

Δειγματοληψία με χρήση Κανονικών Κανονμών και
εφαρμογή στην Καθολική Βελτιστοποίηση

Η ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

υποβάλλεται στην
ορισθείσα από την Γενική Συνέλευση Ειδικής Σύθεσης
του Τμήματος Πληροφορικής Εξεταστική Επιτροπή

από τον

Σωτήρη Χατζή

ως μέρος των Υποχρεώσεων για τη λήψη του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ
ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ
ΣΤΟΥΣ ΕΠΙΣΤΗΜΟΝΙΚΟΥΣ ΥΠΟΛΟΓΙΣΜΟΥΣ

ΙΟΥΛΙΟΣ 2008

ΕΥΧΑΡΙΣΤΙΕΣ

Για την περάτωση της παρούσης εργασίας, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου Ισαάκ Η. Λαγαρή, για την συνεχή του καθοδήγηση και αμέριστη συμπαράστασή του. Επίσης οφείλω ευχαριστίες στον Κ. Βόγκλη που η βοήθειά του ήταν χρήσιμη και καθοριστική σε περιπτώσεις υλοποίησης.

Αφιερώνω την παρούσα εργασία στους γονείς μου για την αμέριστη συμπαράσταση όλων των χρόνων σπουδών μου.

ΠΕΡΙΕΧΟΜΕΝΑ

1	ΓΕΝΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΜΕΘΟΔΩΝ ΚΑΘΟΛΙΚΗΣ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗΣ	5
2	ΣΤΟΧΑΣΤΙΚΕΣ ΜΕΘΟΔΟΙ ΚΑΙ ΚΑΘΟΛΙΚΗ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗ	11
3	ΚΡΙΤΗΡΙΑ ΤΕΡΜΑΤΙΣΜΟΥ ΓΙΑ ΣΤΟΧΑΣΤΙΚΕΣ ΜΕΘΟΔΟΥΣ MULTISTART	16
4	ΚΑΘΟΛΙΚΗ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗ ΜΕ ΧΡΗΣΗ ΚΑΝΟΝΙΚΗΣ ΚΑΤΑΝΟΜΗΣ	24
5	ΠΕΙΡΑΜΑΤΑ ΚΑΙ ΜΕΤΡΗΣΕΙΣ	39
6	ΕΠΙΛΟΓΟΣ	43
7	ΠΑΡΑΡΤΗΜΑ	46

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

1.1	<i>Μέθοδος branch and bound</i>	8
2.1	<i>Παράδειγμα τοπικής αναζήτησης</i>	12
2.2	<i>Παράδειγμα μη τοπικής αναζήτησης</i>	13
3.1	<i>Περιοχή έλξης συνάρτησης Rastrigin</i>	18
4.1	<i>Αποτέλεσμα εκτέλεσης αλγορίθμου multistart στην συνάρτηση Rastrigin, στο πεδίο τιμών [-1.2,1.2]. Παρατηρούμε ότι η αναζήτηση τοπικού ελαχίστου πολλές φορές οδηγεί σημεία στο ίδιο, παραπάνω από μια φορά.</i>	26
4.2	<i>Εμφάνιση επικάλυψης σε τοπική αναζήτηση ελαχίστου.</i>	29
4.3	<i>Επιλογή πιθανού σημείου</i>	33
4.4	<i>Δειγματοληψία στη συνάρτηση Six Hump με χρήση ομοιόμορφης και κανονικής κατανομή</i>	34
4.5	<i>Δειγματοληψία σημείων στη συνάρτηση Rastrigin με χρήση ομοιόμορφης και κανονικής κατανομής</i>	34
4.6	<i>On-line προσδιορισμός παραμέτρων ελάχιστο στο $x^* = [4, 0]^T$</i>	37
7.1	<i>Δειγματοληψία στη συνάρτηση Goldstein με χρήση ομοιόμορφης και κανονικής κατανομή</i>	46
7.2	<i>Δειγματοληψία στη συνάρτηση Booth με χρήση ομοιόμορφης και κανονικής κατανομή</i>	47
7.3	<i>Δειγματοληψία στη συνάρτηση Sphere με χρήση ομοιόμορφης και κανονικής κατανομή</i>	47
7.4	<i>Δειγματοληψία στη συνάρτηση Shubert με χρήση ομοιόμορφης και κανονικής κατανομή</i>	48
7.5	<i>Rastrigin function</i>	50
7.6	<i>Shubert function</i>	51
7.7	<i>Hump function</i>	52
7.8	<i>Grienwank function</i>	53
7.9	<i>Michalevic function</i>	54
7.10	<i>Schwefel function</i>	55
7.11	<i>Goldstein function</i>	56

ΕΥΡΕΤΗΡΙΟ ΑΛΓΟΡΙΘΜΩΝ

1	Πολλαπλή Εκκίνηση	25
2	Adapt Multistart	26
3	Έλεγχος	27
4	Καθολική Ελαχιστοποίησης με χρήση Κανονικής Κατανομής	31
5	Rejection Sampling	32
6	Inverse Rejection Sampling	33
7	Αλγόριθμος Δειγματοληψίας με χρήση κανονικής πολυδιάστατης κατανομής	49

ΠΡΟΛΟΓΟΣ

Η Βελτιστοποίηση είναι ένας κλάδος των Θετικών επιστημών που ασχολείται με την εύρεση βέλτιστων λύσεων σε εφαρμογές όπως σε προβλήματα Οικονομικών Επιστημών, Μαθηματικών και Φυσικής, μοντελοποίησης και σωρεία πολλών προβλημάτων. Η φύση των παραπάνω προβλημάτων είναι αρκετά πολύπλοκη, οπότε η δημιουργία γρήγορων και αποτελεσματικών μεθόδων γίνεται επιτακτική.

Στόχος της εργασίας είναι να χρησιμοποιήσουμε πληροφορία από τοπικές ελαχιστοποιήσεις, έτσι ώστε να δημιουργήσουμε μια πολυδιάστατη κατανομή, από την οποία θα δειγματοληπούμε σημεία για αρχικά σημεία νέων αναζητήσεων. Σκοπός είναι να οδηγήσουμε τα αρχικά σημεία, μέσω της κατανομής, σε περιοχές του χώρου αναζήτησης που δεν έχουν ακόμα εξερευνηθεί.

Στα επόμενα κεφάλαια θα ξεκινήσουμε με μία εισαγωγή στην Καθολική Βελτιστοποίηση αναφέροντας για τα χαρακτηριστικά που συνθέτουν ένα πρόβλημα καθολικής βελτιστοποίησης και στις κυριώτερες μεθόδους καθολικής ελαχιστοποίησης. Στη συνέχεια αναλύουμε το πρόβλημα επιλογής κριτηρίου τερματισμού και τέλος παραθέτουμε τα βήματα του αλγορίθμου multistart, τα βήματα του αλγορίθμου Adapt Multistart και αναλύουμε την πρότασή μας καθώς και πειραματικά αποτελέσματα της εργασίας μας.

Για την περάτωση των πειραμάτων οι συναρτήσεις που επιλέχθηκαν, για την πραγματοποίηση των πειραμάτων είναι κυρίως δύσκολα προβλήματα καθολικής ελαχιστοποίησης.

Η δομή της εργασίας έχει ως εξής:

Κεφάλαιο 1 - Ορίζουμε το καθολικό και τοπικό ελάχιστο και αναφέρουμε τα χαρακτηριστικά των μεθόδων καθολικής βελτιστοποίησης

Κεφάλαιο 2 - Περιγράφουμε τις πιο γνωστές στοχαστικές μεθόδους

Κεφάλαιο 3 - Ανάλυση κριτηρίων τερματισμού που έχουν προταθεί σε βιβλιογραφία

Κεφάλαιο 4 - Ανάλυση και περιγραφή των αλγορίθμων Multistart και Adapt Multistart. Περιγραφή των βημάτων της πρότασής μας.

Κεφάλαιο 5 - Παράθεση αποτελεσμάτων και συμπερασμάτων

Κεφάλαιο 6 - Επίλογος

Παράρτημα

ΚΕΦΑΛΑΙΟ 1

ΓΕΝΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΜΕΘΟΔΩΝ ΚΑΘΟΛΙΚΗΣ ΕΛΑΧΙΣΤΟΠΟΗΣΗΣ

Τα κύρια χαρακτηριστικά που συνθέτουν ένα πρόβλημα βελτιστοποίησης είναι τρία: i) ένα σύνολο αγνώστων ή μεταβλητών, ii) ο τύπος της συνάρτησης που θέλουμε να βελτιστοποιήσουμε και iii) τέλος οι περιορισμοί που πρέπει να πληρεί η επιστρεφόμενη βέλτιστη λύση. Με βάση τα παραπάνω μπορούμε να πούμε ότι βελτιστοποίηση είναι η διαδικασία εύρεσης τιμών που ελαχιστοποιούν ή μεγιστοποιούν την συνάρτηση.

Υπάρχουν διάφοροι τύποι προβλημάτων βελτιστοποίησης. Οι μεταβλητές μπορεί να λαμβάνουν συνεχείς ή διακριτές τιμές. Η συνάρτηση βελτιστοποίησης μπορεί να είναι συνεχής ή διακριτή. Οι περιορισμοί στα διάφορα προβλήματα έχουν διαφορετική μορφή (γραμμικοί ή όχι), ή ακόμα μπορεί να απουσιάζουν. Όλες αυτές οι παραμέτρους συνθέτουν μια ποικιλία στα προβλήματα βελτιστοποίησης, γεγονός που καθιστά το πεδίο μελέτης τους, ευρύ. Σήμερα αποτελεί ένα από τα πλέον σημαντικά θέματα έρευνας στον τομέα των εφαρμοσμένων μαθηματικών και της πληροφορικής.

1.1 Στοιχεία καθολικής ελαχιστοποίησης

Το πρόβλημα της καθολικής βελτιστοποίησης είναι η εύρεση εκείνου του σημείου, έστω x^* , μιας συνάρτησης f με πραγματικές τιμές, τέτοιο ώστε $f(x^*) \leq f(x)$. Δυστυχώς όμως, λίγες μέθοδοι έχουν μελετηθεί για την εύρεση ολικού ελαχίστου σε σύγκριση με την πληθώρα μεθόδων, που υπάρχουν για την εύρεση τοπικού ελαχίστου. Σ' αυτό το σημείο κρίνεται σκόπιμο να ορίσουμε την έννοιά του. Έτσι λοιπόν ένα σημείο $f(x^*)$, με $x^* \in \mathbb{R}^n$ (όπου n , η διάσταση του σημείου), τέτοιο ώστε να υπάρχει μια γειτονία του B του x^* με την ιδιότητα :

$$f(x^*) \leq f(x) \forall x \in B \quad (1.1)$$

Από τον παραπάνω ορισμό είναι φανερό η δυσκολία εύρεσης καθολικού ελαχίστου σε σύγκριση με την εύρεση κάποιου τοπικού ελαχίστου. Αν θεωρήσουμε ότι η συνάρτηση που θέλουμε να ελαχιστοποιήσουμε είναι δύο φορές διαφορίσιμη, τότε οι ικανές συνθήκες για την ύπαρξη ελαχίστου είναι:

1. Το διάνυσμα πρώτης παραγώγου στο σημείο x^* να είναι μηδέν (δηλαδή $\nabla f(x^*) = 0$).
2. Ο Εσσιανός πίνακας στο σημείο x^* ($\nabla^2 f(x^*)$), να είναι θετικά ορισμένος.

Το πρόβλημα της καθολικής βελτιστοποίησης είναι να βρεθεί το μικρότερο δυνατό τοπικό ελάχιστο, το οποίο ονομάζουμε καθολικό. Δηλαδή:

$$f^* = \min_{x \in D} f(x), f : D \subseteq \mathbb{R}^N \rightarrow \mathbb{R} \quad (1.2)$$

Οι παραπάνω συνθήκες για την εύρεση τοπικών ελαχίστων δεν αρκούν για την επιβεβαίωση ότι ένα σημείο αποτελεί και καθολικό ελάχιστο. Το πρόβλημα λοιπόν της καθολικής βελτιστοποίησης, όπως το ορίσαμε είναι άλυτο, αν χρησιμοποιήσουμε μια μέθοδο με πεπερασμένο αριθμό βημάτων. Μπορούμε όμως να λύσουμε προβλήματα καθολικής βελτιστοποίησης, όπου το σύνολο των μεταβλητών είναι πεπερασμένο (δεν είναι για παράδειγμα όλο το σύνολο των πραγματικών αριθμών).

Οι μέθοδοι εύρεσης τοπικού ελαχίστου, δεν μπορούν παρά να μας δώσουν προσεγγιστικές λύσεις. Έτσι αν μπορούμε να θεωρήσουμε ότι ένα πρόβλημα καθολικής ελαχιστοποίησης έχει λυθεί, αν για κάποιο $\epsilon > 0$, έχει βρεθεί ένα στοιχείο για κάποιο από τα ακόλουθα 3 σύνολα:

- 1.

$$A_{x(\epsilon)} = \{x \in S \mid \|x - x^*\| < \epsilon\} \quad (1.3)$$

, όπου $S \subset \mathbb{R}^n$ (το S είναι κυρτό σύνολο).

- 2.

$$A_{f(\epsilon)} = \{x \in S \mid |f(x) - f(x^*)| < \epsilon\} \quad (1.4)$$

Ένα άλλο μέτρο για την αξιοπιστία των αποτελεσμάτων των μεθόδων καθολικής ελαχιστοποίησης δίνεται από τον τύπο 4.6.

$$\phi(t) = m(\{y \in S \mid f(y) < t\})/m(S) \quad (1.5)$$

όπου m είναι το μέτρο Lebesgue (Lebesgue measure) και προκύπτει το σύνολο 4.7.

3.

$$A_{\varphi(\varepsilon)} = \{x \in S \mid \varphi(f(x)) < \varepsilon\} \quad (1.6)$$

1.2 Χαρακτηριστικά μεθόδων καθολικής ελαχιστοποίησης

Οι μέθοδοι καθολικής ελαχιστοποίησης επιχειρούν ολικές και τοπικές αναζητήσεις στις διάφορες περιοχές ενδιαφέροντος, προσπαθώντας να διερευνήσουν το μεγαλύτερο μέρος του πεδίου ορισμού των μεταβλητών. Μία μέθοδος καθολικής ελαχιστοποίησης έχει την ικανότητα να απεγλωβίζεται από τοπικά ελάχιστα με διαφόρους τρόπους, όπως θα δούμε και στη συνέχεια. Οι τοπικές μέθοδοι είναι πολύ χρήσιμες, λόγω της μεγάλης ακρίβειας, που προσφέρουν σε μικρές περιοχές ενδιαφέροντος. Μπορούν λοιπόν να επιστρέψουν με μεγάλη ακρίβεια την τιμή του τοπικού ελαχίστου σε τέτοιες περιοχές.

Γενικά μπορούμε να πούμε ότι οι μέθοδοι καθολικής ελαχιστοποίησης συνδιάζουν τα πλεονεκτήματα της τοπικής ελαχιστοποίησης με την σφαιρική διερεύνηση του πεδίου ορισμού των παραμέτρων. Πολλοί αλγόριθμοι ελαχιστοποίησης απαιτούν μεγάλο υπολογιστικό χρόνο. Έτσι, πολλές φορές η υπολογιστική πολυπλοκότητα κάποιων προβλημάτων φτάνει σε απαγορευτικά επίπεδα. Ωστόσο στις μέρες μας, οι εφαρμογές *παραλλήλων αλγορίθμων*, έχουν μειώσει σημαντικά τον απαιτούμε χρόνο για την εύρεση ικανοποιητικής λύσης, με αντίτιμο βέβαια τις αυξημένες ανάγκες σε εξοπλισμό.

Τα κύρια χαρακτηριστικά που συνθέτουν έναν αποδοτικό αλγόριθμο καθολικής ελαχιστοποίησης, και τον διαχωρίζουν από κάποιον αλγόριθμο εύρεσης τοπικού ελαχίστου είναι τα ακόλουθα:

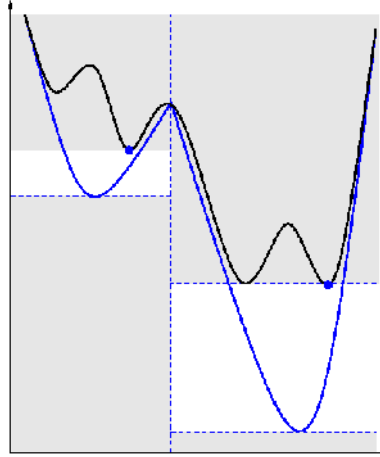
1. Αναπαράσταση του πεδίου αναζήτησης

Το πεδίο αναζήτησης ορίζεται ως το πεδίο ορισμού των παραμέτρων του προβλήματος. Σε κάποια προβλήματα μπορεί να είναι περιορισμένο ή άπειρο, γεγονός που επηρεάζει την υπολογιστική πολυπλοκότητα των αλγορίθμων αναζήτησης. Για παράδειγμα στο γνωστό πρόβλημα του *πλανόδιου πωλητή*, το πεδίο αναζήτησης είναι ένα σύνολο διακριτών τιμών και αποτελείται από το πλήθος των πόλεων. Σε προβλήματα βελτιστοποίησης συνεχών συναρτήσεων, το πεδίο αναζήτησης μπορεί να είναι είτε άπειρο είτε πεπερασμένο.

