+3.063		
		-2.903	-3.186		
		-2.903	-2.595		
Ναι	PI	-2.903	-2.112	-164.703	178
		-2.903	-2.177		
		-2.903	-2.696		
		-2.903	+2.554		
		-2.903	-3.069		
Ναι	UCB	-2.903	+2.687	-128.382	10
		-2.903	+3.351		
		-2.903	-2.828		
		-2.903	-1.649		
		-2.903	+3.031		
Ναι	Thompson Sampling	-2.903	+1.548	-94.258	3
		-2.903	-2.844		
		-2.903	-3.861		
		-2.903	+3.136		
		-2.903	+3.788		

Στη μεσαία διάσταση εκδοχή (5D), η πολυπλοκότητα του προβλήματος αυξάνεται αισθητά. Οι EI και PI συνεχίζουν να επιτυγχάνουν τις χαμηλότερες τιμές της αντικειμενικής συνάρτησης, ωστόσο απαιτούν σημαντικά μεγαλύτερο αριθμό επαναλήψεων μέχρι τη σύγκλιση, γεγονός που αντανακλά την αυξημένη δυσκολία του χώρου αναζήτησης.

Η EI παρουσιάζει ελαφρώς καλύτερη τελική τιμή χωρίς θόρυβο, ενώ η PI εμφανίζεται πιο σταθερή υπό την παρουσία θορύβου, με συγκρίσιμες τελικές τιμές αλλά ελαφρώς μεγαλύτερο αριθμό επαναλήψεων. Το αποτέλεσμα αυτό συνάδει με τη γενικότερη συμπεριφορά των δύο acquisition functions, όπου η EI ευνοεί την εκμετάλλευση, ενώ η PI εμφανίζει πιο ισορροπημένη συμπεριφορά. Αντίθετα, οι UCB και Thompson Sampling συγκλίνουν ταχύτερα, ωστόσο καταλήγουν σε σαφώς χειρότερες λύσεις, γεγονός που υποδεικνύει ότι η έντονη εξερεύνηση δεν είναι αποτελεσματική σε τόσο πολυτροπικά προβλήματα μέσης διάστασης.

Πίνακας 15. Styblinski-Tang (10D) - Χωρίς θόρυβο

Acquisition Functions	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
EI	-2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903	-2.921 -2.712 -0.993 +1.602 -3.505 +2.400 -2.533 +0.402 -0.869 -2.716	-239.790	178
PI	-2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903	-2.588 +1.325 -2.007 -2.567 -3.395 -3.464 +3.147 -1.327 -3.670 +2.873	-269.469	65
UCB	-2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903	+1.570 +2.126 -3.568 -0.803 -2.097 +3.704 -3.790 -2.567 -1.699 +2.184	-210.336	26
Thompson Sampling	-2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903 -2.903	+3.516 -3.107 1.627 -3.315 -3.089 +1.912 -3.942 -4.239 -2.884 +1.316	-230.014	108

Πίνακας 16. Styblinski-Tang (10D) - Με θόρυβο

Acquisition Functions	x_default_best	x_test	min(y)	επανάληψη (σύγκλισης)
EI	-2.903	-3.799	-276.301	133
	-2.903	-2.655		
	-2.903	+3.506		
	-2.903	-2.771		
	-2.903	+1.919		
	-2.903	-2.516		
	-2.903	-2.603		
	-2.903	+3.193		
	-2.903	-3.606		
	-2.903	-1.623		
PI	-2.903	-2.156	-284.989	16
	-2.903	-1.462		
	-2.903	-2.311		
	-2.903	-1.424		
	-2.903	-3.253		
	-2.903	-2.839		
	-2.903	-3.126		
	-2.903	-2.136		
	-2.903	+2.157		
	-2.903	+1.804		
UCB	-2.903	-3.910	-242.175	27
	-2.903	+3.618		
	-2.903	-0.680		
	-2.903	-1.206		
	-2.903	+3.395		
	-2.903	-3.143		
	-2.903	-2.464		
	-2.903	-2.553		
	-2.903	-2.126		
	-2.903	-2.624		
Thompson Sampling	-2.903	+1.686	-234.050	102
	-2.903	-2.558		
	-2.903	-0.875		
	-2.903	-2.564		
	-2.903	-1.485		
	-2.903	+3.168		
	-2.903	-3.065		
	-2.903	-0.720		
	-2.903	-2.119		
	-2.903	+3.275		

Η 10D εκδοχή της συνάρτησης αποτελεί το πιο απαιτητικό σενάριο πειραματικής αξιολόγησης. Σε αυτή την περίπτωση, η PI αναδεικνύεται ως η πιο αποτελεσματική acquisition function, επιτυγχάνοντας τις χαμηλότερες τιμές της αντικειμενικής συνάρτησης τόσο με όσο και χωρίς θόρυβο, και μάλιστα σε αισθητά λιγότερες επαναλήψεις σε σχέση με την EI.

Η EI, αν και καταφέρνει να προσεγγίσει το ολικό ελάχιστο, απαιτεί σημαντικά περισσότερες επαναλήψεις, γεγονός που υποδηλώνει μειωμένη αποδοτικότητα καθώς αυξάνεται η διάσταση του προβλήματος. Το αποτέλεσμα αυτό είναι αναμενόμενο, καθώς η EI βασίζεται σε πιο τοπικές εκτιμήσεις του surrogate model.

Οι UCB και Thompson Sampling παρουσιάζουν εκ νέου ταχύτερη σύγκλιση, ωστόσο οι τελικές λύσεις απέχουν αισθητά από το ολικό ελάχιστο, επιβεβαιώνοντας ότι σε προβλήματα υψηλής διάστασης και έντονης πολυτροπικότητας, οι στρατηγικές αυτές δυσκολεύονται να ισορροπήσουν αποτελεσματικά μεταξύ εξερεύνησης και εκμετάλλευσης.

5.4 Optuna & HyperOpt vs Multidimensional Συναρτήσεις

Στην παρούσα ενότητα πραγματοποιείται η συγκριτική αξιολόγηση των εξατομικευμένων υλοποιήσεων Μπεϋζιανής Βελτιστοποίησης έναντι των καθιερωμένων βιβλιοθηκών αυτοματοποιημένης μηχανικής μάθησης Optuna και HyperOpt. Η ανάλυση εστιάζει στην απόδοση των αλγορίθμων σε περιβάλλον πολυδιάστατων συναρτήσεων αναφοράς (benchmark functions) υπό συνθήκες απουσίας στοχαστικού θορύβου, εξετάζοντας την ακρίβεια της τελικής λύσης, την ταχύτητα σύγκλισης και το υπολογιστικό κόστος.

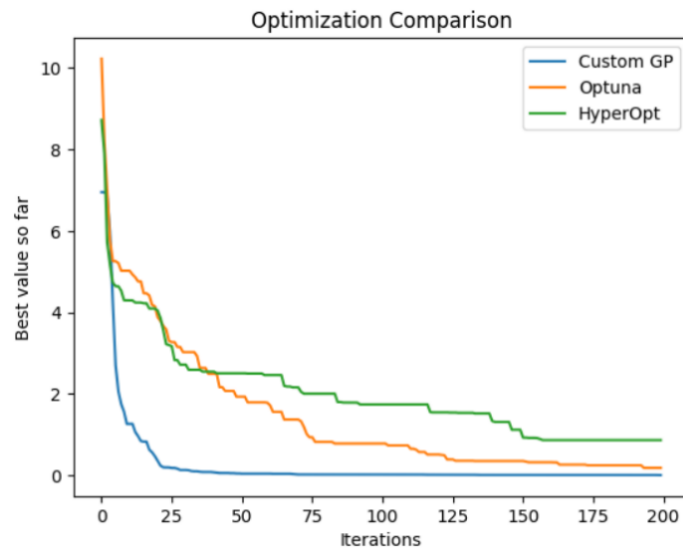
5.4.1 Ackley

Πίνακας 17. Σύγκριση Ackley με Optuna-HyperOpt (χωρίς θόρυβο)

Συνάρτηση	Acquisition Functions	min(y)	Επανάληψη (σύγκλισης)	Μέσος χρόνος εκτέλεσης (s)
Ackley (2D)	PI	1.117	144	27.41
Ackley (2D)	EI	0.011	96	30.68
Ackley (2D)	UCB	8.250	1	27.55
Ackley (2D)	Optuna	0.189	138	1.01
Ackley (2D)	HyperOpt	0.870	102	0.92
Ackley (5D)	EI	5.512	186	31.65
Ackley (5D)	PI	2.068	105	31.87
Ackley (5D)	UCB	8.453	1	30.04
Ackley (5D)	Optuna	2.051	167	2.35
Ackley (5D)	HyperOpt	3.841	147	1.99
Ackley (10D)	EI	4.363	178	19.18
Ackley (10D)	PI	4.127	118	33.86
Ackley (10D)	UCB	8.453	5.2	19.21
Ackley (10D)	Optuna	4.493	183	5.15
Ackley (10D)	HyperOpt	5.241	159	3.86

Ackley (2D)

Στη δισδιάστατη εκδοχή της συνάρτησης, η custom υλοποίηση με Expected Improvement (EI) επιτυγχάνει την καλύτερη τελική τιμή, προσεγγίζοντας πολύ κοντά το ολικό ελάχιστο. Η PI παρουσιάζει σαφώς χειρότερη απόδοση, τόσο ως προς την τελική τιμή αλλά παρόμοια ως προς τον αριθμό επαναλήψεων. Η UCB συγκλίνει σχεδόν άμεσα, ωστόσο η τελική λύση απέχει αρκετά από τη γενική βέλτιστη, γεγονός που επιβεβαιώνει την υπερβολικά εξερευνητική της φύση. Αντίθετα, οι βιβλιοθήκες Optuna και HyperOpt επιτυγχάνουν καλές αλλά όχι βέλτιστες λύσεις, με αισθητά μικρότερο χρόνο εκτέλεσης, λόγω της ελαφρύτερης φύσης του TPE μοντέλου σε σύγκριση με τις Gaussian Processes.

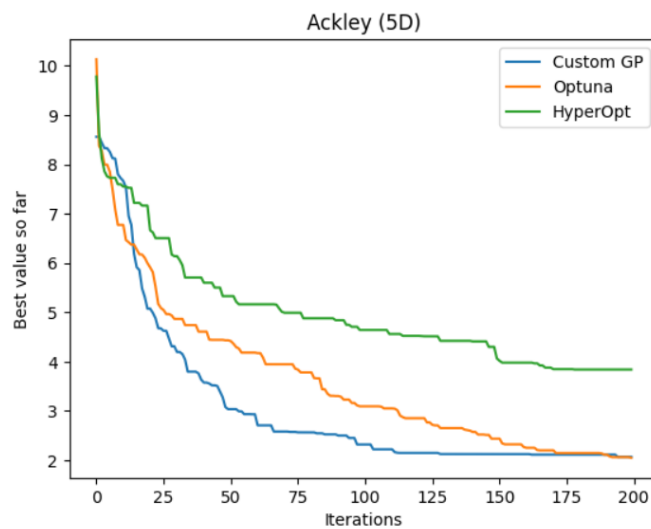


Εικόνα 13. Ackley (2D)-EI vs Optuna/HyperOpt

Ackley (5D)

Καθώς η διάσταση αυξάνεται, παρατηρείται σαφής διαφοροποίηση στη συμπεριφορά των μεθόδων. Η PI της custom υλοποίησης εμφανίζει καλύτερη τελική τιμή από την EI, υποδεικνύοντας μεγαλύτερη αποτελεσματικότητα σε μεσαίας διάστασης προβλήματα. Παράλληλα, απαιτείται μικρότερος αριθμός επαναλήψεων για τη σύγκλιση.

Η Optuna επιτυγχάνει συγκρίσιμη απόδοση με την PI, χρειάζεται μεγαλύτερο αριθμό επαναλήψεων, ενώ υπερέχει σημαντικά στον χρόνο εκτέλεσης. Η HyperOpt παρουσιάζει χειρότερη τελική λύση, αλλά με ταχύτερη σύγκλιση, γεγονός που αναδεικνύει την τάση του TPE να προτιμά την ταχύτητα έναντι της ακρίβειας.