2. Στρατηγικές διαίρεσης (decomposition strategies)

Πολλοί αλγόριθμοι εφαρμόζονται σε όλο το πεδίο ορισμού των παραμέτρων, ενώ άλλοι διαμερίζουν το σύνολο αυτό σε μικρότερα υποσύνολα, όπου και εφαρμόζονται ξεχωριστά σε καθένα από αυτά (σχήμα 1.1).

Η γνωστή τεχνική διακλάδωσης και φράγματος (**branch & bound**), χρησιμοποιεί την παραπάνω φιλοσοφία, απορρίπτοντας υποσύνολα για περαιτέρω αναζήτηση και υποδιαίρωντας κάποια άλλα, που ενδεχομένως περιέχουν την λύση του προβλήματος. Με αυτή τη μεθοδολογία μπορούμε αν οδηγηθούμε σε ανάπτυξη παραλλήλων αλγορίθμων.



Σχήμα 1.1: Μέθοδος *branch and bound*

3. Πρόβλεψη της κατεύθυνσης διερεύνησης στο πεδίο ορισμού των παραμέτρων

Το θέμα της επιλογής κατάλληλης κατεύθυνσης αναζήτησης με τελικό στόχο το πραγματικό καθολικό ελάχιστο είναι πολύ σημαντικό, αφού η εξαντλητική αναζήτηση του πεδίου ορισμού σε προβλήματα συνεχών συναρτήσεων κρίνεται απαγορευτική. Σε μερικές μεθόδους, υπολογίζεται το σύνολο τιμών της συνάρτησης σε ένα συγκεκριμένο πεδίο ορισμού και προτιμούνται αυτές οι περιοχές, που το κάτω άκρο του διαστήματος τιμών είναι μικρότερο. Στη συνέχεια, το διάστημα αυτό υποδιαιρείται, ελέγχεται πάλι το κάτω άκρο και ούτο καθ'εξής. Σε άλλες περιπτώσεις η τιμή της συνάρτησης του γειτονικού σημείου παρέχει πληροφορία για την κατεύθυνση αναζήτησης. Στους γενετικούς αλγόριθμους, η τιμή καταλληλότητας κάθε μελους του πληθυσμού (*fitness score*) είναι καθοριστική στην επιλογή του για την επόμενη γενιά. Έτσι λοιπόν, η αναζήτηση κατευθύνεται προς εκείνα τα μέλη με την μεγαλύτερη τιμή καταλληλότητας.

4. Μηχανισμοί απεγκλωβισμού από τοπικά ελάχιστα

Για την εύρεση ενός καθολικού ελαχίστου, ο αλγόριθμος αναζήτησης πρέπει να έχει την δυνατότητα απεγκλωβισμού από τα διάφορα τοπικά ελάχιστα. Για παράδειγμα η μέθοδος *simulated annealing* δέχεται και σημεία που έχουν μεγαλύτερη τιμή συνάρτησης από την τρέχουσα ελάχιστη, ιδίως στα πρώτα βήματα του αλγορίθμου, ώστε να επιτευχθεί ευρεία αναζήτηση σε όλο το πεδίο ορισμού των παραμέτρων. Οι γενετικοί αλγόριθμοι ξεφεύγουν από τα διάφορα τοπικά ελάχιστα, συνδυάζοντας τα διάφορα μέλη του πληθυσμού με κάποια πιθανότητα. Έτσι προκύπτουν νέα σημεία που οι τιμές τους δεν έχουν καμία σχέση με τις προηγούμενες τιμές. Η ποικιλία λοιπόν που προστίθεται σε κάθε πληθυσμό, εκφράζει τον απεγκλωβισμό από τα διάφορα τοπικά ελάχιστα. Επίσης, οι υβριδικές μέθοδοι, που αποτελούνται από κάποια τοπική μέθοδο ελαχιστοποίησης κι έναν αλγόριθμο αναζήτησης, υπακούουν στην παραπάνω φιλοσοφία.

5. Κριτήρια τερματισμού.

Όλοι οι αλγόριθμοι θεωρούν ότι θα βρουν το καθολικό ελάχιστο μετά από άπειρες εφαρμογές τους. Αυτό όμως στην πραγματικότητα δεν είναι εφικτό. Ο αλγόριθμός μας θα πρέπει να τερματίζει σε εύλογο χρόνο, οπότε εισάγεται ένα *trade off* μεταξύ απόδοσης και εγγύησης εύρεσης καθολικού ελαχίστου. Γίνεται επιτακτική η ανάγκη, οι αλγόριθμοι εύρεσης ελαχίστου να τερματίζουν μετά από κάποιο χρονικό διάστημα, δύνοντας λύσεις ικανοποιητικής προσέγγισης. Ο βαθμός προσέγγισης της πραγματικής λύσης του ολικού ελαχίστου είναι τις περισσότερες φορές ανάλογος με τον χρόνο εκτέλεσης του αλγορίθμου. Τα γενικότερα κριτήρια τερματισμού είναι:

- i. Ένα καθορισμένο πλήθος αποτιμήσεων της συνάρτησης.
- ii. Εσαγωγή προκαθορισμένου αριθμού βημάτων του αλγορίθμου.

1.3 Κατηγορίες μεθόδων καθολικής ελαχιστοποίησης

Οι μέθοδοι που έχουν αναπτυχθεί για την εύρεση των τοπικών ελαχίστων διακρίνονται σε 2 μεγάλες κατηγορίες:

- **Αιτιοκρατικές (Deterministic) μέθοδοι**
- **Στοχαστικές (Stochastic) μέθοδοι**

Οι κατηγορίες αυτές διακρίνται ανάλογα με το αν χρησιμοποιούμε ή όχι στοχαστικά στοιχεία.

1.3.1 Αιτιοκρατικές (Deterministic) μέθοδοι

Οι μέθοδοι που ανήκουν σε αυτή την κατηγορία, βρίσκουν το καθολικό ελάχιστο, εφαρμόζοντας μια εξαντλητική αναζήτηση σε όλο το πεδίο ορισμού των μεταβλητών της συνάρτησης. Τα αποτελέσματα είναι εγγυημένα, αλλά για να εφαρμοστούν πρέπει η συνάρτηση ελαχιστοποίησης να πληρεί κάποιες προϋποθέσεις. Η πιο δημοφιλής είναι η συνθήκη **Lipschitz**(1.7)

Για κάποιο σταθερό L , για όλα τα x και x^* να ισχύει

$$|f(x) - f(x^*)| \leq L \|x - x^*\| \quad (1.7)$$

Δυστυχώς όμως η εξέταση του παραπάνω κριτηρίου σε επίπεδο υλοποίησης δεν είναι πάντα εύκολη υπόθεση κι εφικτή υπόθεση. Επιπρόσθετα, οι μέθοδοι αυτές καταναλώνουν μεγάλο υπολογιστικό χρόνο μέχρι την ικανοποίηση κάποιου κριτηρίου τερματισμού. Συνήθως, παρατηρείται εκθετική αύξηση του χρόνου με την αύξηση του πλήθους των μεταβλητών του προβλήματος. Οι μέθοδοι διαστημάτων (interval) είναι οι πιο δημοφιλείς εκπρόσωποι αυτής της κατηγορίας μεθόδων. Χρησιμοποιούν στοιχεία της αριθμητικής διαστημάτων για τον υπολογισμό των διαστημάτων των τιμών της συνάρτησης, σε συγκεκριμένα διαστήματα των μεταβλητών της.

1.3.2 Στοχαστικές μέθοδοι

Οι στοχαστικές μέθοδοι, περιλαμβάνουν συνήθως δύο μεγάλες φάσεις:

1. Η **ολική φάση (Global phase)**, όπου η συνάρτηση αποτιμάται, σε έναν τυχαίο αριθμό σημείων
2. Η **τοπική φάση (local phase)**, όπου κάποια από τα σημεία του αρχικού δείγματος αποτελούν αρχικά σημεία για μεθόδους τοπικής ελαχιστοποίησης.

1.3.3 Σύγκριση των μεθόδων

Το μεγάλο μειονέκτημα των στοχαστικών μεθόδων σε σύγκριση με τις αιτιοκρατικές είναι ότι ανακαλύπτουν το καθολικό ελάχιστο με κάποια πιθανότητα μέσα από έναν πεπερασμένο αριθμό επαναλήψεων. Αυτό σημαίνει ότι εφόσον σε πρακτικές εφαρμογές το ασυμπτωτικό όριο (άπειρες επαναλήψεις) δεν είναι εφικτό, θα πρέπει να τερματιστεί η διαδικασία πρόωμα, διακινδυνεύοντας ωστόσο την πλήρη επιτυχία της. Παρ'όλα αυτά υπό κάποιες προϋποθέσεις, που αφορούν την κατανομή των σημείων του αρχικού δείγματος και την συνάρτηση, η πιθανότητα να έχουμε εγγυημένες λύσεις του προβλήματος πλησιάζει τη μονάδα, όσο αυξάνεται το πλήθος των σημείων του αρχικού δείγματος. Επιπλέον παρατηρούμε, ότι μια μέθοδος, η οποία δεν περιλαμβάνει την τοπική φάση, είναι λιγότερη αποδοτική από κάποια άλλη που την περιλαμβάνει. Όπως και στην περίπτωση των αιτιοκρατικών μεθόδων το μεγάλο πρόβλημα που τίθεται, είναι η εύρεση κάποιου κατάλληλου κριτηρίου τερματισμού. Οι στοχαστικές μέθοδοι, τερματίζουν κυρίως με ένα κριτήριο που πιστοποιεί σε μεγάλο βαθμό ότι η πιθανότητα για την ορθότητα των αποτελεσμάτων πλησιάζει τη μονάδα. Υπάρχουν βέβαια και ειδικά κριτήρια τερματισμού, που σχετίζονται με την ιδιαίτερη φύση των μεθόδων και είναι κατάλληλα μόνο για αυτές.

ΚΕΦΑΛΑΙΟ 2

ΣΤΟΧΑΣΤΙΚΕΣ ΜΕΘΟΔΟΙ ΚΑΙ ΚΑΘΟΛΙΚΗ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗ

Στο παρόν κεφάλαιο θα αναφερθούμε επιγραμματικά στις πιο γνωστές στοχαστικές μεθόδους και θα επιχειρήσουμε μια θεωρητική μελέτη, ούτως ώστε να διαφανεί η ποιότητα, η ορθότητα και η αξιοπιστία των αποτελεσμάτων τους. Οι μεθοδολογίες που αναφέρονται είναι από τις βασικότερες που έχουν προταθεί.

2.1 Τυχαία αναζήτηση (Random Search)

Η τυχαία αναζήτηση είναι η πιο απλή μέθοδος ελαχιστοποίησης. Λαμβάνεται ένα αρχικό δείγμα σημείων και απλά αποτιμάται η τιμή της συνάρτησης f στα σημεία αυτά. Η επιλογή των σημείων γίνεται με βάση μια ομοιόμορφη κατανομή και έχει αποδειχθεί ότι όσο αυξάνεται το πλήθος του δείγματος, τόσο αυξάνεται και η πιθανότητα να βρούμε το καθολικό ελάχιστο. Ωστόσο μια τέτοια αύξηση επιφέρει και αύξηση του χρόνου που απαιτείται από τη μέθοδο για να μας δώσει ικανοποιητικά αποτελέσματα. Σήμερα η επιλογή αυτών των μεθόδων ενδείκνυται μόνο σε συναρτήσεις με λίγα ολικά ελάχιστα. Ενδιαφέρον παρουσιάζουν οι μέθοδοι της *ελεγχόμενης τυχαίας αναζήτησης* (*controlled random search*), όπου το νέο σημείο στο οποίο αποτιμάται η συνάρτηση, προκύπτει μελετώντας τη γνώση που έχουμε για τις τιμές της συνάρτησης από σημεία, που έχουμε ήδη εξετάσει. Ο πιο δημοφιλής εκπρόσωπος αυτής της κατηγορίας είναι ο αλγόριθμος του Price[11].

2.2 Πολλαπλή εκκίνηση (Multistart)

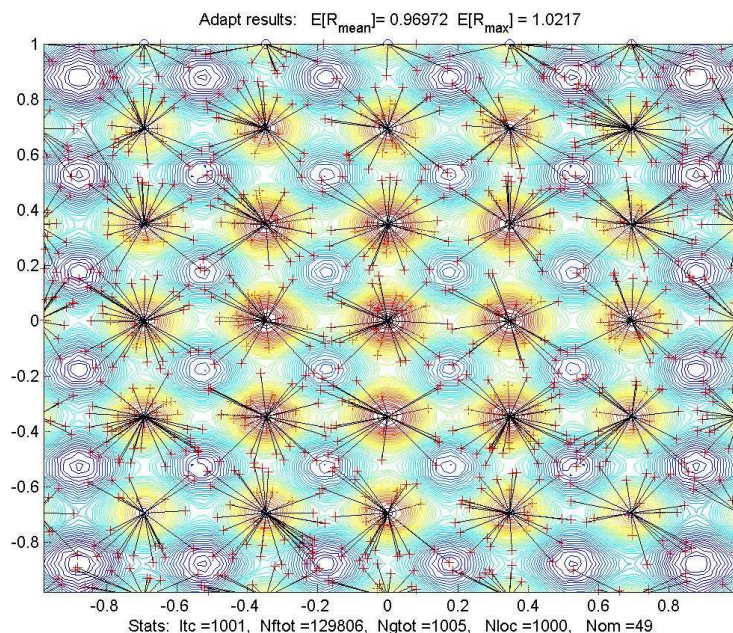
Λόγω των σημαντικών μειονεκτημάτων που παρουσιάζει η προηγούμενη μέθοδος, έχουν προταθεί διάφορες επεκτάσεις. Η μέθοδος που θα περιγράψουμε, αρχίζει με ένα τυχαίο αρχικό σύνολο σημείων, έστω S (ομοιόμορφα κατανεμημένο), περιλαμβάνει όμως και διάφορες τοπικές αναζητήσεις. Αυτό προϋποθέτει κάποια μέθοδο τοπικής ελαχιστοποίησης, η οποία αρχίζοντας από ένα αρχικό σημείο $x \in S$, επιστρέφει ένα τοπικό ελάχιστο x^* στην

γειτονιά του x . Ανάλογα με τη φύση της συνάρτησης επιλέγουμε την κατάλληλη μέθοδο τοπικής ελαχιστοποίησης. Η πιο απλή μέθοδος που συνδυάζει και τις δύο φάσεις που περιγράψαμε παραπάνω είναι η μέθοδος **πολλαπλής εκκίνησης (MULTISTART)**. Εδώ η μέθοδος τοπικής ελαχιστοποίησης εφαρμόζεται σε όλα τα σημεία του δείγματος, και το **μικρότερο από τα τοπικά ελάχιστα** που επιστρέφονται μετά από διαδοχικές κλήσεις, αποτελεί το καθολικό ελάχιστο.

Ακόμα κι αν αυτή η μέθοδος είναι πιο αποδοτική από την τυχαία αναζήτηση, υπάρχουν αρκετά μειονεκτήματα. Το κριότερο αφορά της επιλογής του καταλληλότερου κριτηρίου τερματισμού. Το πρόβλημα αντιμετωπίζεται μέσω μία στοχαστικής προσέγγισης, η οποία βασίζεται σε μία *Μπευζιανή μελέτη* που αφορά τον αριθμό των τοπικών ελαχίστων και το μέγεθος κάθε διαφορετικής **περιοχής έλξης (region of attraction)**. Η περιοχή $R(x^*)$ ορίζεται ως το σύνολο όλων των σημείων του S , τα οποία αν αποτελούν τα αρχικά σημεία για μία μέθοδο ελαχιστοποίησης P , θα οδηγήσουν στο ίδιο τοπικό ελάχιστο x^* .