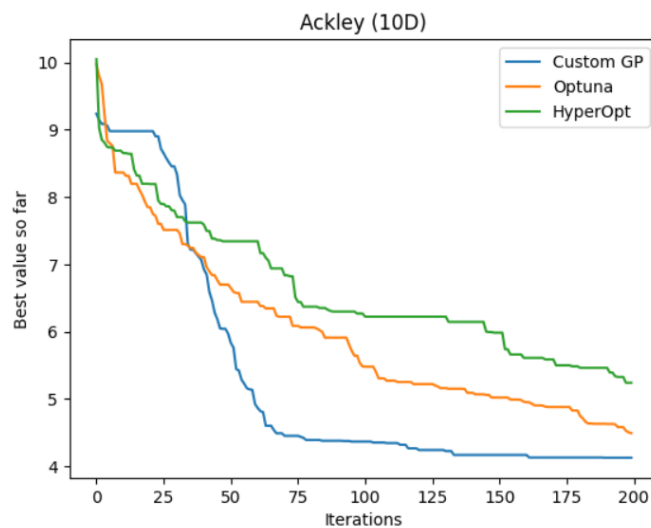


Εικόνα 14. Ackley (5D)-PI vs Optuna/HyperOpt

Ackley (10D)

Στη δεκαδιάστατη περίπτωση, όλες οι μέθοδοι παρουσιάζουν μειωμένη ακρίβεια, κάτι αναμενόμενο λόγω της αυξημένης δυσκολίας του προβλήματος. Η PI επιτυγχάνει την καλύτερη τελική τιμή μεταξύ των custom acquisition functions, ενώ η EI απαιτεί περισσότερες επαναλήψεις για συγκρίσιμη απόδοση.

Η Optuna εμφανίζει παρόμοια συμπεριφορά με την EI, τόσο ως προς την ποιότητα της λύσης όσο και ως προς τον αριθμό επαναλήψεων, ενώ η HyperOpt παρουσιάζει ελαφρώς χειρότερη τελική τιμή αλλά με χαμηλότερο υπολογιστικό κόστος. Και στις τρεις διαστάσεις, η UCB αποτυγχάνει να προσεγγίσει το ολικό ελάχιστο, παρά την ταχεία σύγκλιση.



Εικόνα 15. Ackley (10D)-PI vs Optuna/HyperOpt

5.4.2 Booth-Colville-Three hump Camel

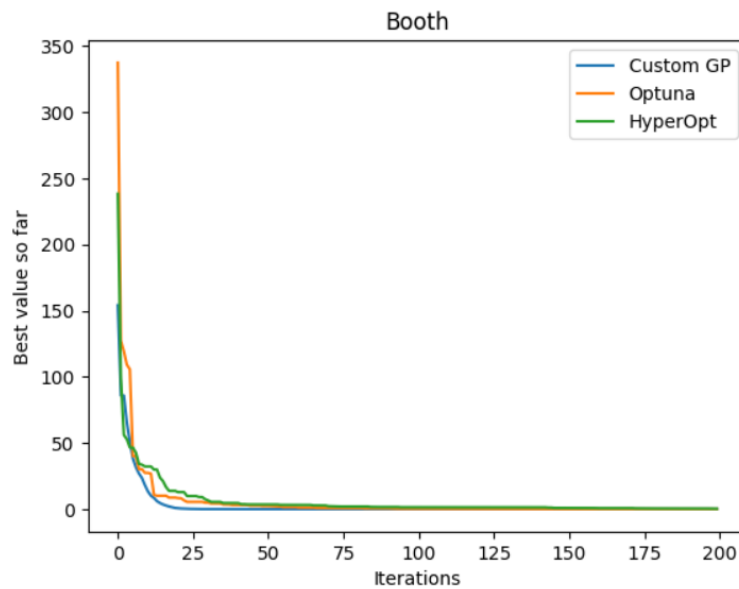
Πίνακας 18. Σύγκριση Booth-Colville-Three hump Camel με Optuna-HyperOpt (χωρίς θόρυβο)

Συνάρτηση	Acquisition Functions	min(y)	Επανάληψη (σύγκλισης)	Μέσος χρόνος εκτέλεσης (s)
Booth	EI	0.0002	134	28.93
Booth	PI	0.0003	109	27.80
Booth	UCB	103.7956	1	28.57
Booth	Optuna	0.0147	150	0.97
Booth	HyperOpt	0.2736	141	0.81
Colville	EI	70.1819	143	31.19
Colville	PI	56.7974	143	29.26
Colville	UCB	61.1777	1	28.05
Colville	Optuna	21.7862	165	1.82
Colville	HyperOpt	145.8486	151	1.46
Three-hump Camel	EI	0.0001	144	28.66
Three-hump Camel	PI	0.0000	108	29.76
Three-hump Camel	UCB	7.4134	4	28.22
Three-hump Camel	Optuna	0.0036	124	0.88
Three-hump Camel	HyperOpt	0.0826	103	0.87

Booth

Για τη συνάρτηση Booth, οι custom μέθοδοι με EI και PI επιτυγχάνουν εξαιρετικά χαμηλές τιμές ελάχιστου κόστους, πολύ κοντά στο θεωρητικό ολικό ελάχιστο ($\min(y) \approx 0$), με σύγκλιση σε λιγότερες επαναλήψεις από Optuna και HyperOpt. Αντίθετα, η χρήση της UCB αποτυγχάνει να προσεγγίσει το ελάχιστο, συγκλίνοντας ουσιαστικά από την πρώτη επανάληψη σε μη ικανοποιητική λύση, γεγονός που υποδεικνύει υπερβολική εκμετάλλευση.

Οι βιβλιοθήκες Optuna και HyperOpt επιτυγχάνουν αποδεκτά αποτελέσματα, ωστόσο με σαφώς υψηλότερες τελικές τιμές $\min(y)$ σε σχέση με τις custom μεθόδους EI και PI. Παρ' όλα αυτά, παρουσιάζουν πολύ μικρότερο χρόνο εκτέλεσης, επιβεβαιώνοντας το πλεονέκτημά τους σε υπολογιστική αποδοτικότητα.