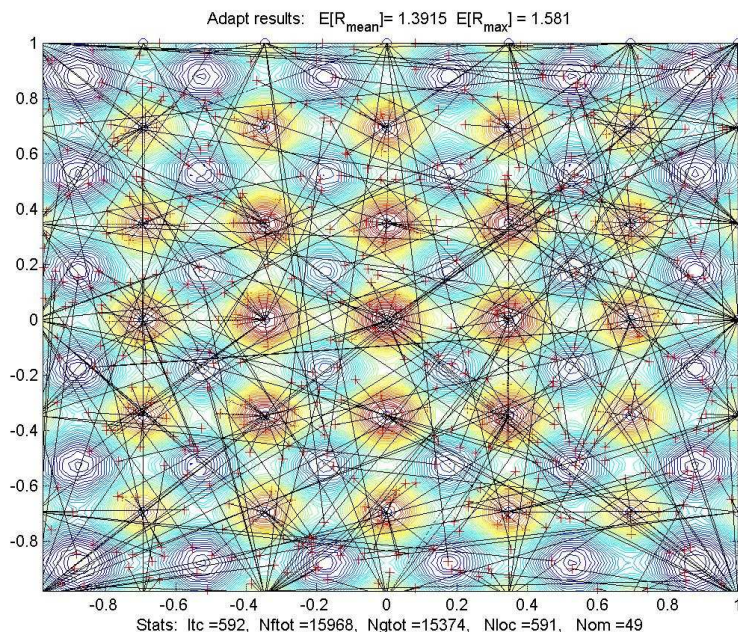
Μια ιδιότητα που θέλουμε να έχει η μέθοδος τοπικής ελαχιστοποίησης προκειμένου οι περιοχές έλξης να είναι συνεχόμενες, είναι να είναι πραγματικά τοπική. Ουσιαστικά εννοούμε να είναι σαν την Gradient descent με απειροελάχιστο βήμα, δηλαδή να ακολουθούμε την κατεύθυνση της παραγώγου με πολύ μικρό βήμα. (βλέπε 2.1 και 2.2)

Στην πρώτη εικόνα η τοπική ελαχιστοποίηση δεν εμφανίζει άλματα και έτσι η περιοχές έλξης είναι ευδιάκριτες. Κάτι τέτοιο δεν γίνεται και στη δεύτερη περίπτωση όπου ξεκινώντας από ένα σημείο του πεδίου ορισμού καταλήγουμε στην άλλη άκρη, με αποτέλεσμα να παρακάμπουμε γειτονικά ελάχιστα. Αν γνωρίζουμε το πλήθος των τοπικών ελαχίστων της



Σχήμα 2.1: Παράδειγμα τοπικής αναζήτησης

συνάρτησης που θέλουμε να ελαχιστοποιήσουμε, καθώς και το μέγεθος κάθε περιοχής ενδιαφέροντος (δηλαδή το μέτρο της ακτίνας αυτής της περιοχής, με κέντρο το τοπικό ελάχιστο),



Σχήμα 2.2: Παράδειγμα μη τοπικής αναζήτησης

τότε μπορούμε να αναλύσουμε ποιοτικά τα επιστρεφόμενα αποτελέσματα. Δυστυχώς όμως η γνώση αυτών των παραμέτρων δεν είναι δυνατόν να υπάρχει εκ των προτέρων. Θεωρούμε λοιπόν τυχαίες κατανομές για τα σημεία του αρχικού δείγματος και για τα σημεία που θα ανήκουν στην ίδια περιοχή ενδιαφέροντος (εκ των προτέρων πιθανότητα). Στη συνέχεια με βάση τα αποτελέσματα της μεθόδου και χρησιμοποιώντας τον κανόνα του *Bayes* υπολογίζουμε την εκ των υστέρων πιθανότητα, η επόμενη τοπική αναζήτηση να οδηγήσει σε τοπικό ελάχιστο που έχουμε ήδη υπολογίσει και την εκ των υστέρων πιθανότητα να έχουμε ήδη βρει όλα τα τοπικά ελάχιστα της συνάρτησης. Περισσότερα στοιχεία για την θεωρητική μελέτη των αποτελεσμάτων της μεθόδου πολλαπλής εκίνησης, αναπτύσσονται από τον Rinnooy Kann[3],[4].

Παρά το γεγονός όμως ότι η θεωρητική μελέτη της μεθόδου πολλαπλής εκίνησης είναι εφικτή, η παραπάνω μέθοδος δεν μπορεί να θεωρηθεί αποδοτική. Η κύρια αιτία είναι ότι λόγω της φύσης της αναζήτησης τοπικού ελαχιστού, καταλήγουμε πολλές φορές να υπολογίζουμε σημεία που έχουμε ήδη βρει.

2.3 Μέθοδοι Ομαδοποίησης (Clustering)

Οι μέθοδοι ομαδοποίησης [4] είναι από τους πιο αποτελεσματικούς κι αποδοτικούς αλγόριθμους, που έχουν προταθεί για καθολική ελαχιστοποίηση. Η βασική ιδέα πίσω από αυτές τις μεθόδους είναι, ότι αρχίζοντας από ένα δείγμα ομοιόμορφα κατανομημένων σημείων, δημιουργούμε ομάδες από σημεία, που ανήκουν στην ίδια περιοχή έλξης, έτσι ώστε να εφαρμόζεται η μέθοδος τοπικής ελαχιστοποίησης μόνο μία φορά σε μία τέτοια περιοχή.

Η διαδικασία της ομαδοποίησης γίνεται με διάφορους τρόπους. Η βασική ιδέα όμως παρα-

μένει η ίδια. Σε κάθε ομάδα (cluster) μπορεί να υπάρχει ένα σημείο πυρήνας (seed point). Ως πυρήνας μπορεί να θεωρηθεί το σημείο εκείνο του δείγματος που έχει τη χαμηλότερη τιμή συνάρτησης, ή, ακόμα καλύτερα, το τοπικό ελάχιστο που λάβαμε αρχίζοντας τοπική ελαχιστοποίηση από εκείνο το σημείο. Τα υπόλοιπα σημεία προστίθενται στους διάφορους clusters με βάση κάποιο κανόνα για την ομαδοποίηση, μέχρι να ικανοποιηθεί κάποιο κριτήριο τερματισμού. Οι περισσότερες μέθοδοι διαφέρουν στην επιλογή αυτού του κανόνα. Οι *Rinnoy Kan* και *Timmer* [4] περιγράφουν μια μέθοδο ομαδοποίησης κι επίσης περιγράφουν συνθήκες για τις οποίες όλα τα τοπικά ελάχιστα θα βρεθούν μετά από ένα πεπερασμένο αριθμό επαναλήψεων, με πιθανότητα ένα. Το σημαντικό μειονέκτημα αυτών των μεθόδων είναι ότι δεν παρουσιάζουν καλή απόδοση στην περίπτωση συναρτήσεων με πάρα πολλά τοπικά ελάχιστα. Για αυτές τις συναρτήσεις χρειάζονται δείγματα πολλών σημείων, ώστε να βρεθούν με αυξημένη πιθανότητα αυτά τα τοπικά ελάχιστα.

2.4 Simulated annealing

Η μέθοδος αυτή [8],[9],[10], έχει εμπνευστεί από τη γνωστή διαδικασία του annealing στην φυσική για την απόκτηση καταστάσεων χαμηλής ενέργειας κι από τις διαδικασίες εύρεσης ελαχίστων λύσεων σε προβλήματα διακριτής βελτιστοποίησης. Σε γενικές γραμμές υπολογίζεται σε κάποιο σημείο x' η τιμή της συνάρτησης και θεωρώντας ένα γειτονικό σημείο x'' κάνουμε τον εξής έλεγχο:

Αν ισχύει

$$f(x'') - f(x') < 0 \quad (2.1)$$

τότε δεχόμαστε το καινούργιο σημείο. Αλλιώς το νέο σημείο το δεχόμαστε με πιθανότητα

$$e^{-(f(x'')-f(x'))/T} \quad (2.2)$$

Η θετική παράμετρος T είναι η *θερμοκρασία* η οποία μειώνεται με σταθερό ρυθμό, όπως γίνεται στην διαδικασία του annealing. Αρχικά, σε κατάσταση υψηλής θερμοκρασίας, η πιθανότητα να δεχτούμε σημεία, στα οποία η τιμή της συνάρτησης αυξάνεται, είναι αρκετά μεγάλη. Το γεγονός αυτό κάνει την παραπάνω διαδικασία να έχει χαρακτήρα καθολικής ελαχιστοποίησης. Όσο η θερμοκρασία μειώνεται, η αναζήτηση περιορίζεται σε πολύ μικρές περιοχές του πεδίου ορισμού των μεταβλητών. Σε πολύ χαμηλές θερμοκρασίες η αναζήτηση περιορίζεται σε μια πολύ μικρή περιοχή, όπου είναι πλέον αδύνατον να απεγκλωβιστούμε από κάποιο τοπικό ελάχιστο. Είναι φανερό ότι όσο μειώνεται η θερμοκρασία, τόσο καλύτερα προσεγγίζουμε το καθολικό ελάχιστο, οπότε ο εγκλωβισμός στην τελική φάση είναι επιθυμητός.

2.5 Γενετικού αλγόριθμοι

Οι γενετικοί αλγόριθμοι [7] προέρχονται από γενικότερη ιδέα της εξέλιξης των βιολογικών μοντέλων. Οι γενετικοί αλγόριθμοι διαχειρίζονται έναν αρχικό πληθυσμό από ενδεχόμενες λύσεις κι εφαρμόζοντας την αρχή της επιβίωσης του πιο ισχυρού, παράγουν καλύτερες προσεγγίσεις λύσεων, όσο συνεχίζεται η δημιουργία καινούργιων γενεών. Σε κάθε γενεά, ο παραγόμενος πληθυσμός, προκύπτει από τον προηγούμενο είτε μέσω της επιλογής (επιβίωση του καταλληλότερου μέλους του πληθυσμού) είτε με την εφαρμογή διαφόρων γενετικών τελεστών (όπως *μετάλλαξη και διασταύρωση*) εμπνευσμένοι από τον χώρο της βιολογικής εξέλιξης. Αυτό έχει ως αποτέλεσμα τα μέλη του πληθυσμού της επόμενης γενεάς να είναι καλύτερα σε σύγκριση με τους προγόνους τους.

2.6 Αναζήτηση με απαγορεύσεις (Tabu Search)

Η μέθοδος αυτή κρίνεται αρκετά επιτυχημένη για συγκεκριμένα προβλήματα της καθολικής βελτιστοποίησης (όπως το πρόβλημα του πλανόδιου πωλητή - *Travelling salesman problem*). Η βασική ιδέα της μεθόδου είναι η διερεύνηση του πεδίου ορισμού των παραμέτρων της συνάρτησης, μέσω μίας ακολουθίας κινήσεων. Η κίνηση που επιλέγεται από κάποιο σημείο σε κάποιο άλλο, είναι εκείνη που έχει αποδειχτεί από τα προηγούμενα βήματα, ότι μπορεί να οδηγήσει σε χαμηλότερες τιμές της συνάρτησης. Επίσης κατά την επιλογή αυτών των κινήσεων, κάποιες από αυτές θεωρούνται **απαγορευμένες (tabu)**, ώστε να αποφεύγονται κύκλοι κι εγκλωβισμός σε τοπικά ελάχιστα. Κατά την εξέλιξη του αλγορίθμου κρατείται ένα ιστορικό με τις έως τώρα κινήσεις και τις λύσεις στις οποίες οδηγηθήκαμε με βάση αυτές. Σύμφωνα λοιπόν με αυτό το ιστορικό, μία κίνηση μπορεί να χαρακτηριστεί ως **απαγορευμένες (tabu)**, αν έχει γίνει αρκετές φορές ή σχετικά πρόσφατα. Επίσης για να διασφαλίσουμε ότι δεν περιοριζόμαστε σε ένα περιορισμένο αριθμό κινήσεων, σε κάποια στάδια του αλγορίθμου δεχόμαστε και κινήσεις που έχουν χαρακτηριστεί ως **απαγορευμένες** [5],[6].

ΚΕΦΑΛΑΙΟ 3

ΚΡΙΤΗΡΙΑ ΤΕΡΜΑΤΙΣΜΟΥ ΓΙΑ ΣΤΟΧΑΣΤΙΚΕΣ ΜΕΘΟΔΟΥΣ MULTISTART

Ένα από τα κυριότερα θέματα των στοχαστικών μεθόδων είναι τα κριτήρια τερματισμού τους. Επειδή οι μέθοδοι αυτής της κατηγορίας ανακαλύπτουν το καθολικό ελάχιστο με κάποια πιθανότητα μέσα από έναν πεπερασμένο αριθμό επαναλήψεων δεν είναι εγγυημένη η επιτυχία της εφαρμογής τους. Αυτό σημαίνει ότι εφόσον σε πρακτικές εφαρμογές το ασυμπτωτικό όριο (άπειρες επαναλήψεις) δεν είναι εφικτό, θα πρέπει να τερματιστεί η διαδικασία πρώιμα, διακινδυνεύοντας ωστόσο την πλήρη επιτυχία της.

Έχουν αναπτυχθεί διάφοροι "κανόνες τερματισμού" που στοχεύουν στο να διατηρούν μεγάλη πιθανότητα επιτυχούς αποτελέσματος και ταυτόχρονα να μην απαιτούν υψηλό υπολογιστικό κόστος. Στην βιβλιογραφία έχουν εμφανιστεί αρκετοί κανόνες όπως τον Zieliński [12], Betro et al [13], Rinnooy Kan [14], Hart[16]. Μια βιβλιογραφική επισκόπηση των κανόνων τερματισμού δίνεται στο [15].

Στην επόμενη παράγραφο γίνεται μια σύντομη αναφορά στα κριτήρια των Boender, Rinnooy Kan και Zieliński και στην συνέχεια περιγράφουμε τρεις νέους κανόνες, ήτοι τον κανόνα του "διπλάσιου κουτιού" (double box), τον κανόνα "των παρατηρήσιμων μεγεθών" και τέλος "των αναμενόμενων ελαχίστων".

3.1 Τα κριτήρια των Rinnooy Kan, Boender και Zieliński

Σύμφωνα με το κριτήριο τερματισμού που διατυπώθηκε από τον Kan [17], αν για την έρεση w ελαχίστων έχουν γίνει t τοπικές αναζητήσεις, τότε μια εκτίμηση για το πλήθος των τοπικών ελαχίστων της συναρτήσεως δίνεται από τον τύπο:

$$\hat{w} = \frac{w(t-1)}{t-w-2} \quad (3.1)$$

Σε πολυδειγματικές μεθόδους καθολικής ελαχιστοποίησης όπως η Topographical Multi-level Single Linkage Clustering [18], το t στην παραπάνω εξίσωση αντικαθίσταται από το σύνολο των δειγματοληψιών που έχει πραγματοποιήσει ο αλγόριθμος. Ο αλγόριθμος ελαχιστοποίησης διακόπτεται όταν:

$$\hat{w} - w < \frac{1}{2} \quad (3.2)$$

Ωστόσο κάνοντας πράξεις στην ανισότητα διαπιστώνεται πως πρέπει αν ισχύει ότι:

$$t > 2w^2 + 3w + 2 \quad (3.3)$$

Με άλλα λόγια θα πρέπει ο αριθμός των τοπικών ελαχιστοποιήσεων (ή δειγμάτων σε άλλες περιπτώσεις) να είναι ανάλογος του τετραγώνου των ελαχίστων που έχουν βρεθεί, προκειμένου να σταματήσει ο αλγόριθμος. Αν έχουν βρεθεί πχ. 1000 ελάχιστα θα απαιτηθούν 10^6 επιπλέον σημεία κατά την διάρκεια των πρόσθετων τοπικών ελαχιστοποιήσεων. Ο κανόνας αυτός δεν είναι ιδιαίτερα αποδοτικός και επίσης δεν εκμεταλλεύεται την γνώση που αποκτά από την δειγματοληψία. Ο Zieliński [12] εκτιμά το ποσοστό του χώρου αναζήτησεως που δεν έχει καλυφθεί με την δειγματοληψία ως:

$$P(w) = \frac{w(w+1)}{t(t-1)} \quad (3.4)$$

Σ'αυτήν την περίπτωση η μέθοδος καθολικής ελαχιστοποίησης διακόπτεται όταν:

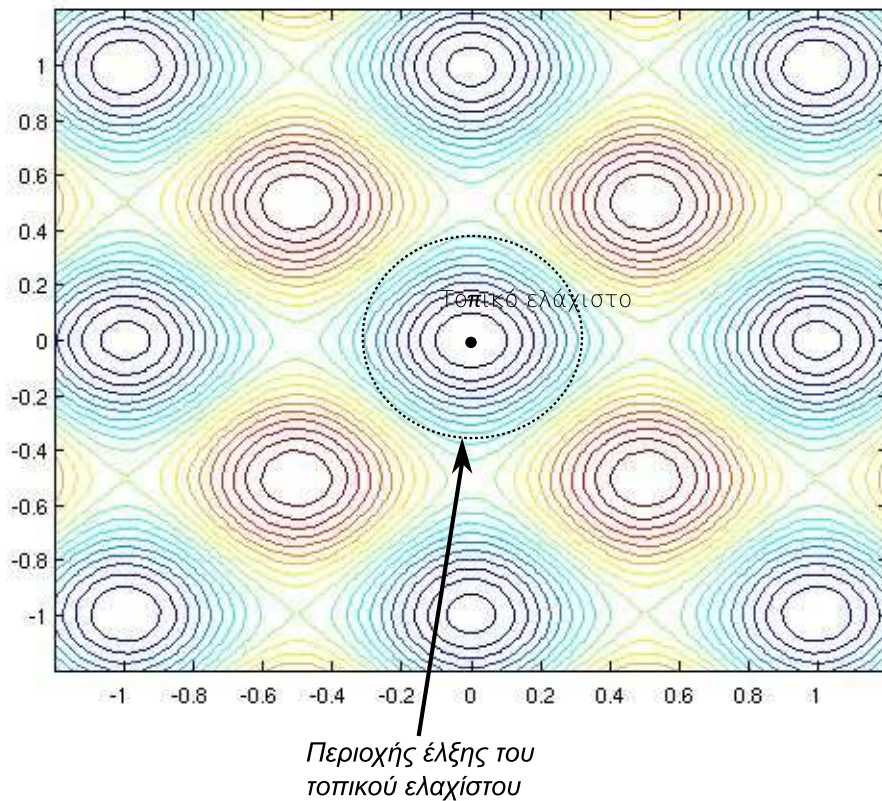
$$P(w) \leq \varepsilon \quad (3.5)$$

όπου ε ένας μικρός θετικός αριθμός, πχ. $\varepsilon = 10^{-4}$.

3.2 Ο κανόνας του διπλού κουτιού (double box)

Προσπαθούμε μέσα τις πληροφορίες που εξάγουμε από τα τοπικά ελάχιστα να το "δημιουργήσουμε" το κριτήριο τερματισμού. Μία trivial λύση αντιμετώπισης του προβλήματος είναι να τερματίζουμε τον αλγόριθμο όταν θα έχουμε βρει όλα τα τοπικά ελάχιστα της συνάρτησης. Μία τέτοια πληροφορία όμως δεν είναι δυνατόν να την έχουμε εκ των προτέρων. Έτσι προσπαθούμε με διαφορετικό τρόπο να εκμεταλευτούμε την πληροφορία που μας δίνει κάθε τοπικό ελάχιστο που βρίσκουμε κατά τη διάρκεια της αναζήτησης.

Έστω ότι έχουμε μια συνάρτηση f , ένα σημείο x και μία μέθοδο ελαχιστοποίησης E και καταλήγω στο τοπικό ελάχιστο y^* . Ορίζουμε το σύνολο $A(y) = \{x | L(x) = y\}$ που ονομάζεται περιοχής έλξης. Όπως βλέπουμε και από την εικόνα 4.3, τα σύνολα $A(y_i)$ είναι



Σχήμα 3.1: Περιοχή έλξης συνάρτησης *Rastrigin*

ανοιχτά σύνολα και στα άκρα έχω τοπικά μέγιστα. Εύλογα μπορούμε να παρατηρήσουμε ότι κάθε $A(y)$ εξαρτάται i) από το εύρος του y , ii) από τη συνάρτηση f και iii) από τη συνάρτηση ελαχιστοποίησης. Επίσης βλέπουμε ότι οι περιοχές $A(y_i)$ είναι ξένες μεταξύ τους, δηλαδή δεν έχουν σημεία τομής. Τέλος οι περιοχές έλξης έχουν ένα μέτρο που ανάλογα με τη μορφή του σύνολο τιμών διαφέρει. Έτσι σε μία διάσταση το μέτρο είναι το μήκος, στις 2 είναι το εμβαδόν, ενώ στις 3 είναι ο όγκος.

Έστω ότι θέλω να εφαρμόσω τη μέθοδο *multistart* σε μια συνάρτηση που έχει k τοπικά ελάχιστα. και έστω ότι επιλέγω το ελάχιστο που ανήκει στην περιοχή A_4 , δηλαδή το y_4 . Το ερώτημα που τίθεται σε αυτό το σημείο είναι ποια θα είναι η πιθανότητα να βρω το y_4 ξεκινώντας από ένα τυχαίο σημείο. Με την προϋπόθεση ότι τα σημεία επιλέγονται μέσω ομοιόμορφης κατανομής, η πιθανότητα να βρω το y_4 την πρώτη φορά είναι:

$$P(y_4) = \frac{m(A_4)}{m(S)} \quad (3.6)$$

άρα η πιθανότητα να μην το βρω μετά από μια προσπάθεια είναι

$$1 - P(y_4), \quad (3.7)$$

ενώ η πιθανότητα να μην το βρω μετά από 2 προσπάθειες είναι:

$$(1 - P(y_4))^2 \quad (3.8)$$

Τελικά μετά από k προσπάθειες η πιθανότητα να μην βρω το ελάχιστο θα είναι:

$$(1 - P(y_4))^k, \quad (3.9)$$

οπότε η πιθανότητα να το βρω μετά από k προσπάθειες θα είναι:

$$1 - (1 - P(y_4))^k \quad (3.10)$$

Ακόμα και μετά από αρκετές προσπάθειες για την εύρεση του minimum ($k \gg 1$) βλέπουμε ότι:

$$1 - (1 - P(y_4))^k = 1 - P(y_4)^k \xrightarrow{k \rightarrow \infty} 1, \quad (3.11)$$

δηλαδή η πιθανότητα να βρω το y_4 , είναι κοντά στο 1. Συνεπώς ασυμπτωτικά θα βρω το τοπικό ελάχιστο.. Για κάθε τοπικό ελάχιστο, η αντίστοιχη πιθανότητα ισούται με:

$$P_i = \frac{m(A_i)}{m(S)} \quad (3.12)$$

Ορίζω την ποσότητα

$$\hat{U} = 1 - \sum_{i=1}^w \frac{m(A_i)}{m(S)} \quad (3.13)$$

όπου w είναι ο αριθμός των ελαχίστων που έχω βρει. Στα πειράματα που πραγματοποιούνται είναι ανάγκη να γνωρίζουμε το τμήμα του χώρου αναζητήσεως που δεν έχει καλυφθεί από τις τοπικές ελαχιστοποιήσεις που έχουν πραγματοποιηθεί. Αν το τμήμα αυτό είναι μικρό, τότε η μέθοδος καθολικής ελαχιστοποίησης μπορεί να τερματιστεί. Το ποσοστό της περιοχής αναζητήσεως που δεν έχει καλυφθεί δίνεται από την ποσότητα:

$$\hat{U} = 1 - \sum_{i=1}^w \frac{m(A_i)}{m(S)} \quad (3.14)$$

Στην παραπάνω εξίσωση, η ποσότητα $\frac{m(A_i)}{m(S)}$ προσεγγίζεται από την ποσότητα $\frac{t_i}{T}$, όπου t_i είναι το πλήθος των εφαρμογών της μεθόδου ελαχιστοποίησης που οδήγησαν στο τοπικό ελάχιστο y_i^* και T είναι ο συνολικός αριθμός εφαρμογών της τοπικής ελαχιστοποίησης. Δυστυχώς σε κάθε επανάληψη η ποσότητα \hat{U} έχει την τιμή 0 μιας και

$$\sum_{i=1}^w t_i = T$$

και σαν συνέπεια η παραπάνω εξίσωση δεν μπορεί να χρησιμοποιηθεί για την προσέγγιση του τμήματος του χώρου αναζητήσεως που δεν έχει ακόμα καλυφθεί. Για τον σκοπό αυτόν, κατασκευάζεται ένας μεγαλύτερος χώρος αναζητήσεως S_2 που περιέχει τον αρχικό χώρο S , έτσι ώστε να ισχύει $m(S_2) = 2m(S)$. Επίσης ένα και μοναδικό (και ψεύτικο) ελάχιστο θεωρείται πως υπάρχει στην περιοχή $A_0 = S_2 - S$. Ο όγκος της περιοχής έλξης για αυτό το ελάχιστο είναι $m(A_0) = m(S)$.

$$U = 1 - \sum_{i=1}^n \frac{m(A_i)}{m(S)} \Rightarrow U = 1 - \sum_{i=1}^n \frac{\frac{1}{2}m(A_i)}{m(S)} \Rightarrow U = 1 - \sum_{i=1}^n \frac{2m(A_i)}{m(S)} \Rightarrow U = 1 - 2 \sum_{i=1}^n \frac{m(A_i)}{m(S)} \quad (3.15)$$

Επομένως η ποσότητα \hat{U} προσεγγίζεται ως:

$$\hat{U} = 1 - 2 \sum_{i=1}^w \frac{t_i}{T} \quad (3.16)$$

Σε κάθε επανάληψη, N σημεία συλλέγονται ομοιόμορφα τυχαία στον χώρο S_2 . Έστω ότι μετά από k επαναλήψεις ο συνολικός αριθμός των σημείων που έχουν ληφθεί στον χώρο S_2 είναι M_k και από αυτά τα k_N βρίσκονται στον χώρο S . Η ποσότητα $\delta_k = \frac{k_N}{M_k}$, έχει μια μέση τιμή

$$\langle \delta \rangle_k = \frac{1}{k} \sum_{i=1}^k \delta_k$$

, η οποία τείνει στην ποσότητα $\frac{m(S)}{m(S_2)} = \frac{1}{2}$ καθώς ο μετρητής επαναλήψεων k τείνει στο άπειρο. Η διακύμανση της ίδιας ποσότητας δίνεται από τον τύπο:

$$\sigma_k^2(\delta) = \langle \delta^2 \rangle_k - \langle \delta \rangle_k^2 \quad (3.17)$$

η οποία τείνει στο 0 καθώς το k τείνει στο άπειρο. Η ποσότητα αυτή είναι περισσότερο ομαλά μεταβαλλόμενη από την μέση τιμή και επομένως πιο κατάλληλη για να χρησιμοποιηθεί για τον σχηματισμό ενός κριτηρίου τερματισμού. Η καθολική μέθοδος ελαχιστοποίησης όταν ισχύουν οι συνθήκες:

$$\left| \langle \delta \rangle - \frac{1}{2} \right| \leq \sigma_k(\delta), \sigma_k(\delta) \leq p \sigma_{last}^2(\delta) \quad (3.18)$$

όπου $\sigma_{last}^2(\delta)$ η τιμή της διακύμανσης της ποσότητας δ_k στην επανάληψη που βρέθηκε το τελευταίο τοπικό ελάχιστο και p είναι μια ποσότητα στο διάστημα $(0, 1)$ και χρησιμοποιείται σαν ένας συμβιβασμός ανάμεσα σε μια εξαντλητική αναζήτηση ($p \rightarrow 0$) και σε μία αναζήτηση βελτιστοποιημένη για ταχύτητα ($p \rightarrow 1$). Η προτεινόμενη λύση για αυτήν την παράμετρο είναι η τιμή 0.5. Ο αλγόριθμος έχει ως εξής:

- 1 Αρχικά θέτουμε $\alpha = 0$
- 2 Δειγματοληψία N σημείων με την διαδικασία που διατυπώθηκε προηγουμένως.
- 3 Υπολογισμός της ποσότητας $\sigma^2(\delta)$.
- 4 Εκτέλεση μιας επαναλήψεως της διαδικασίας καθολικής ελαχιστοποίησης χρησιμοποιώντας τα σημεία που ελήφθησαν προηγουμένως. Ανάλογα με τη χρησιμοποιούμενη μέθοδο (Multistart, TMLSL, κτλ.) αυτά τα σημεία μπορούν να χρησιμοποιηθούν σαν σημεία εκκινήσεως μεθόδου τοπικής ελαχιστοποίησης ή να εισαχθούν σε κάποια ομάδα.
- 5 Αν βρεθεί τουλάχιστον ένα νέο ελάχιστο, θέσε $\alpha = p \sigma^2(\delta)$ και μετάβαση στο βήμα 2.
- 6 Τερματισμός αν $\left| \langle \delta \rangle - \frac{1}{2} \right| \leq \sigma_k(\delta)$ και $\sigma^2(\delta) \leq \alpha$, διαφορετικά μετάβαση στο βήμα 2.

3.3 Ο κανόνας των παρατηρησίμων μεγεθών

Ο κανόνας αυτός βασίζεται στην σύγκλιση των αναμενόμενων τιμών ορισμένων μεγεθών που υπολογίζονται κατά τη διάρκεια της διαδικασίας, στις ασυμπτωτικές του τιμές. Γίνεται μια σύγκριση ανάμεσα στη ποσότητα t_i (το πλήθος των ελαχιστοποιήσεων που οδήγησαν στο ελάχιστο y_i^*) και στην αναμενόμενη τιμή για αυτή την ποσότητα.

Έστω πως έχουν βρεθεί ω τοπικά ελάχιστα και από τις συνολικά T επαναλήψεις t_1 φορές βρέθηκε το ελάχιστο 1, t_2 το ελάχιστο 2 κ.ο.κ. Προφανώς ισχύει ότι $t_i \geq 1, \forall i = 1, \dots, \omega$. Επίσης συμβολίζουμε με n_J το πλήθος των επαναλήψεων που μεσολάβησαν για την εύρεση του J ελαχίστου από το προηγούμενο. Με βάση τα παραπάνω και λαμβάνοντας υπόψη ότι χρησιμοποιείται ομοιόμορφη δειγματοληψία ο αναμενόμενος αριθμός L_J^ω των τοπικών αναζητήσεων που οδηγήθηκαν στο ελάχιστο J την στιγμή που εντοπίζεται για πρώτη φορά το ελάχιστο ω είναι:

$$L_J^\omega = L_J^{\omega-1} + (n_\omega - 1) \frac{m(A_i)}{m(S)} \quad (3.19)$$

όπου $m(A)$ είναι το μέτρο Lebesgue του συνόλου A . Είναι προφανές πως μία τέτοια ποσότητα δεν είναι δυνατόν να εκτιμηθεί αναλυτικά, αλλά μπορεί να εκτιμηθεί ασυμπτωτικά με την χρήση των ποσοτήτων L_i . Καθώς το πλήθος των τοπικών αναζητήσεων αυξάνει θα ισχύει ότι:

$$\frac{t_i}{\sum_{j=1}^{\omega} t_j} \simeq \frac{m(A_i)}{m(S)} \quad (3.20)$$

Συνδιάζουμε τους 2 παραπάνω τύπους και προκύπτει ότι:

$$L_J^\omega = L_J^{\omega-1} + (n_\omega - 1) \frac{t_i}{\sum_{j=1}^{\omega-1} t_j} = L_J^{\omega-1} + (n_\omega - 1) \frac{t_i}{\sum_{j=1}^{\omega-1} n_i} \quad (3.21)$$

και ισχύει ότι $n_1 = 1, J \leq \omega - 1$ και $L_J^{(\omega)} = 1$. Ο παραπάνω τύπος ισχύει όταν έχει βρεθεί το ω ελάχιστο για πρώτη φορά. Ωστόσο μετά από την εύρεση αυτού του ελαχίστου είναι πιθανόν να έχουν γίνει και άλλες τοπικές αναζητήσεις που να είναι άκαρπες, έστω \mathbf{K} . Σε αυτήν την περίπτωση ο αναμενόμενος αριθμός εμφανίσεων του ελαχίστου \mathbf{J} δίνεται από:

$$L_J^{(\omega)}(K) = L_J^{(\omega)}(K - 1) + \frac{t_j}{K + \sum_{i=1}^W n_i} \quad (3.22)$$

με την ιδιότητα ότι $L_J^{(\omega)}(0) = L_J^{(\omega)}$. Όπως και στην πρώτη τεχνική τερματισμού που παρουσιάστηκε σε αυτό το κεφάλαιο, έτσι και εδώ το κριτήριο τερματισμού που χρησιμοποιήθηκε βασίστηκε στην διακύμανση της ποσότητας:

$$E_2(W, K) = \frac{1}{W} \sum_{J=1}^W \left(\frac{L_J^{(W)}(K) - t_J}{\sum_{i=1}^W t_i} \right)^2 \quad (3.23)$$

Η μέθοδος καθολικής ελαχιστοποίησης διακόπτεται όταν:

$$E_2(\omega, k) < \sigma_k(E_2), \sigma^2(E_2) < p\sigma_{last}^2(E_2) \quad (3.24)$$

όπου $\sigma_{last}^2(E_2)$ είναι η διακύμανση της ποσότητας E_2 όπως αυτή υπολογίστηκε την τελευταία φορά που βρέθηκε καινούργιο ελάχιστο και η ποσότητα p έχει την ίδια λειτουργία με την περίπτωση της μεθόδου του διπλασίου κουτιού.

3.4 Ο κανόνας των αναμενόμενων ελαχίστων

Η τεχνική αυτή βασίζεται στην προσπάθεια υπολογισμού του πλήθους των ελαχίστων μιας συνάρτησης. Η μέθοδος διακόπτεται όταν θεωρήσει πως δεν υπάρχουν πλέον άλλα ελάχιστα να καλύψει.

Έστω ότι με P_m^l συμβολίζεται η πιθανότητα ύστερα από m εφαρμογές της μεθόδου τοπικής ελαχιστοποίησης να έχουν βρεθεί l τοπικά ελάχιστα. Επίσης έστω ότι με π_k συμβολίζεται η πιθανότητα σε μια εφαρμογή της μεθόδου να βρει το ελάχιστο y_k^* . Αυτή η πιθανότητα είναι γνωστή εκ των προτέρων και δίνεται από τον τύπο:

$$\pi_k = \frac{m(A_k)}{m(S)} \quad (3.25)$$

Η πιθανότητα P_m^l μπορεί να υπολογιστεί αναδρομικά από τον τύπο:

$$P_m^l = \left(1 - \sum_{i=1}^{l-1} \pi_i\right) P_{m-1}^{l-1} + \left(\sum_{i=1}^l \pi_i\right) P_{m-1}^l \quad (3.26)$$

Παρατηρούμε ότι $P_0^1 = 0$, $P_1^1 = 1$. Στην γενική περίπτωση θα ισχύει ότι $P_m^l = 0$ όταν $l > m$ και $P_m^0 = 0, \forall m \geq 1$. Η λογική πίσω από την εξίσωση της πιθανότητας του καινούργιου ελαχίστου, είναι ότι μετά από m εφαρμογές τοπικής ελαχιστοποίησης να βρεθούν l ελάχιστα, συνδέεται με την πιθανότητα στις $(m-1)$ εφαρμογές είτε να έχουν βρεθεί τα $(l-1)$ τοπικά ελάχιστα (και το l τοπικό ελάχιστο ανακαλύπτεται στην επόμενη εφαρμογή) είτε να έχουν βρεθεί ακριβώς l τοπικά ελάχιστα και να μην βρεθεί στην επόμενη προσπάθεια. Η ποσότητα

$$\sum_{i=1}^l \pi_i$$

, είναι η πιθανότητα ότι ένα από τα l ελάχιστα βρίσκεται σε μία προσπάθεια και ομοίως η ποσότητα

$$1 - \sum_{i=1}^{l-1} \pi_i$$

είναι η πιθανότητα να μην βρεθεί κανένα από τα $l-1$ ελάχιστα σε μία προσπάθεια. Επιπλέον, αφού οι ποσότητες P_m^l ορίζει πιθανότητες θα πρέπει αν ισχύει ο τύπος:

$$\sum_{i=1}^m P_m^i = 1 \quad (3.27)$$

Για να αποδειχτεί ο παραπάνω τύπος ας θεωρήσουμε την ποσότητα

$$s_l = \sum_{i=1}^l \pi_i$$

. Εφαρμόζοντας το άθροισμα και στα δύο μέλη της ισότητας για την πιθανότητα P_m^l και έχουμε:

$$\sum_{l=1}^m P_m^l = \sum_{l=1}^m P_{m-1}^{l-1} - \sum_{i=1}^m s_{l-1} P_{m-1}^{l-1} + \sum_{l=1}^m s_l P_{m-1}^l \quad (3.28)$$

Μιας και $P_{m-1}^0 = 0, P_{m-1}^m = 0$ παρατηρούμε ότι:

$$\sum_{l=1}^m P_m^l = \sum_{l=1}^{m-1} P_{m-1}^l$$

Από φαίνεται ότι το άθροισμα

$$\sum_{l=1}^m P_m^l$$

δεν εξαρτάται από την ποσότητα m , οπότε θέτοντας $m = 1$ στην εξίσωση 3.27 επαληθεύουμε την εξίσωση 3.28.