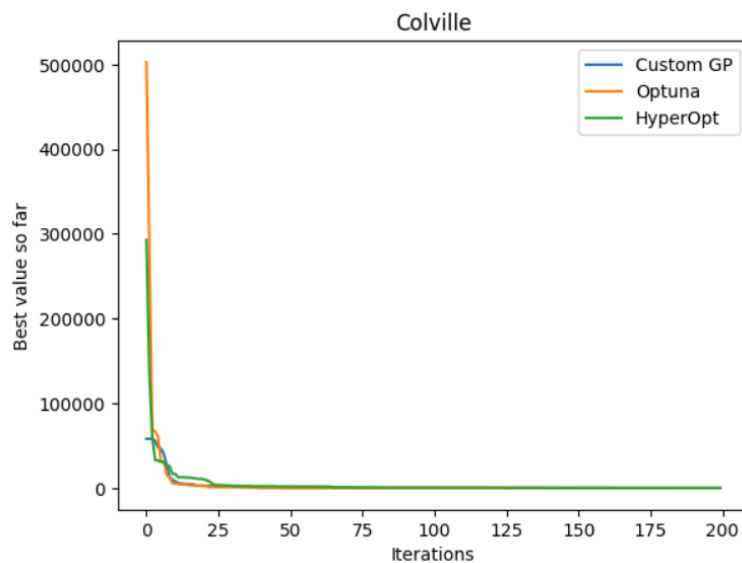


Εικόνα 16. Booth-El vs Optuna/HyperOpt

Colville

Η συνάρτηση Colville, ως πολυδιάστατη και έντονα μη γραμμική, αποδεικνύεται σημαντικά πιο απαιτητική. Οι custom υλοποιήσεις με EI και PI παρουσιάζουν δυσκολία στην προσέγγιση του ολικού ελαχίστου, με σχετικά υψηλές τελικές τιμές $\min(y)$, ενώ η UCB αποτυγχάνει εκ νέου να βελτιώσει τη λύση.

Αξιοσημείωτο είναι ότι η Optuna επιτυγχάνει την καλύτερη συνολική απόδοση για τη συγκεκριμένη συνάρτηση, παρουσιάζοντας χαμηλότερη τιμή $\min(y)$ από όλες τις custom μεθόδους, γεγονός που υποδηλώνει ότι οι ενσωματωμένοι μηχανισμοί δειγματοληψίας (π.χ. TPE) είναι ιδιαίτερα αποτελεσματικοί σε σύνθετα τοπία αναζήτησης. Η HyperOpt, αν και ταχύτερη, εμφανίζει τη χειρότερη τελική απόδοση στη συγκεκριμένη περίπτωση.

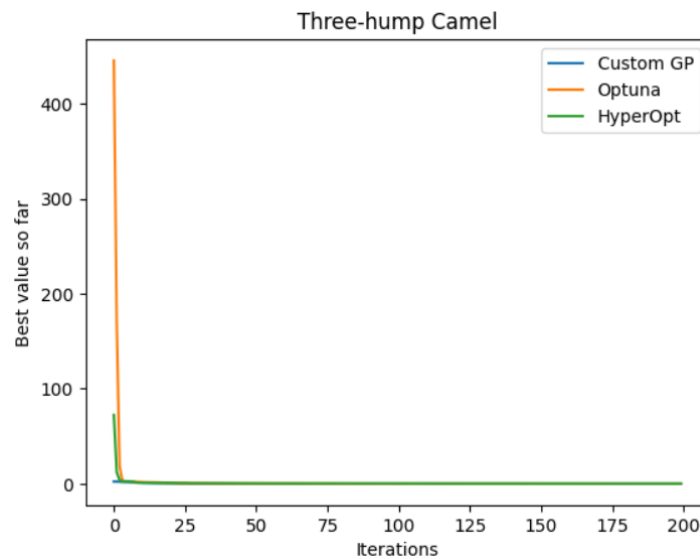


Εικόνα 17. Colville-El vs Optuna/HyperOpt

Three-hump Camel

Για τη συνάρτηση Three-hump Camel, οι custom μέθοδοι με EI και PI επιτυγχάνουν εξαιρετικά ακριβή αποτελέσματα, με $\min(y)$ σχεδόν μηδενικό και ικανοποιητική ταχύτητα σύγκλισης. Αντίθετα, η UCB αποτυγχάνει και πάλι να εντοπίσει το σωστό ελάχιστο, συγκλίνοντας πολύ νωρίς σε υποβέλτιστες λύσεις.

Οι βιβλιοθήκες Optuna και HyperOpt εμφανίζουν καλή αλλά σαφώς κατώτερη ακρίβεια σε σχέση με τις custom EI και PI, αν και με σημαντικά μικρότερο υπολογιστικό κόστος. Το αποτέλεσμα αυτό αναδεικνύει ότι, για απλές και ομαλές συναρτήσεις, οι κλασικές acquisition functions μπορούν να είναι πιο ακριβείς από γενικού σκοπού αυτοματοποιημένους βελτιστοποιητές.



Εικόνα 18. Three-hump Camel-Optuna/HyperOpt

5.4.3 Styblinski-Tang

Πίνακας 19. Σύγκριση Styblinski-Tang με Optuna-HyperOpt (χωρίς θόρυβο)