Ο αναμενόμενος αριθμός ελαχίστων που έχουν βρεθεί ύστερα από m εφαρμογές της μεθόδου δίνεται από:

$$\langle \omega \rangle = \sum_{l=1}^m l P_m^l \quad (3.29)$$

και η διακύμανση από τον τύπο:

$$\sigma(\omega)_m^2 = \sum_{l=1}^m l^2 P_m^l - \left(\sum_{l=1}^m l P_m^l \right)^2 \quad (3.30)$$

Η ποσότητα

$$\pi_k = \frac{m(A_k)}{m(S)} \quad (3.31)$$

δεν μπορεί να υπολογιστεί ακριβώς παρά μόνο προσεγγιστικά. Μια προσέγγιση που μπορεί να πραγματοποιηθεί, η οποία εξαρτάται από το πλήθος εφαρμογών της μεθόδου τοπικής ελαχιστοποίησης (m) δίνεται από τον τύπο:

$$\pi_i^{(m)} = \frac{L_i^{(m)}}{m} \quad (3.32)$$

όπου $L_i^{(m)}$ είναι το πλήθος των εφαρμογών της μεθόδου τοπικής ελαχιστοποίησης που οδήγησαν στο ελάχιστο y^* . Έτσι η νέα μορφή της εξίσωσης για τον υπολογισμό της πιθανότητας είναι:

$$P_m^l = \left(1 - \sum_{i=1}^{l-1} \pi_i^{(m-1)} \right) P_{m-1}^{l-1} + \left(\sum_{i=1}^l \pi_i^{(m-1)} \right) P_{m-1}^l \quad (3.33)$$

Και στον νέο κανόνα τερματισμού, η καθολική μέθοδος τερματίζει με ένα κριτήριο που βασίζεται στην διακύμανση μιας ποσότητας και πιο συγκεκριμένα όταν:

$$|\langle \omega \rangle - \omega| \leq \sigma(\omega)_m, \sigma^2(\omega)_m < p\sigma^2(\omega)_{last} \quad (3.34)$$

ΚΕΦΑΛΑΙΟ 4

ΚΑΘΟΛΙΚΗ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗ ΜΕ ΧΡΗΣΗ ΚΑΝΟΝΙΚΗΣ ΚΑΤΑΝΟΜΗΣ

Μέχρι στιγμής έχουμε παρουσιάσει μεθόδους τοπικής ελαχιστοποίησης και έχουμε καλύψει το πρόβλημα της επιλογής κριτηρίου τερματισμού. Σ' αυτό το κεφάλαιο θα αναλύσουμε τους αλγόριθμους πολλαπλής εκκίνησης καθώς και μια παραλλαγή του, την Adapt multistart και θα παρουσιάσουμε την πρότασή μας για καθολική βελτιστοποίηση η οποία βασίζεται στην επιλογή σημείων από πολυδιάστατη κατανομή. Θα αναλύσουμε το τρόπο δειγματοληψίας και των τριών αλγορίθμων και ενημέρωση των παραμέτρων πιθανοφάνειας.

4.1 Πολλαπλή Εκκίνηση - Multistart

Η μέθοδος πολλαπλής εκκίνησης είναι από της πιο διαδεδομένες μεθόδους καθολικής ελαχιστοποίησης. Ξεκινάμε επιλέγοντας ένα σημείο x από το σύνολο τιμών της συνάρτησης και εκτελούμε μια τοπική αναζήτηση σ' αυτό το σημείο. Αν το σημείο y που μας επιστρέφει η αναζήτηση είναι καινούργιο ελάχιστο, το κρατάμε αλλιώς συνεχίζουμε την αναζήτηση έως ότου να πετύχουμε το κριτήριο τερματισμού (αλγόριθμος 1).

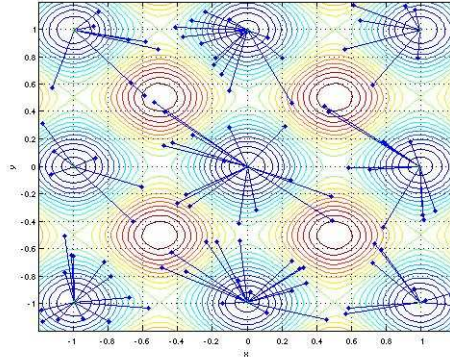
- 1 Επιλέγω σημείο x από $S \subseteq \mathbb{R}^n$
- 2 Εφαρμόζω τοπική ελαχιστοποίηση (L) που μας οδηγεί στο σημείο y
- 3 $k = 1$
- 4 Επαναληπτικά:
ΕΠΙΛΟΓΗ ΣΗΜΕΙΟΥ x
Αν $(x \in \bigcup_{i=1}^k A_i)$ τότε επιλέγω καινούργιο x
αλλιώς $k = k + 1$ και $y_k = L(x)$
ΕΛΕΓΧΟΣ ΓΙΑ "ΚΡΙΤΗΡΙΟ ΤΕΡΜΑΤΙΣΜΟΥ"
ΕΠΙΛΟΓΗ ΚΑΙΝΟΥΡΓΙΟΥ ΣΗΜΕΙΟΥ

Αλγόριθμος 1: Πολλαπλή Εκκίνηση

4.2 Adapt Multistart

Η επιλογή των σημείων που θα ξεκινήσουμε τοπική ελαχιστοποίηση γίνεται με ομοιόμορφη κατανομή. Όμως σε συναρτήσεις με μεγάλο αριθμό ελαχίστων, η αναζήτηση είναι πολύ απαιτητική και αυτό γιατί, κατά τη διάρκεια της αναζήτησης μπορεί να πέσουμε σε ελάχιστο που έχουμε ήδη βρει παραπάνω από μία φορές κάτι που φαίνεται στην εικόνα 4.1.

Επίσης κάθε φορά για να επιλέξω καινούργιο σημείο x παρατηρώ ότι ελέγχω αν το x ανήκει στην περιοχή έλξης του ελαχίστου y . Οι Λαγαρής και Βόγλης στην πρότασή τους [19] προτείνουν μέσω το αλγόριθμό τους, Adapt Multistart, μια διαφορετική προσέγγιση με χρήση πιθανοτήτων. Διαφοροποιούν τον αλγόριθμο έτσι ώστε να ελέγχουν αν το καινούργιο σημείο δεν ανήκει σε περιοχές έλξης ελαχίστων (αλγόριθμος 2).



Σχήμα 4.1: Αποτέλεσμα εκτέλεσης αλγορίθμου *multistart* στην συνάρτηση *Rastrigin*, στο πεδίο τιμών $[-1.2, 1.2]$. Παρατηρούμε ότι η αναζήτηση τοπικού ελαχίστου πολλές φορές οδηγεί σημεία στο ίδιο, παραπάνω από μια φορά.

1 Αρχικοποίηση:

$$k = 1$$

Επιλέγω $x \in S$

$$y_k = L(x)$$

2 Τερματισμός: Σταματάω σύμφωνα με το κριτήριο τερματισμού

3 Έλεγχος:

Αν $x \notin \bigcup_{i=1}^k A(y_i)$ τότε

$$y = L(x), k = k + 1, y_k = k + 1$$

4 Επανάληψη ΕΛΕΓΧΩ ΚΡΙΤΗΡΙΟ ΤΕΡΜΑΤΙΣΜΟΥ

Αλγόριθμος 2: Adapt Multistart

Το πρόβλημα του υπολογισμού των περιοχών έλξης κάθε ελαχίστου παραμένει μιας και με την εύρεση του καινούργιου ελαχίστου y δεν έχω υπολογίσει την περιοχή έλξης του. Γενικά ο υπολογισμός της περιοχής έλξης (μέγεθος, σχήμα) ενός ελαχίστου, δεδομένου μόνο το ελάχιστο, είναι δύσκολος. Γι'αυτό τροποποιούνε τον αλγόριθμο και εισάγουν πιθανότητες για να εκτιμήσουν την περιοχή έλξης. Η αλλαγές γίνονται στο κομμάτι ελέγχου του κώδικα (αλγόριθμος 3).

Έτσι με αυτό τον τρόπο δεν χρειάζεται να υπολογίζουμε τις περιοχές έλξης παρά μόνο να ελέγξουν μέσω πιθανότητας αν θα ξεκινήσουν τοπική ελαχιστοποίηση. Το πρόβλημά μας τώρα έγγυται στην εκτίμηση της πιθανότητας. Επιλέγω σημεία $x \in S$ και για κάθε σημείο x_i που οδηγεί στο ελάχιστο y_i υπολογίζω την ποσότητα:

$$R_i = \text{maximum} |x_j^{(i)} - y_i| \quad (4.1)$$

δηλαδή δημιουργούμε σφαίρες γύρω από το ελάχιστο με ακτίνα ίση με τη μεγαλύτερη απόσταση από το σημείο x_i . Έστω το ελάχιστο y_i έχει ακτίνα R και έστω σημείο x . Αν

1. Έλεγχος:

Υπολόγισε \mathbf{p} έτσι ώστε $x \notin \bigcup_{i=1}^k A(y_i)$

Επιλέγουμε $\xi \in (0, 1)$

Αν $(\xi \leq p)$ τότε

$y = L(x)$

Αν $(y \notin y_1, y_2, y_3, \dots, y_n)$ τότε

$k = k + 1$

$y_k = y$

Αλγόριθμος 3: Έλεγχος

$|y - x| < R$ τότε το x έχει μια αρκετά μεγάλη πιθανότητα να ανήκει στην περιοχή έλξης του x , δηλαδή στο $A(y)$. Εάν $(\vec{y} - \vec{x})^T \nabla f(x) \geq 0$ τότε η κατεύθυνση είναι αύξουσα, που σημαίνει ότι αν κάνω τοπικές αναζητήσεις, θα βρω τιμές μεγαλύτερες από πριν. Έστω $z = \frac{|y-x|}{R} \in (0, 1)$ όταν $|y - x| < R$. Ορίζω τότε την πιθανότητα το $x \notin A(y)$ ως εξής:

$$P(x \notin A(y)) = \begin{cases} 1, \text{ αν } z > 1 \text{ ή } (\vec{y} - \vec{x})^T \nabla f(x) \geq 0 \\ \varphi(z, l) [1 + \frac{(\vec{y} - \vec{x})^T \nabla f(x)}{|\vec{y} - \vec{x}|^T |\nabla f(x)|}], \text{ άλλο} \end{cases}$$

όπου l είναι το πλήθος που έχουμε βρει το ελάχιστο y , κατά τη διάρκεια της τοπικής ελαχιστοποίησης. Βλέπουμε ότι η παραπάνω πιθανότητα γίνεται 1 όταν το διάνυσμα $\vec{y} - \vec{x}$ είναι κάθετο στο $\nabla f(x)$, δηλαδή $\cos(\nabla f(x), \vec{y} - \vec{x}) = 0$ και 0 όταν έχουν αντίθετες κατευθύνσεις, δηλαδή $\cos(\nabla f(x), \vec{y} - \vec{x}) = -1$.

Το πιθανοτικό μοντέλο $\varphi(z, l)$ που χρησιμοποιούμε για τον υπολογισμό της πιθανότητας, θέλουμε να ικανοποιεί τις παρακάτω ιδιότητες:

1. $\lim_{z \rightarrow 0} \varphi(z, l) = 0$
2. $\lim_{z \rightarrow 1} \varphi(z, l) = 1$
3. $\lim_{z \rightarrow \infty} \varphi(z, l) = 0$
4. $0 \leq \varphi(z, l) \leq 1$

Η συνάρτηση που ικανοποιεί τις παραπάνω συνθήκες είναι η 4.2.

$$\varphi(z, l) = ze^{-l^2(z-1)^2} \quad (4.2)$$

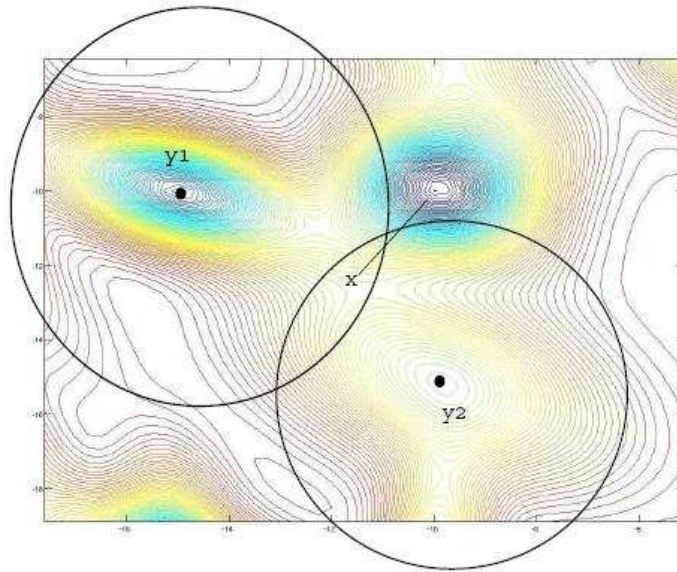
το οποίο ικανοποιεί και τις παραπάνω συνθήκες.

Μέχρι στιγμής υπολογίσαμε την πιθανότητα το $x \notin A(y)$. Όταν έχω y_1, y_2, \dots, y_k ελάχιστα θα έχω αντίστοιχα A_1, A_2, \dots, A_k περιοχές έλξης, οπότε ψάχνουμε την πιθανότητα το x να μην ανήκει σε καμία περιοχή έλξης, δηλαδή:

$$x \notin A_1 \cup x \notin A_2 \dots \cup x \notin A_k \Rightarrow x \notin \bigcup_{i=1}^k A_i \Rightarrow P(x \notin \bigcup_{i=1}^k A_i) = \prod_{i=1}^k P(x \notin A_i) \quad (4.3)$$

Όμως με τον παραπάνω τρόπο που υπολογίζουμε την πιθανότητα για ένα καινούργιο σημείο να μην ανήκει σε μια περιοχή έλξης δημιουργείται ένα πρόβλημα. Συγκεκριμένα, κάθε περιοχή έλξης είναι μια σφαίρα με ακτίνα ίση με την απόσταση $\vec{y} - \vec{x}$. Με αυτό τον τρόπο έχουμε την πιθανότητα εμφάνισης επικάλυψης, δηλαδή ένα σημείο που πιθανόν να κατέληγε σε καινούργιο ελάχιστο να βρίσκεται μέσα σε περιοχές έλξης ελαχίστων που έχουμε ήδη βρει, με αποτέλεσμα η τοπική ελαχιστοποίηση να μην ξεκινήσει.

Ας παρατηρήσουμε την εικόνα 4.2 Αν πάρω ένα σημείο που ανήκει στην περιοχή έλξης



Σχήμα 4.2: Εμφάνιση επικάλυψης σε τοπική αναζήτηση ελαχίστου.

και του y_1 αλλά και του y_2 θα έπρεπε να υπολογίσουμε το $p(x \notin A_1)p(x \notin A_2)$ το οποίο θα είναι πολύ μικρό με αποτέλεσμα, για το συγκεκριμένο σημείο, να μην ξεκινήσουμε τοπική ελαχιστοποίηση.

Για να λύσουμε το παραπάνω θέτουμε:

$$P(x \notin \bigcup_{i=1}^k A_i) = \prod_{i=1}^k P(x \notin A_i) \approx P(x \notin A_n) \quad (4.4)$$

, όπου $A_n = A(y_n)$ και y_n το κοντινότερο ελάχιστο.

4.3 Τοπική ελαχιστοποίηση με χρήση κανονικής κατανομής

Είμαστε έτοιμοι να αναλύσουμε την πρόταση μας. Ο αλγόριθμός μας προτείνει μια καινούργια μέθοδο δειγματοληψίας που ενισχύει τις ήδη υπάρχουσες.