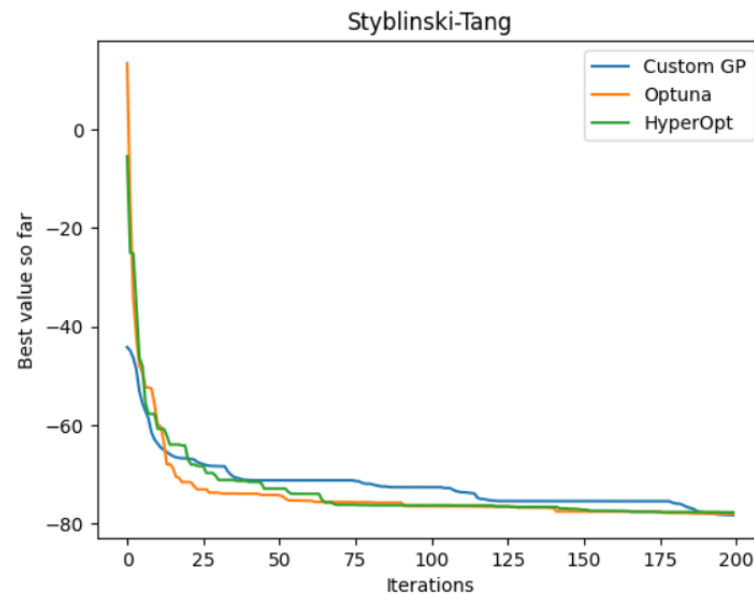
Συνάρτηση	Acquisition Functions	min(y)	Επανάληψη (σύγκλισης)	Μέσος χρόνος εκτέλεσης (s)
Styblinski-Tang (2D)	EI	-78.330	151	27.32
Styblinski-Tang (2D)	PI	-78.332	127	26.44
Styblinski-Tang (2D)	UCB	-56.823	10	28.28
Styblinski-Tang (2D)	Optuna	-78.044	166	0.87
Styblinski-Tang (2D)	HyperOpt	-77.764	161	0.91
Styblinski-Tang (5D)	EI	-178.755	135	32.11
Styblinski-Tang (5D)	PI	-172.130	132	30.48
Styblinski-Tang (5D)	UCB	-111.79	7	28.80
Styblinski-Tang (5D)	Optuna	-180.536	154	2.27
Styblinski-Tang (5D)	HyperOpt	-163.524	126	2.00
Styblinski-Tang (10D)	EI	-248.799	97	31.99
Styblinski-Tang (10D)	PI	-249.492	122	33.18
Styblinski-Tang (10D)	UCB	-238.686	68	28.73
Styblinski-Tang (10D)	Optuna	-310.227	159	4.79
Styblinski-Tang (10D)	HyperOpt	-277.966	107	3.72

Ο Πίνακας 19 παρουσιάζει τη σύγκριση των custom υλοποιήσεων Bayesian Optimization (EI, PI, UCB) με τις βιβλιοθήκες Optuna και HyperOpt για την Styblinski-Tang συνάρτηση, σε τρεις διαστάσεις (2D, 5D, 10D).

Styblinski-Tang 2D

Στη 2D περίπτωση, οι custom μέθοδοι EI και PI επιτυγχάνουν τις καλύτερες τιμές min(y) (~ -78.33), οι οποίες είναι πιο ακριβείς από αυτές που παρέχουν οι βιβλιοθήκες Optuna και HyperOpt. Η UCB υπολείπεται σημαντικά, φτάνοντας σε υψηλότερα min(y), γεγονός που δείχνει πρόωρη σύγκλιση σε μη βέλτιστες λύσεις.

Παράλληλα, οι βιβλιοθήκες Optuna και HyperOpt εμφανίζουν σημαντικά μικρότερο χρόνο εκτέλεσης, με την HyperOpt να ολοκληρώνει σχεδόν άμεσα. Αυτό αναδεικνύει τη συνήθη εμπορική τάση: μικρή απώλεια ακριβείας έναντι σημαντικής κερδισμένης υπολογιστικής ταχύτητας.

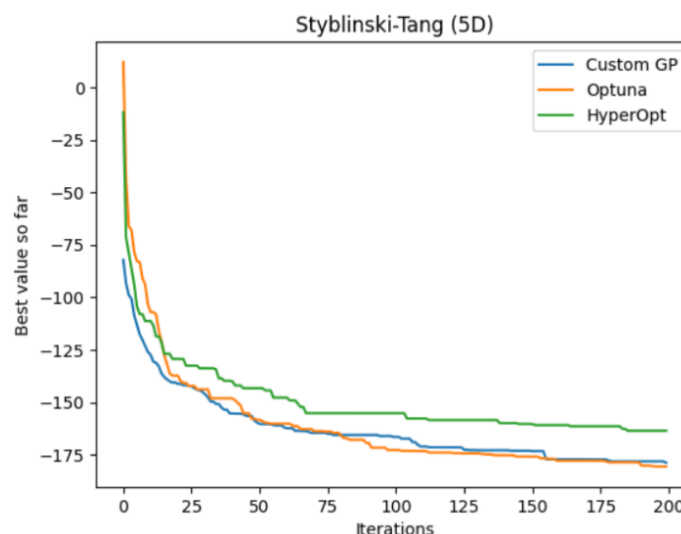


Εικόνα 19. Styblinski-Tang (2D)-PI vs Optuna/HyperOpt

Styblinski-Tang 5D

Η αύξηση της διάστασης δυσκολεύει την προσαρμογή των custom acquisition functions. Οι EI και PI παραμένουν σχετικά ισχυρές, επιτυγχάνοντας $\min(y) \sim -172$ έως -178 , με αριθμό επαναλήψεων γύρω στις 130–135. Η UCB συνεχίζει να εμφανίζει μειωμένη ακρίβεια, με πολύ μικρότερο αριθμό επαναλήψεων.

Η Optuna και η HyperOpt προσεγγίζουν σε μεγαλύτερη κλίμακα τις αντίστοιχες τιμές, με μικρότερο αριθμό επαναλήψεων και ασυναγώνιστο χρόνο εκτέλεσης (~ 2 s). Παρά την απώλεια ακρίβειας, η ταχύτητα και η σταθερότητα αυτών των βιβλιοθηκών γίνεται εμφανής σε υψηλότερες διαστάσεις.

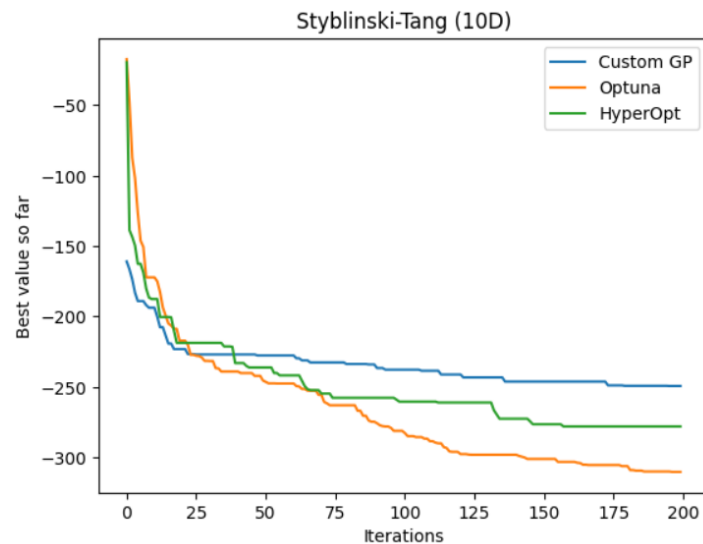


Εικόνα 20. Styblinski-Tang (5D)-Optuna/HyperOpt

Styblinski–Tang 10D

Στην περίπτωση των 10 διαστάσεων, οι custom μέθοδοι παραμένουν ανταγωνιστικές. Η PI ξεπερνά ελαφρώς την EI, επιτυγχάνοντας το χαμηλότερο $\min(y)$ (~ -249.49), ενώ η UCB υπολείπεται και πάλι σημαντικά.

Η Optuna και η HyperOpt συνεχίζουν να διατηρούν μικρότερο αριθμό επαναλήψεων και ταχύτατο χρόνο εκτέλεσης, προσφέροντας έναν πρακτικό συμβιβασμό μεταξύ ακρίβειας και υπολογιστικού κόστους σε υψηλές διαστάσεις.



Εικόνα 21. Styblinski-Tang (10D)-PI vs Optuna/HyperOpt

5.5 Agglomerative Clustering Αλγόριθμος

Προχωρώντας στην εφαρμογή σε πραγματικά προβλήματα Μηχανικής Μάθησης, η αρχική φάση της διαδικασίας βασίζεται και πάλι στα αποτελέσματα που παρήγαγε η custom υλοποίησή μας κατά τη ρύθμιση των υπερπαραμέτρων του αλγορίθμου Ιεραρχικής Συσταδοποίησης (Agglomerative Clustering). Στόχος ήταν η αυτόματη εύρεση του βέλτιστου συνδυασμού του αριθμού των συστάδων, της μετρικής απόστασης και του κριτηρίου σύνδεσης (linkage), με σκοπό τη μεγιστοποίηση του Silhouette Score. Συγκεκριμένα, αναλύσαμε και συγκρίναμε τους διαφορετικούς συνδυασμούς τεχνικών κωδικοποίησης με τις custom Γκαουσιανές Διαδικασίες (Gaussian Processes - GP), καθώς και με τις υλοποιήσεις των πλαισίων Optuna και HyperOpt. Τα αποτελέσματα παρατίθενται στον παρακάτω πίνακα:

Πίνακας 20. Σύγκριση Custom GP's με Optuna-HyperOpt (Agglomerative Clustering)

Encoder	Acquisition Functions	Silhouette Score	Επανάληψη (σύγκλισης)	Μέσος χρόνος εκτέλεσης (s)
onehot	EI	0.582	21	1.18
hash	EI	0.582	22	1.24
target	EI	0.582	18	1.13
onehot	HyperOpt	0.582	60	0.93
hash	HyperOpt	0.582	60	0.97
target	HyperOpt	0.582	60	0.98
onehot	Optuna	0.582	39	0.49
hash	Optuna	0.580	15	0.49
target	Optuna	0.582	18	0.51
onehot	PI	0.582	6	1.09
hash	PI	0.581	33	0.91
target	PI	0.582	13	1.25
onehot	Thompson Sampling	0.582	21	1.11
hash	Thompson Sampling	0.582	22	1.06
target	Thompson Sampling	0.582	14	1.1

Στα πειράματα με Agglomerative Clustering, παρατηρούμε ότι όλες οι μέθοδοι και οι encoders καταλήγουν σε παρόμοιο silhouette score (~0.582), υποδεικνύοντας ότι η ποιότητα του clustering είναι γενικά ανεξάρτητη από την acquisition function ή τον τρόπο κωδικοποίησης. Ωστόσο, οι μέθοδοι διαφέρουν στην ταχύτητα σύγκλισης και στο runtime: η Custom GP με PI ή Thompson Sampling φτάνει γρήγορα σε καλό αποτέλεσμα σε λίγα iterations (6–22), αλλά έχει μεγαλύτερο μέσο runtime (~1.1–1.2 s), ενώ η Optuna προσφέρει γρήγορο runtime (~0.49–0.51 s) με μέτρια convergence (15–39 iterations) και το HyperOpt είναι πιο σταθερό αλλά πιο αργό στη σύγκλιση (~60 iterations, ~0.93–0.98 s). Οι encoders δεν επηρεάζουν σημαντικά τα τελικά scores, με μικρές αποκλίσεις κυρίως στο

convergence. Συνολικά, για γρήγορη σύγκλιση σε λίγα iterations προτιμάται η Custom GP με κατάλληλη acquisition function, ενώ για σταθερό και γρήγορο runtime η Optuna είναι η πιο αποτελεσματική επιλογή.

5.6 Σχολιασμός των Πειραματικών Αποτελεσμάτων

Η πειραματική αξιολόγηση κατέδειξε τις διαφορές μεταξύ των custom υλοποιήσεων της Μπεϋζιανής Βελτιστοποίησης και των βιβλιοθηκών Optuna και HyperOpt σε προβλήματα διαφορετικής διάστασης και πολυπλοκότητας. Η μελέτη περιελάμβανε μονοδιάστατες συναρτήσεις, όπως οι f1–f4, και πολυδιάστατες benchmark συναρτήσεις, όπως Ackley, Booth, Colville, Three-hump Camel και Styblinski–Tang, με και χωρίς θόρυβο.

Στα 1D προβλήματα, οι custom acquisition functions Expected Improvement (EI) και Probability of Improvement (PI) επέδειξαν εξαιρετική ακρίβεια, επιτυγχάνοντας τιμές $\min(y)$ πολύ κοντά στις πραγματικές βέλτιστες λύσεις. Η UCB, λόγω της τάσης της για έντονη εκμετάλλευση, παρουσίασε μεγαλύτερη διακύμανση και πρόωρη σύγκλιση σε μη βέλτιστες λύσεις. Οι βιβλιοθήκες Optuna και HyperOpt ολοκλήρωσαν τις εκτελέσεις σημαντικά ταχύτερα, αλλά με μικρή απώλεια ακρίβειας σε σχέση με τις custom μεθόδους.