Όπως είδαμε στον αλγόριθμο Adapt Multistart, η επιλογή ενός σημείου για την εκίνηση τοπικής ελαχιστοποίησης γίνεται μέσω της πιθανότητας $P(x \notin A_n)$. Στην πρόταση που θα παρουσιάσουμε χρησιμοποιείται Κανονική κατανομή (Gaussian distribution). Συγκεκριμένα μετά από κάθε εύρεση τοπικού ελαχίστου y^* , αναθέτουμε στο σημείο την συνάρτηση κατανομής f που δίνεται από τον τύπο 4.5.

$$f = \frac{1}{\sqrt{2\pi}} \frac{1}{|\Sigma|} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (4.5)$$

όπου μ = μέση τιμή και Σ = ο πίνακας διασποράς.

Βασιστήκαμε στην κανονική κατανομή για δύο λόγους. Πρώτον έχει παραμέτρους που μπορούν αν εκτιμηθούν και δεύτερον παρουσιάζει τοπικές ιδιότητες. Δοθέντων N σημείων x_1, x_2, \dots, x_N μπορούμε να υπολογίσουμε τη μέση τιμή και τον πίνακα διασποράς ως:

$$\mu = E(X) = \frac{1}{N} \sum_{i=1}^N x_i \quad (4.6)$$

$$\Sigma = E((X - \mu)(X - \mu)^T) = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T \quad (4.7)$$

Τα βήματα του αλγορίθμου έχουν ως εξής:

Αρχικοποίηση:

Επιλέγω τυχαίο σημείο και εφαρμόζω τοπική ελαχιστοποίηση και καταλήγω στο ελάχιστο y^* και αναθέτω την τιμή f_1 , όπως την ορίσαμε παραπάνω, υπολογίζω τον πίνακα διασποράς και την μέση τιμή.

Σε κάθε επανάληψη γίνονται τα εξής:

- 1 Δειγματοληψία: Ένα σημείο x επιλέγεται εάν για κάποιο $\xi \in (0, 1)$ ισχύει $\xi \times \max > f$, όπου f είναι η τιμή της συνάρτησης κατανομής και \max είναι η μέγιστη τιμή πιθανότητας, στο σημείο x .
- 2 Ενημέρωση: Για κάθε ελάχιστο που βρίσκω ενημερώνω τα εξής:
 - i Τα εκτιμώμενα μ_i και Σ_i
 - ii Μετρητή ρ_i πόσες φορές βρέθηκε το ελάχιστο
 - iii Τη θέση και την τιμή του ελαχιστου
- 3 Έλεγχος κριτηρίου τερματισμού

Αλγόριθμος 4: Καθολική Ελαχιστοποίησης με χρήση Κανονικής Κατανομής

4.3.1 Δειγματοληψία

Οι μέθοδοι δειγματοληψίας είναι πολύ σημαντικές διότι με την βοήθειά τους εγγυώνται την κάλυψη όλου του χώρου λύσεων. Στην βιβλιογραφία έχουν προταθεί διάφορες μέθοδοι όπως:

1. Ομοιόμορφη κατανομή
2. Κανονική κατανομή (βλ. Simulated Annealing)
3. Quasi-Random κατανομές
4. Importance Sampling
5. Rejection Sampling

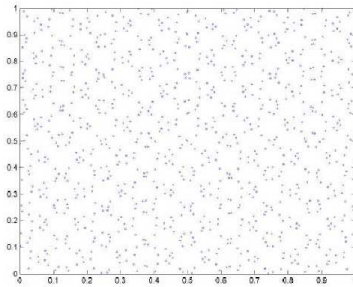
Στην ομοιόμορφη κατανομή μπορούμε να βρούμε μία αναλυτική έκφραση και έτσι για X, U, V σημεία επιλεγμένα στο $U(0, 1)$ μπορούμε να έχουμε:

- Κανονική κατανομή $N(0, 1)$: $Y = \sqrt{-2\ln(U)} \cos(2\pi V)$
- Εκθετική κατανομή με παράμετρο λ : $Y = -\frac{\ln(X)}{\lambda}$
- β -κατανομή με παράμετρο 1 και ν : $Y = 1 - X^{1/\nu}$

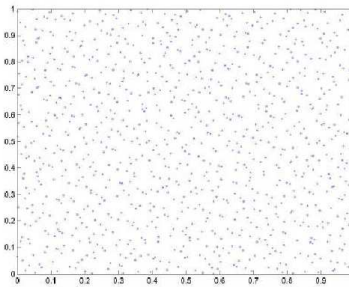
Οι Quasi-Random κατανομές παράγονται από ακολουθιακούς αλγορίθμους που λαμβάνουν υπόψη προηγούμενα σημεία και μετά από άπειρες επαναλήψεις προσεγγίζουν την ομοιόμορφη. Μερικές ακολουθίες αυτής της κατηγορίας είναι οι Halton, οι Sobol και οι Niederreiter. Παρ'όλου που προσεγγίζουν την ομοιόμορφη κατανομή η κάλυψη του χώρου γίνεται

καλύτερα και αυτό μπορούμε να το δούμε παρακάτω.

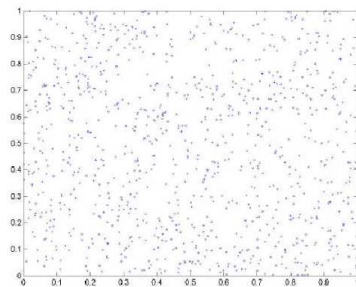
Παρατηρούμε πως η ομοιόμορφη κατανομή αφήνει ακάλυπτες περιοχές ενώ οι Quasi-



(a) Sobol sequence



(b) Niederreiter



(c) Uniform

Random κατανομές καλύπτουν πιο "ομοιόμορφα" τον χώρο.

Το rejection sampling χρησιμοποιείται για δειγματοληψία από πολύπλοκες κατανομές $f(x)$. Χρησιμοποιεί μια βοηθητική κατανομή $g(x)$ με την προϋπόθεση ότι $f(x) < Mg(x)$, όπου $M > 1$. Τα βήματα του αλγορίθμου rejection sampling δίνονται στον αλγόριθμο 5

Στην υλοποίησή μας θέτουμε όπου $f(x) = N(x)$, $g(x) = U(0, 1)$ και $M = \max N(x)$

* Δειγματοληπτώ x από $g(x)$ και u από $U(0, 1)$

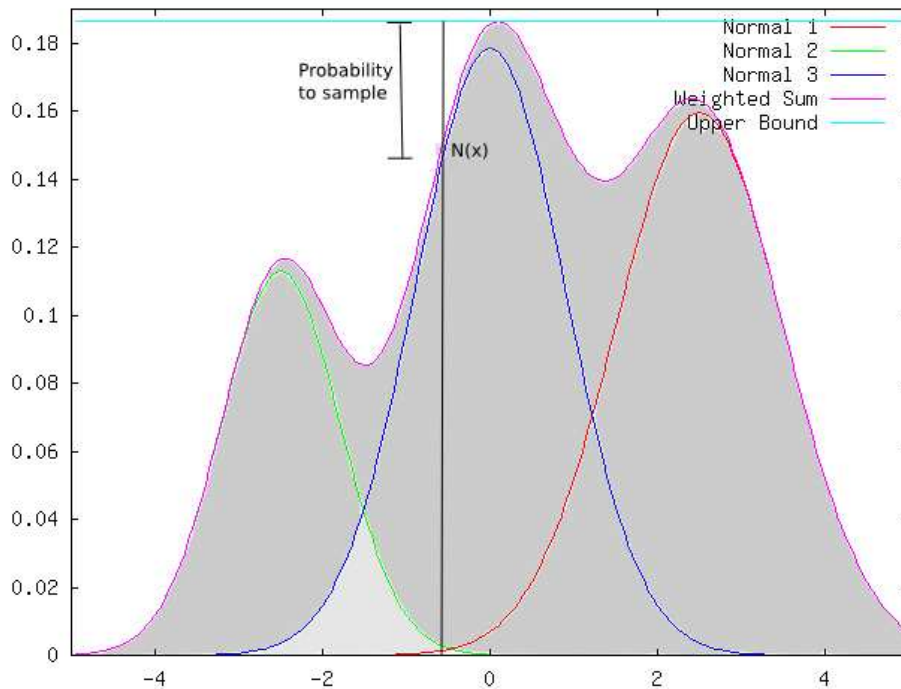
* Ελέγχω $u < \frac{f(x)}{Mg(x)}$.

ο Αν ισχύει, αποδέχομαι το x as a realization of $f(x)$;

ο Αν όχι, απορίπτω την τιμή του x και επαναλαμβάνω τη δειγματοληψία.

Αλγόριθμος 5: Rejection Sampling

Η μέθοδος δειγματοληψίας που χρησιμοποιήσαμε στην πρότασή μας στην ουσία είναι το inverse rejection sampling. Έστω ότι κατά τη διάρκεια τοπικής ελαχιστοποίησης στο διάστημα $[a, b]$ έχουμε βρει 3 τοπικά ελάχιστα και έστω ότι το x είναι υποψήφιο σημείο για έναρξη τοπικής ελαχιστοποίησης. Κάθε y_1, y_2, y_3 ακολουθεί την κατανομή $N(\mu, \Sigma)$.



Σχήμα 4.3: Επιλογή πιθανού σημείου

Ο αλγόριθμος επιλέγει το καινούργιο σημείο για να ξεκινήσουμε τοπικές ελαχιστοποίηση όταν για κάποιο $\xi \in (0, 1)$ ισχύει $\xi \times \max > f$. Η ποσότητα \max είναι η μέγιστη τιμή του γραμμικού συνδιασμού των κατανομών του σημείου x , ενώ η N είναι η τιμή της κατανομής στο αντίστοιχο σημείο.

Αναλυτικά ο αλγόριθμος δειγματοληψίας είναι ο 6.

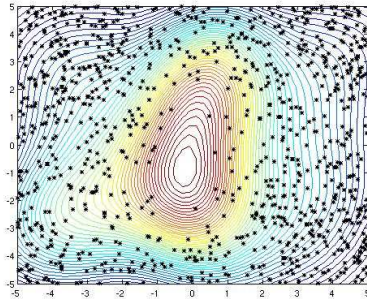
Στο σχήμα 4.3, με κόκκινη γραμμή συμβολίζεται ο γραμμικός συνδιασμός των κατανο-

```
function sample_of_gaussians(Xmin, mx, xl, xu) while(1)
    testx <- uniform(xl, xu)
    f <- sum_of_gaussians(Xmin, testx);
    ksi = uniform(0, 1);
    if (ksi * mx > f)
        x = testx; /* Δέχομαι το σημείο */
        return
    end
end
```

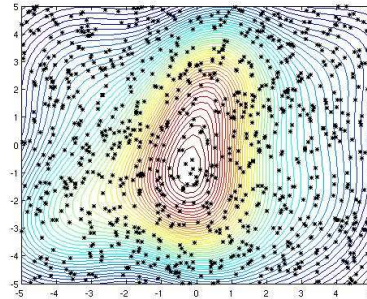
Αλγόριθμος 6: Inverse Rejection Sampling

μών. Στο σημείο x βρίσκουμε τη μέγιστη τιμή του γραμμικού συνδιασμού και ελέγχουμε τον "χώρο" που βρίσκεται ανάμεσα από τη μέγιστη τιμή της κατανομής και την τιμή της κατανομής στο σημείο x . Πάνω και κάτω από την κόκκινη γραμμή. Όσο μεγαλύτερο "χώρο" έχουμε τόσο μεγαλύτερη είναι η πιθανότητα να δεχθούμε το σημείο.

Η μέθοδος αποδίδει πολύ καλά και απορρίπτει σημεία που θα μας οδηγούσαν στο ίδιο τοπικό ελάχιστο. Αν δούμε τα παρακάτω παραδείγματα που αφορούν τις συναρτήσεις Six Hump και Rastrigin θα δούμε εμφανείς διαφορές στο τρόπο επιλογής σημείων σε σύγκριση με μία ομοιόμορφη κατανομή. Στα παραδείγματα που παραθέτουμε έχουμε επιλέξει 200 τυχαία

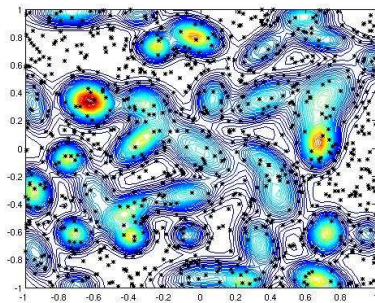


(a) *Six Hump* με κανονική κατανομή

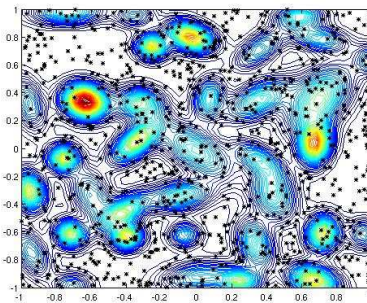


(b) *Six Hump* με χρήση ομοιόμορφη κατανομή

Σχήμα 4.4: Δειγματοληψία στη συνάρτηση Six Hump με χρήση ομοιόμορφης και κανονικής κατανομής



(a) *Rastrigin* με κανονική κατανομή



(b) *Rastrigin* με ομοιόμορφη κατανομή

Σχήμα 4.5: Δειγματοληψία σημείων στη συνάρτηση Rastrigin με χρήση ομοιόμορφης και κανονικής κατανομής

σημεία και ξεκινήσαμε τοπικές ελαχιστοποιήσεις. Στην συνέχεια επιλέξαμε σημεία με βάση τον αλγόριθμό μας αλλά και με ομοιόμορφη κατανομή. Οι εικόνες απεικονίζουν την κατανομή των συναρτήσεων. Οι περιοχές με σκούρο χρώμα έχουν υψηλότερες καμπάνες, δηλαδή το συγκεκριμένο ελάχιστο έχει βρεθεί πολλές φορές. Έτσι στην επιλογή ενός καινούργιου σημείου θέλουμε να αποφύγουμε αυτές τις περιοχές καθώς μια τοπική ελαχιστοποίηση θα μας οδηγήσει σε ελάχιστο που έχουμε ήδη υπολογίσει.

Βλέπουμε ότι ο αλγόριθμός μας αποφεύγει τις σκούρες περιοχές, κατευθύνοντας τα πιθανά σημεία που η πιθανότητα εύρεσης καινούργιου ελαχίστου είναι μεγάλη, ενώ η ομοιόμορφη κατανομή όχι.

Παρ'όλα αυτά παρατηρούμε ότι, κατά τη διάρκεια της δειγματοληψίας, υπάρχει μια μεγάλη συγκέντρωση σημείων προς τα όρια της συνάρτησης κάτι που μπορεί να καθυστερήσει την εκτέλεση του αλγορίθμου μας καθώς για εκείνα τα σημεία θα ξεκινήσει τοπική ελαχιστοποίηση. Περισσότερα αποτελέσματα πειραμάτων παραθέτουμε στο παράρτημα.

4.3.2 Online εκτίμηση παραμέτρων κανονικής κατανομής

Η ενημέρωση των εκτιμητών μεγίστης πιθανοφάνειας γίνεται κατά τη διάρκεια εκτέλεσης του αλγορίθμου. Για κάθε ελάχιστο που υπολογίζουμε ενημερώνουμε τις παρακάτω ποσότητες ως εξής:

$$\mu_i = \mu_{i-1} + \alpha_i(x_i - \mu_{i-1}) \quad (4.8)$$

$$\Sigma_i = \Sigma_{i-1} + \alpha_i(x_i - \mu_i)(x_i - \mu_i)^T, \alpha_i \in (0, 1), \mu_0 = (0, 0, \dots, 0)^T, \Sigma_0 = \alpha I_n \quad (4.9)$$

Η ποσότητα α_i είναι ο ρυθμός μάθησης της κατανομής και μπορεί να είναι είτε μεταβλητού μεγέθους είτε σταθερού. Στην περίπτωση του μεταβλητού μεγέθους το α_i ισούται με $\frac{1}{x(\min) \cdot Count}$, δηλαδή το πλήθος των φορών που έχουμε βρει το συγκεκριμένο ελάχιστο. Αυτό έχει ως αποτέλεσμα να δίνει περισσότερο βάρος στα πρώτα σημεία που έχουν καταλήξει στο συγκεκριμένο ελάχιστο. Στην περίπτωση που είναι σταθερού μεγέθους επιλέγουμε ένα μικρό αριθμό της τάξεως 10^{-4} . Αυτή η ποσότητα όμως, αναλόγως τη συνάρτηση, διαφέρει και πρέπει να εκτιμηθεί πειραματικά.

Η ενημέρωση του πίνακα διασποράς είναι Rank-1, οπότε μέσα από Cholesky decomposition έχουμε ότι $\Sigma = LL^T$, άρα σε κάθε επανάληψη αποθηκεύουμε μόνο τον παράγοντα L (όπου είναι ένας πίνακας τριγωνικός, θετικά ορισμένος) και η ενημέρωσή του γίνεται και αυτή σε Rank-1. Με αυτόν τον τρόπο επιτυγχάνουμε ενημέρωση του πίνακα L σε $O(n^2)$. Επίσης η ορίζουσα του πίνακα διασποράς μπορεί αν υπολογιστεί πολύ εύκολα ως:

$$\det(\Sigma) = \prod_{i=1}^N L_{ii}^2 \quad (4.10)$$

όπου είναι το γινόμενο του τετραγώνου των στοιχείων της διαγωνίου του πίνακα L .

Τέλος για τον υπολογισμό του εκθετικού παράγοντα αυτό που έχουμε να κάνουμε είναι να αντικαταστήσουμε τον πίνακα διασποράς Σ με LL^T και έχουμε:

$$(x - \mu)^T \Sigma^{-1} (x - \mu) = (L^{-1}(x - \mu))^T (L^{-1}(x - \mu)) \quad (4.11)$$

Αν θέσουμε $L^{-1}(x - \mu) = y \Rightarrow Ly = x - \mu$, αυτό που μας μένει είναι να λύσουμε ως προς y και στην συνέχεια να πολλαπλασιάσουμε με τον ανάστροφό του.

Η ενημέρωση του πίνακα διασποράς μπορεί να γίνει με δύο τρόπους:

1. Αρχικοποίηση με τον Εσσιανό πίνακα υπολογισμένο στο σημείο από όπου ξεκινήσαμε την τοπική ελαχιστοποίηση
2. Αρχικοποίηση με τον μοναδιαίο πίνακα I_n πολλαπλασιασμένο με μία μικρή ποσότητα a_i

Στην περίπτωση που αρχικοποιούμε με τον Εσσιανό πίνακα επιτυγχάνουμε την καλύτερη αναπαράσταση της μορφής του ελαχίστου αντλώντας πληροφορίες από τα ιδιοδυνασμάτα

του πίνακα, ενώ στην δεύτερη περίπτωση προσπαθούμε η κατανομή να είναι εντοπισμένη κοντά στο ελάχιστο. Στην ουσία δημιουργώ την ελάχιστη δυνατή περιοχή έλξης. Και εδώ η ποσότητα a_i που αναφέραμε πρέπει να υπολογιστεί πειραματικά.

Οπότε το τελικό μοντέλο κατανομής είναι ο γραμμικός συνδυασμός των κατανομών που έχουμε υπολογίσει για κάθε ελάχιστο που βρήκαμε.

Ο τύπος είναι:

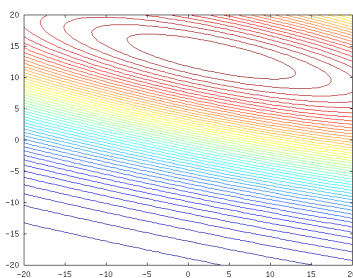
$$N(x) = \sum_{i=1}^M \pi_i N(x; \mu_i, \Sigma_i), \quad \sum_{i=1}^M \pi_i = 1 \quad (4.12)$$

Η ποσότητα π_i ισούται με

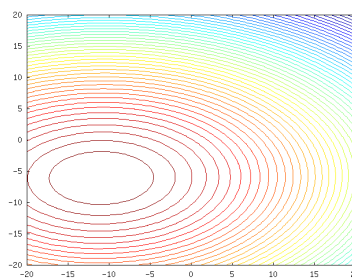
$$\pi_i = \frac{\rho_i}{\sum_{i=1}^M \rho_i} \quad (4.13)$$

, δηλαδή είναι το πλήθος εύρεσης του ελαχίστου y_i προς τον συνολικό αριθμό τοπικών ελαχιστοποιήσεων που πραγματοποιήσαμε. Η ποσότητα αυτή θέλουμε να ισούται με τη μονάδα ούτως ώστε το ολοκλήρωμα της κατανομής $N(x)$ να είναι ίσο με 1.

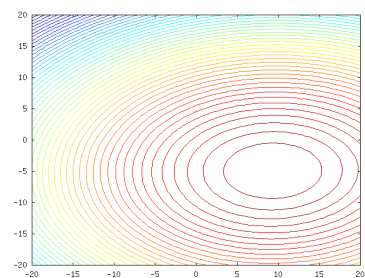
Στο παρακάτω παράδειγμα ξεκινήσαμε έναν αλγόριθμο ελαχιστοποίησης με τη μέθοδο μας και βλέπουμε κατά τη διάρκεια εκτέλεσης πως αλλάζει η μορφή της κατανομής και πως αυτή καταλήγει να είναι ισοδύναμη με τη μορφή του ελαχίστου.



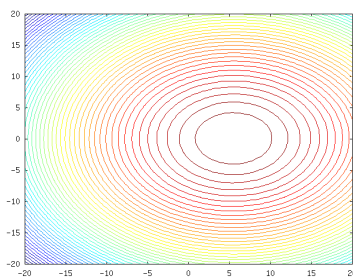
(a) Iter 1



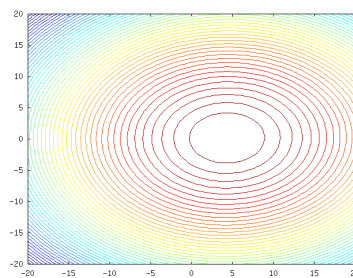
(b) Iter 3



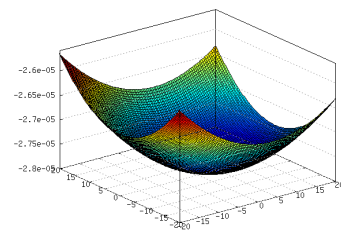
(c) Iter 6



(d) Iter 9



(e) Iter 12



(f) Μορφή ελαχίστου

Σχήμα 4.6: On-line προσδιορισμός παραμέτρων ελάχιστο στο $x^* = [4, 0]^T$

4.3.3 Δειγματοληψία ως μέθοδος τερματισμού

Κατά τη διάρκεια του inverse rejection sampling είναι λογικό να θεωρήσουμε ότι όταν ο γραμμικός συνδιασμός των κανονικών κατανομών γεμίσει τον χώρο θα πρέπει να σταματήσουμε την αναζήτηση. Αυτό γιατί κατά τη διάρκεια της επιλογής νέου σημείου, κάθε υποψήφιο σημείο θα απορίπτεται καθώς το γινόμενο $\xi \cdot \max$ δεν θα μπορεί να γίνει μεγαλύτερο του f (όπου $\xi \in U(0, 1)$, \max μέγιστη τιμή του αθροίσματος κατανομών και f τιμή του αθροίσματος κατανομών στο υποψήφιο σημείο). Οπότε ο αλγόριθμός μας μπορεί να τερματίζει όταν απορρίπτονται πολλά σημεία κατά το inverse rejection sampling (threshold based rule). Το μειονέκτημα σε αυτή την περίπτωση είναι ότι το φράγμα των αποτυχημένων δοκιμών θα πρέπει να εκτιμηθεί για κάθε συνάρτηση.

Ένα δεύτερο κριτήριο θα είναι όταν το πηλίκο

$$\frac{N(x)}{\max \{N(x)\}g(x)} \quad (4.14)$$

πλησιάζει στην μονάδα που δηλώνει ότι η πιθανότητα να βρούμε νέο σημείο αναζήτησης πλησιάζει στο 0. Με αυτό το κριτήριο ενώ δεν χρειάζεται να εκτιμήσουμε τον αριθμό αποτυχιών, όπως αναφέραμε, ο αλγόριθμός πρέπει να πάρει αρκετά δείγματα ούτως ώστε να μπορέσει να ικανοποιήσει το κριτήριο.

ΚΕΦΑΛΑΙΟ 5

ΠΕΙΡΑΜΑΤΑ ΚΑΙ ΜΕΤΡΗΣΕΙΣ

Στο παρόν κεφάλαιο παραθέτουμε πειραματικά αποτελέσματα που πραγματοποιήθηκαν σε περιβάλλον Linux (Ubuntu 7.3). Το σύστημά μας απαρτιζόταν από επεξεργαστή Intel Pentium 4 με συχνότητα 2.8 GHz με 1 GB RAM. Η υλοποίηση των αλγορίθμων και των συναρτήσεων έγινε σε γλώσσα προγραμματισμού C και χρησιμοποιήθηκε το περιβάλλον βελτιστοποίησης Merlin [1].

Σε κάθε πείραμα μετρήσαμε τον αριθμό τοπικών αναζητήσεων που χρειάστηκε ο αλγόριθμος για την εύρεση συγκεκριμένου αριθμού ελαχίστων. Η στήλη Uni αναφέρεται στο πλήθος επαναλήψεων χρησιμοποιώντας την ομοιόμορφη κατανομή ως μέθοδος δειγματοληψίας. Η στήλη $Norm5$ αναφέρεται στο πλήθος επαναλήψεων δειγματοληπώντας 5 αρχικά σημεία μέσω ομοιόμορφης κατανομής και στη συνέχεια χρησιμοποιήσαμε την πρότασή μας. Αντιστοίχως το ίδιο ισχύει για τις στήλες $Norm10$, $Norm20$, $Norm50$ και $Norm100$.

Όπως βλέπουμε (πίνακας 5), η βελτίωση που παρατηρείται είναι της τάξεως του 15 τοις εκατό, που είναι μια αρκετά σημαντική μείωση. Βλέπουμε πως ο αλγόριθμός μας εκμεταλεύεται την πληροφορία που εξάγει από τα τοπικά ελάχιστα και με αυτόν τον τρόπο δεν επαναλαμβάνουμε αναζητήσεις ελαχίστων σε περιοχές που έχουμε ήδη εξερευνήσει. Ακόμα και σε συναρτήσεις με μικρό αριθμό ελαχίστων (όπως η Bra, Hyper και Matyas) παρατηρείται βελτίωση

Function	Number of Local Searches					
	Uni	Norm5	Norm10	Norm20	Norm50	Norm100
Rast	6829.2	5853.4	5817.1	6113.6	5744.4	6566.4
Tube	879.3	735.6	941.5	719.9	641	846.8
Bird	932	731.6	790.8	785.2	589.6	689.4
Griew	1733.9	1509.7	1635.6	1685.6	1796.4	1636.5
Six-Hump	252.3	192.9	239.8	264.8	206.3	237.7
Bra	28	28.8	32.3	22.4	30.5	22.4
Hyper	14	13.3	15.1	13.1	14	14
Matyas	7.2	7.8	7.9	7.3	7.4	7.7

Πίνακας 5.1: Πίνακας αποτελεσμάτων με πίνακα διασποράς αρχικοποιημένο με τον Εσσιανό και μεταβλητός ρυθμός μάθησης

Στην περίπτωση που ο ρυθμός μάθησης παραμένει σταθερός (πίνακας 5.2), ίσος με 10^{-4} , δεν παρατηρούμε την ίδια συμπεριφορά στα αποτελέσματα των πειραμάτων. Όπως παρατηρούμε στον παρακάτω πίνακα, σε αρκετές περιπτώσεις αντί για μείωση των τοπικών αναζητήσεων έχουμε αύξηση (βλέπε συνάρτηση Rast, Griew, Tube).

Η συμπεριφορά αυτή οφείλεται στην επιλογή του ρυθμού μάθησης που ανά συνάρτηση διαφέρει καθώς αλλάζει η μορφή των ελαχίστων και το πεδίο τιμών.

	Number of Local Searches					
Function	Uni	Norm5	Norm10	Norm20	Norm50	Norm100
Rast	6469.9	5484.6	7122	6050.9	6261	6351
Tube	813	1002.1	891.3	884.1	973.2	899.2
Bird	457.8	290.3	321.9	203	220.6	244.4
Griew	1605.7	1624.3	1554.8	1695.1	1496.1	1765.2
Six-Hump	229.5	260.3	230.6	281.3	231.7	207.8
Bra	22	21.6	20.6	19.6	29.2	19.6
Hyper	14.7	14	13.8	13.1	14.3	14.1
Matyas	7.7	8.2	7.5	8.2	7.9	6.9

Πίνακας 5.2: Πίνακας αποτελεσμάτων με πίνακα διασποράς αρχικοποιημένο με τον Εσσιανό και σταθερός ρυθμός μάθησης

Ακριβώς την ίδια συμπεριφορά παρατηρούμε και στην περίπτωση που η αρχικοποίηση του πίνακα διασποράς είναι ίσος με τον μοναδιαίο πίνακα πολλαπλασιαζόμενο με μία μικρή ποσότητα της τάξεως του 10^{-4} (πίνακες 5.3 και 5.4). Και στις 2 περιπτώσεις όπου ο ρυθμός μάθησης είναι είτε μεταβλητός είτε σταθερός, βλέπουμε ότι ο αριθμός των τοπικών αναζητήσεων αυξάνεται αντί να μειώνεται.

	Number of Local Searches					
Function	Uni	Norm5	Norm10	Norm20	Norm50	Norm100
Rast	5940.2	7055.3	5659.5	6105.7	5908.7	7210.6
Tube	872.1	814.5	997.7	1089.8	908.7	892.7
Bird	189.2	471.9	292.2	263.8	393.2	331.4
Griew	1848.9	1602.2	1875.7	1667.1	1686.5	1806.9
Six-Hump	221.5	232	203.1	258.1	231	108.7
Bra	22.8	21.6	23.7	22	27.8	24.9
Hyper	14.4	13.9	14.4	13.4	14.1	13.1
Matyas	7.7	7	8.2	7.7	7.2	7.5

Πίνακας 5.3: Πίνακας διασποράς αρχικοποιημένος με τον μοναδιαίο και μεταβλητός ρυθμός μάθησης

Function	Number of Local Searches					
	Uni	Norm5	Norm10	Norm20	Norm50	Norm100
Rast	5153.3	6248.7	7255.8	5523.7	6709.1	5764.6
Tube	766.3	913	726.5	1026.2	1169.1	925.2
Bird	225.5	280.3	287.9	249.4	384.1	288.1
Griew	1815.5	1774.3	1903.7	1674.7	1772.5	1749.4
Six-Hump	234.4	288.4	280.3	224	236.4	196.1
Bra	21.7	22.8	22.6	22.9	23.6	24.7
Hyper	14.1	13.9	14	13.9	14.3	13
Matyas	7.6	7.7	8.2	7.7	7.8	6.9

Πίνακας 5.4: Πίνακας διασποράς αρχικοποιημένος με τον μοναδιαίο και σταθερός ρυθμός μάθησης

Όπως βλέπουμε η μέθοδος δειγματοληψίας που προτείνουμε υπερέρχει όταν ο πίνακας διασποράς αρχικοποιείται με τον Εσσιανό πίνακα και ο ρυθμός μάθησης είναι μεταβλητός. Σε οποιαδήποτε άλλη περίπτωση ο αλγόριθμος δεν αποδίδει το ίδιο καλά με τον Multistart. Αυτό οφείλεται στο γεγονός ότι κάθε συνάρτηση έχει διαφορετική μορφή και πλήθος ελαχίστων, οπότε οι παράμετροι που αφορούν τον πίνακα διασπορας και τον ρυθμό μάθησης πρέπει να εκτιμηθούν πειραματικά. Σε συναρτήσεις όπου το πλήθος των ελαχίστων είναι μικρό κάτι τέτοιο δεν αποτελεί πρόκληση. Όταν όμως η διάσταση και το πλήθος ελαχίστων του προβλήματος αυξάνουν, μια τέτοια εκτίμηση κρίνεται δύσκολη και ασύμφορη.

ΚΕΦΑΛΑΙΟ 6

ΕΠΙΛΟΓΟΣ

Η παρούσα διατριβή ασχολήθηκε με την καθολική ελαχιστοποίηση συναρτήσεων. Χρησιμοποιήσαμε μία καινούργια μέθοδο δειγματοληψίας σημείων από κανονικές κατανομές. Περιγράφεται μια διαδικασία που είναι υλοποιημένη σε γλώσσα C και χρησιμοποιήθηκε το περιβάλλον βελτιστοποίησης Merlin. Όπως μπορεί κάποιος να διαπιστώσει, από τα πειράματα, η μέθοδος υπερέχει σε σύγκριση με την μέθοδο πολλαπλής εκκίνησης.

Κατά τη διάρκεια εκπόνησης της παρούσας εργασίας ανακαλύψαμε αρκετές καινοτομίες που μπορούν να χρησιμοποιηθούν για μελλοντικές εργασίες. Η χρήση κατανομών είναι ένα από τα σημαντικότερα σημεία καθώς μας δίνεται η δυνατότητα χρήσης εναλλακτικών πλην της κανονικής κατανομής. Επίσης η ενημέρωση των παραμέτρων της κατανομής γίνονται *on line*, με αποτέλεσμα οι παράμετροι της κανονικής κατανομής να αλλάζουν κατά τη διάρκεια του αλγορίθμου και έτσι η επιλογή σημείων για αναζήτηση τοπικών ελαχίστων να γίνεται με βάση τα ήδη υπάρχοντα.

Όπως αναφέραμε μέσω Cholesky decomposition καταφέραμε να έχουμε ενημέρωση των παραμέτρων πιθανοφάνειας σε $O(n^2)$.

Ένα ακόμα σημαντικό στοιχείο είναι η δυνατότητα χρήσης αρχικοποιημένου Εσσιανού πίνακα για αρχική προσέγγιση και επιλογή μεταξύ μεταβλητού και σταθερού ρυθμού μάθησης. Τελειώνοντας, δεν πρέπει να παραλείψουμε την χρήση του *inverse sampling* καθώς μέσα από αυτή τη διαδικασία είμαστε σε θέση να εξάγουμε ένα κριτήριο τερματισμού.