Στα πολυδιάστατα προβλήματα, η εικόνα διαφοροποιείται ανάλογα με τη διάσταση και την πολυπλοκότητα της συνάρτησης:

- Στην Ackley, η custom PI επιτυγχάνει τις χαμηλότερες τιμές $\min(y)$ σε υψηλές διαστάσεις, ενώ η EI παραμένει πολύ ανταγωνιστική. Η UCB εμφανίζει έντονη υπερβολική εκμετάλλευση και συχνά μη βέλτιστες τιμές, ιδιαίτερα σε υψηλές διαστάσεις (10D). Οι βιβλιοθήκες Optuna και HyperOpt, αν και πιο γρήγορες, δείχνουν μικρή απώλεια ακρίβειας σε σύγκριση με τις custom μεθόδους.
- Στις συνεχείς, χαμηλής διάστασης συναρτήσεις όπως Booth και Three-hump Camel, οι custom μέθοδοι EI και PI συγκλίνουν σταθερά πολύ κοντά στο global minimum, ενώ οι βιβλιοθήκες παραμένουν ταχύτερες αλλά ελαφρώς λιγότερο ακριβείς.
- Στις πολύπλοκες υψηλής διάστασης συναρτήσεις όπως Colville και Styblinski–Tang, η PI φαίνεται να υπερτερεί ελαφρώς της EI σε όρους ακρίβειας, ενώ η UCB υπολείπεται σημαντικά. Παράλληλα, Optuna και HyperOpt διατηρούν σαφώς μικρότερο χρόνο εκτέλεσης και μικρότερο αριθμό επαναλήψεων, προσφέροντας έναν πρακτικό συμβιβασμό μεταξύ ακρίβειας και υπολογιστικού κόστους.

Συνολικά, η σύγκριση δείχνει ότι:

1. *Custom EI και PI*: Υπερέχουν σε ακρίβεια, ιδιαίτερα σε προβλήματα χαμηλής έως μέτριας διάστασης και σε συνεχείς συναρτήσεις. Οι μέθοδοι αυτές διατηρούν σταθερή απόδοση, αλλά απαιτούν υψηλότερο υπολογιστικό κόστος και μεγαλύτερο αριθμό επαναλήψεων.
2. *Custom UCB*: Προσφέρει γρήγορη σύγκλιση, αλλά με σημαντική απώλεια ακρίβειας, ειδικά σε υψηλές διαστάσεις.
3. *Optuna και HyperOpt*: Παρέχουν πολύ χαμηλούς χρόνους εκτέλεσης και μικρότερο αριθμό επαναλήψεων, γεγονός που τα καθιστά ιδανικά για εφαρμογές όπου η ταχύτητα είναι κρίσιμη. Η απώλεια ακρίβειας είναι μικρή για τις περισσότερες 1D και μέσης διάστασης συναρτήσεις, αλλά αυξάνεται για πολύπλοκες και υψηλής διάστασης συναρτήσεις.
4. *Agglomerative Clustering*: Οι custom acquisition functions EI/PI προσφέρουν καλύτερη ποιότητα συσταδοποίησης και σταθερότερη σύγκλιση σε σχέση με τις TPE-based μεθόδους, ενώ η επιλογή κωδικοποίησης και linkage επηρεάζει ελαφρώς την ταχύτητα και ακρίβεια.

Η συστηματική απώλεια ακρίβειας που παρουσιάζουν τα βιομηχανικά πλαίσια Optuna και HyperOpt έναντι των custom μεθόδων βασισμένων σε Γκαουσιανές Διαδικασίες (Gaussian Processes - GP) εντοπίζεται στον υποκείμενο αλγόριθμο βελτιστοποίησης που χρησιμοποιούν, ο οποίος είναι ο Tree-structured Parzen Estimator (TPE). Ο TPE είναι μια μέθοδος που δεν μοντελοποιεί άμεσα την ίδια την αντικειμενική συνάρτηση, αλλά την πυκνότητα πιθανότητας των δεδομένων, χωρίζοντάς τα στοχαστικά σε «καλές» και «κακές» παρατηρήσεις. Επιπλέον, για λόγους υπολογιστικής απλοποίησης, ο TPE εισάγει την παραδοχή ότι οι υπερπαραμέτροι είναι ανεξάρτητες μεταξύ τους. Αυτή η προσέγγιση επιτρέπει στα Optuna και HyperOpt να διαχειρίζονται τον χώρο αναζήτησης με εξαιρετική ταχύτητα και χαμηλό υπολογιστικό κόστος, ωστόσο αδυνατεί να αποτυπώσει σύνθετες, μη γραμμικές αλληλεπιδράσεις (interactions) και συσχετίσεις μεταξύ των μεταβλητών, ιδιαίτερα σε πολυδιάστατα τοπία.

Αντίθετα, η custom υλοποίηση Γκαουσιανών Διαδικασιών που αναπτύχθηκε μοντελοποιεί καθολικά την αντικειμενική συνάρτηση και αποτυπώνει πλήρως τη συνδιακύμανση ολόκληρου του χώρου αναζήτησης μέσω του επιλεγμένου πυρήνα (kernel). Διατηρώντας ακέραιη τη στατιστική πληροφορία και τη συσχέτιση των παραμέτρων, οι custom συναρτήσεις απόκτησης (κυρίως η Expected Improvement) εκτελούν μια σαφώς πιο λεπτομερή, συστηματική και ποιοτική εξερεύνηση (exploration) του χώρου. Αυτό επιτρέπει στο custom πλαίσιο να εντοπίζει με μεγαλύτερη ακρίβεια το καθολικό βέλτιστο (global minimum), εκεί όπου ο στοχαστικός χαρακτήρας του TPE τείνει να παγιδευτεί σε τοπικά ακρότατα χάριν της υπολογιστικής ταχύτητας.

Συμπερασματικά, η επιλογή του κατάλληλου εργαλείου εξαρτάται από τη φύση του προβλήματος και τους στόχους της βελτιστοποίησης: για μέγιστη ακρίβεια και λεπτομερή έρευνα του χώρου αναζήτησης, οι custom συναρτήσεις EI/PI είναι προτιμότερες. Για υψηλή ταχύτητα και χαμηλό υπολογιστικό κόστος, οι βιβλιοθήκες Optuna και HyperOpt αποτελούν πρακτική επιλογή, ιδιαίτερα σε πολυδιάστατα ή κατηγορικά προβλήματα.

6. Συμπεράσματα

Στην παρούσα εργασία μελετήθηκε η Μπεϋζιανή Βελτιστοποίηση (Bayesian Optimization, BO) και η εφαρμογή της σε μια σειρά από προβλήματα βελτιστοποίησης, τόσο μονοδιάστατα όσο και πολυδιάστατα. Η μελέτη περιλάμβανε τόσο custom υλοποιήσεις acquisition functions (Expected Improvement, Probability of Improvement, Upper Confidence Bound, Thompson Sampling), όσο και τη σύγκριση με βιβλιοθήκες ευρείας χρήσης όπως Optuna και HyperOpt.

Η πειραματική αξιολόγηση απέδειξε ότι η BO είναι ιδιαίτερα αποτελεσματική στην εύρεση των βέλτιστων λύσεων για προβλήματα συνεχών και ομαλών συναρτήσεων, με τις custom μεθόδους EI και PI να υπερέχουν σε ακρίβεια, ειδικά σε χαμηλής και μέσης διάστασης προβλήματα. Η UCB, ενώ επιταχύνει τη σύγκλιση, εμφανίζει συχνά πρόωρη σύγκλιση σε μη βέλτιστες λύσεις, ιδιαίτερα σε υψηλές διαστάσεις.

Οι βιβλιοθήκες Optuna και HyperOpt παρουσίασαν σαφώς χαμηλότερο χρόνο εκτέλεσης και απαιτούσαν λιγότερες επαναλήψεις για σύγκλιση, καθιστώντας τις πρακτικές επιλογές για προβλήματα υψηλής διάστασης ή όταν το υπολογιστικό κόστος είναι κρίσιμο. Ωστόσο, σε ορισμένες περιπτώσεις, η ακρίβεια των λύσεων τους ήταν ελαφρώς χαμηλότερη σε σύγκριση με τις custom μεθόδους, ιδίως σε προβλήματα που απαιτούν λεπτομερή εξερεύνηση του χώρου αναζήτησης.

Η μελέτη ανέδειξε επίσης τη σημασία της επιλογής κατάλληλου surrogate model και της acquisition function, καθώς αυτά επηρεάζουν άμεσα τη σύγκλιση και την ακρίβεια. Οι Gaussian Processes αποδείχθηκαν ιδανικές για συνεχείς και χαμηλής διάστασης συναρτήσεις, ενώ οι μέθοδοι TPE (Optuna/HyperOpt) και SMAC προσφέρουν καλύτερη απόδοση σε κατηγορικά ή υψηλής διάστασης προβλήματα, όπου οι Gaussian Processes δυσκολεύονται.

Συνολικά, η εργασία επιβεβαιώνει ότι η Μπεϋζιανή Βελτιστοποίηση αποτελεί ένα ισχυρό εργαλείο για τη βελτιστοποίηση πολύπλοκων συναρτήσεων, ενώ η προσεκτική επιλογή των παραμέτρων και των surrogate models είναι κρίσιμη για την αποτελεσματική εφαρμογή της. Η επέκταση σε υψηλής διάστασης προβλήματα με περιορισμούς, καθώς και η διερεύνηση διαφορετικών acquisition functions και samplers, αποτελούν ενδιαφέρουσες κατευθύνσεις για μελλοντική έρευνα.

Βιβλιογραφία

1. Nocedal, J., & Wright, S. J. (2006). *Numerical Optimization* (2nd ed.). Springer.
2. Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.
3. Spall, J. C. (2003). *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Wiley.
4. Rios, L. M., & Sahinidis, N. V. (2013). Derivative-free optimization: A review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3), 1247–1293.
5. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & de Freitas, N. (2016). Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE*, 104(1), 148–175.
6. Frazier, P. I. (2018). A Tutorial on Bayesian Optimization. *arXiv preprint arXiv:1807.02811*.
7. Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. *NIPS*.
8. Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. *NIPS*.
9. Brochu, E., Cora, V. M., & De Freitas, N. (2010). *A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning*. arXiv preprint arXiv:1012.2599.
10. Mockus, J. (1978). *Application of Bayesian Approach to Numerical Methods of Global and Stochastic Optimization*. *Journal of Global Optimization*, 4(4), 347–365.
11. Jones, D. R., Schonlau, M., & Welch, W. J. (1998). *Efficient Global Optimization of Expensive Black-Box Functions*. *Journal of Global Optimization*, 13(4), 455–492.
12. Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.
13. J. Mockus, *Bayesian Approach to Global Optimization*, Kluwer Academic Publishers, 1989.
14. Yang, Z., & Shami, A. (2020). *On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice*. *Neurocomputing*, 415, 295–316.
15. J. Bergstra, D. Yamins and D. D. Cox, “Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures,” *International Conference on Machine Learning (ICML)*, 2013.
16. Hutter, F., Hoos, H. H., & Leyton-Brown, K. (2011). *Sequential Model-Based Optimization for General Algorithm Configuration (SMAC)*. *LION 5 Conference*.
17. Zhong, Y., Liu, W., & Yao, Q. (2023). *Weighted Expected Improvement with Dynamic Trade-offs*. *arXiv preprint arXiv:2306.04262*.

18. T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.
19. M. Feurer, A. Klein, S. Falkner, J. T. Springenberg, F. Hutter and F. Hutter, "Auto-Sklearn: Efficient and Robust Automated Machine Learning," *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.
20. F. Hutter, H. H. Hoos and K. Leyton-Brown, "An Efficient Approach for Assessing Hyperparameter Importance," *International Conference on Machine Learning (ICML)*, 2014.