Οι μελλοντικές επεκτάσεις στο συγκεκριμένο θέμα είναι πολλές, καθώς η καθολική βελτιστοποίηση αποτελεί ένα σημαντικό αντικείμενο έρευνας. Η βελτίωση της *on line* ενημέρωσης των παραμέτρων θα μπορέσει να δώσει περαιτέρω αύξηση στην απόδοση του αλγορίθμου. Επίσης περαιτέρω μελέτη μπορεί να γίνει για την επιλογή κριτηρίων τερματισμού (όπως επιλογή καλύτερου *threshold* για το πλήθος αποτυγχημένων δοκιμών), με βάση την κανονική κατανομή.

ΒΙΒΛΙΟΓΡΑΦΙΑ

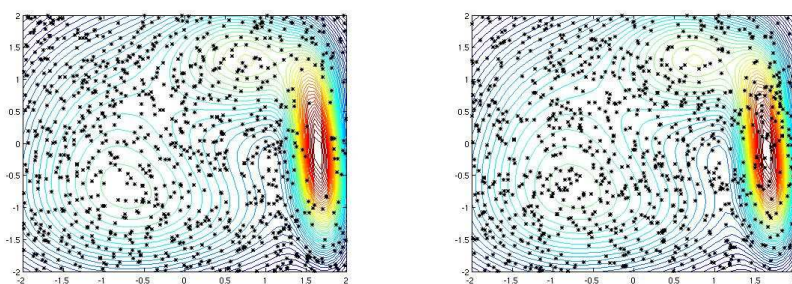
- [1] D.G..Papageorgiou, I.N.Demitropoulow, I.E.Lagaris. "Merlin-3.0 A Mutlidimensional optimixation enviroment". *Comp.Phys.Commun.*:227-249, 1998.
- [2] D.G..Papageorgiou, I.N.Demitropoulow, I.E.Lagaris. "The Merlin Control Language for strategic optimization". *Comp.Phys.Commun.*:250-270, 1998.
- [3] A.H.G. Rinnoy KAn, G.T Timmer. "Global Optimization". G.L. Nemhauser et al., Eds, *Handbooks in OR&MS*, Vol.1, 1982
- [4] A.H.G. Rinnoy Kan, G.T Timmer."Stochastic Methods for Global Optimization", *American Journal of Mathematical and Management Scieances*, 1983
- [5] F.Glover."Tabu Search - Part I". *ORSA Journal on Computing*2(1);4-34,1989
- [6] F.Glover."Tabu Search - Part II". *ORSA Journal on Computing*2(1);4-34,1990
- [7] D.E. Goldberg. "Genetic Algorithms in search optimization and Machine Learning". New-York, Addison-Wesley, 1989
- [8] S. Kiprkpatrick, C.D.Gelatt and M.P.Vecchi. "Optimization by simulated annealing". *Science* 220:671-680, 1983
- [9] P.J.M van Laarhoven and E.H.L. Aarts. "Simulated Annealing:Theory and Applications". D. Riedel, Bonston, 1997
- [10] L. Ingber."Simulated Annealing: Practice versus theory". *J.Mathl.Comput.Modeling* 18:29-57,1993
- [11] W.L.PRICE "A controlled random search procedure for global optimization". In *Towras Global Optimization2*, L.C.W.Dixon and G.P.Szego (eds.), North Holland, Amsterdam, 1974
- [12] R. Zieliński, A statistical estimate of the structure of multiectremal problems, *Mathemetical Programming* 21 (1981), 348-356
- [13] B. Betró and F. Schoen, Optimal and sub-optimal stopping rules for the Multistart algorithm in global optimization, *Mathematicl Programming* 57 (1992), 445-458

- [14] C.G.E. Boender, A.H.G. Rinnoy Kan, Bayesian stopping rules for multistart global optimization methods, *Mathematical Programming* 37 (1987), 469-480
- [15] C.G.E. Boender, A.H.G. Rinnoy Kan, Stochastic methods, in *Handbook of Global Optimization* (Horst R. and Pardalos P.M. eds.), Kluwer, Dordrecht (1995), 829-871
- [16] Hart W.E., Sequential stopping rules for random optimization methods with application to multistart local search, *Siam J. Optim* 9(1998), 270-290
- [17] C.G.E. Boender, A.H.G Rinnoy Kan, Bayesian stopping rules for multistart global optimization methods, *Mathematical Programming* 37(1987), 59-80)
- [18] M.M Ali and C. Storey, Topographical Multilevel Single Linkage, *Journal of Global Optimization* 5 (1994), 349-358
- [19] Towards Ideal Multistart: A stochastic approach for locating all local minima inside bounded domain

ΚΕΦΑΛΑΙΟ 7

ΠΑΡΑΡΤΗΜΑ

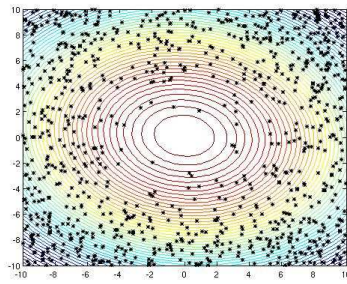
Παρακάτω βλέπουμε τα αποτελέσματα από πειράματα δειγματοληψίας που πραγματοποιήσαμε στις συναρτήσεις Goldstein, Booth, Sphere, Shubert.



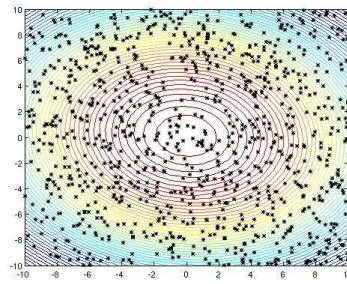
(a) *Goldstein με κανονική κατανομή*

(b) *Goldstein με χρήση ομοιόμορφη κατανομή*

Σχήμα 7.1: Δειγματοληψία στη συνάρτηση Goldstein με χρήση ομοιόμορφης και κανονικής κατανομή

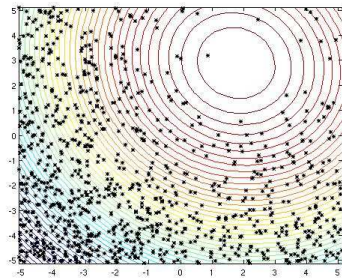


(a) *Booth* με κανονική κατανομή

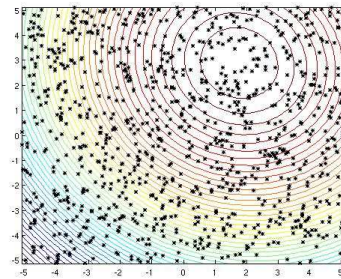


(b) *Booth* με χρήση ομοιόμορφης κατανομής

Σχήμα 7.2: Δειγματοληψία στη συνάρτηση Booth με χρήση ομοιόμορφης και κανονικής κατανομής

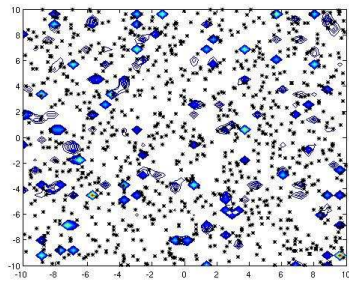


(a) *Sphere* με κανονική κατανομή

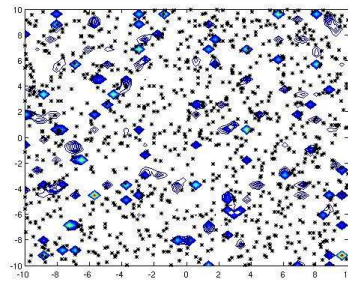


(b) *Sphere* με χρήση ομοιόμορφης κατανομής

Σχήμα 7.3: Δειγματοληψία στη συνάρτηση Sphere με χρήση ομοιόμορφης και κανονικής κατανομής



(a) *Shubert με κανονική κατανομή*



(b) *Shubert με χρήση ομοιόμορφη κατανομή*

Σχήμα 7.4: Δειγματοληψία στη συνάρτηση Shubert με χρήση ομοιόμορφης και κανονικής κατανομής

```

αρχικοποίηση objfun, nom, xl, xu, k, init_hess, Xmin = record {y,
yval, Count, Mean, Sigma} while(κριτήριο_τερματισμού)
  if i<k
    x <- uniform(xl, xu) /* Κάνε πρώτα k δειγματοληψίες από ομοιόμορφη */
  else
    mx <- max_of_gaussians(Xmin);
    x <- sample_of_gaussians(Xmin, mx, xl, xu);
  end
[y, yval, fcall, hessian] <- LocalSearch(objfun, x, xl, xu); /* Εκτέλεσε τοπική ελαχιστοποίηση */
im <- new_minimum(y, Xmin, eps); /* Έλεγχος αν βρέθηκε νέο ελάχιστο */
if im == -1
  nom = nom + 1
  Xmin(nom).Count = 1
  Xmin(nom).Mean = (x + y)/2;
  if init_hess == 1
    Xmin(nom).Sigma = 1/det(hessian) * hessian;
  else
    Xmin(nom).Sigma = alpha * eye(n);
  end
else
  Xmin(im).Count = Xmin(im).Count + 1;
  if const_rate == 1
    a <- const_a
  else
    a <- (1/Xmin(im).Count)
  Xmin(im).Mean = Xmin(im).Mean + a * (x - Xmin(im).Mean);
  Xmin(im).Sigma = Xmin(im).Sigma + a * ((x - Xmin(im).Mean)*(x - Xmin(im).Mean)');
end
end

```

Αλγόριθμος 7: Αλγόριθμος Δειγματοληψίας με χρήση κανονικής πολυδιάστατης κατανομής

Συναρτήσεις Για την εκτέλεση πειραμάτων χρησιμοποιήθηκαν οι παρακάτω συναρτήσεις, που αποτελούν ένα αντιπροσωπευτικό δείγμα για δοκιμές σε προβλήματα ελαχιστοποίησης:

1. Rastrigin

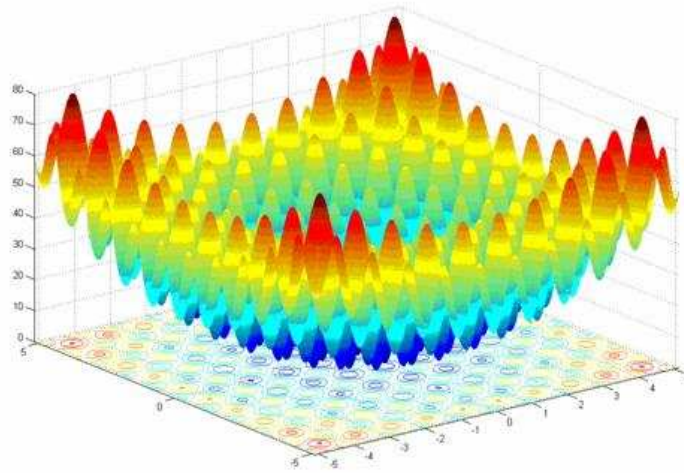
Αριθμός μεταβλητών: n

Συνάρτηση:

$$f(x) = 10n + \sum_{i=1}^n (x_i^2 - 10\cos(2\pi x_i))$$

Πεδίο τιμών: $x_i \in [-5.12, 5.12]$

Ολικό ελάχιστο: $x^* = (0, \dots, 0), f(x^*) = 0$



Σχήμα 7.5: *Rastrigin function*

2. Shubert

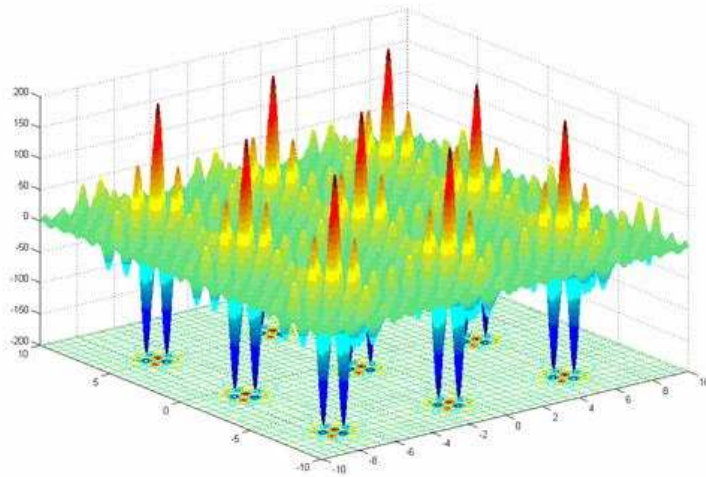
Αριθμός μεταβλητών: 2

Συνάρτηση:

$$f(x) = \left(\sum_{i=1}^5 \text{icos}((i+1)x_1 + i) \right) \left(\sum_{i=1}^5 \text{icos}((i+1)x_2 + i) \right)$$

Πεδίο τιμών: $x_i \in [-10.0, 10.0], i = 1, 2, \dots, n$

Ολικό ελάχιστο: $f(x^*) = -186.7309$



Σχήμα 7.6: *Shubert function*

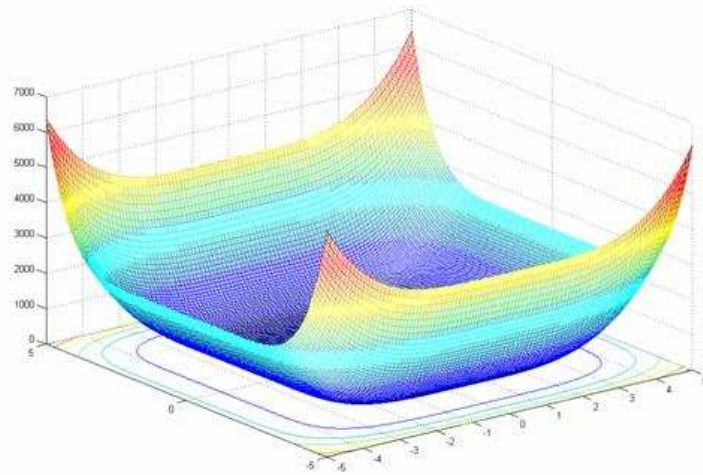
3. Hump

Αριθμός μεταβλητών: 2

Συνάρτηση: $f(x) = 4x_1^2 + 2.1x_1^4 + \frac{1}{3}x_1^8 - 4x_2^2 + 4x_2^4$

Πεδίο τιμών: $x_i \in [-5, 5]$

Ολικό ελάχιστο: $x^* = (0.0898, -0.7126), (-0.0898, 0.7126), f(x^*) = 0$



Σχήμα 7.7: *Hump function*

4. Griewank

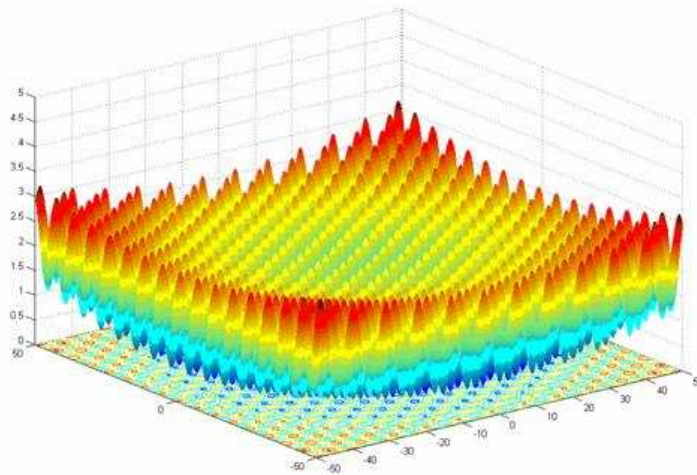
Αριθμός μεταβλητών: n

Συνάρτηση:

$$f(x) = \sum_{i=1}^n \frac{x_i^2}{4000} - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$$

Πεδίο τιμών: $x_i \in [-600, 600], i = 1, 2, \dots, n$

Ολικό ελάχιστο: $x^* = (0, \dots, 0), f(x^*) = 0$



Σχήμα 7.8: *Griewank function*

5. Michalevic

Αριθμός μεταβλητών: n

Συνάρτηση:

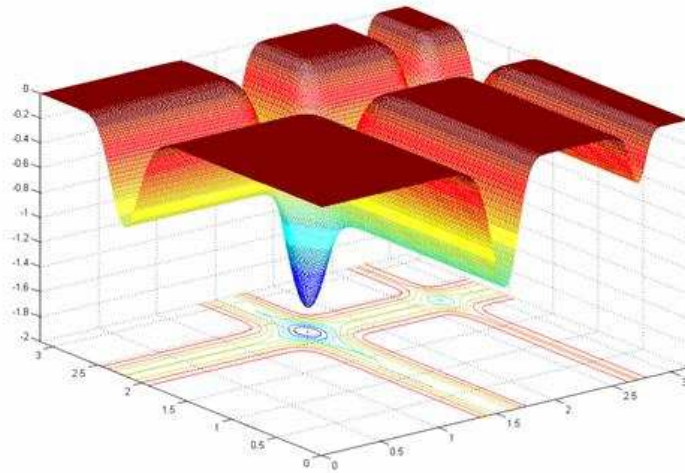
$$f(x) = - \sum_{i=1}^n \sin(x_i) \left(\sin\left(\frac{i x_i^2}{\pi}\right) \right)^{2m}, m = 10$$

Πεδίο τιμών: $x_i \in [0, \pi], i = 1, 2, \dots, n$

Ολικό ελάχιστο: για n=2, $f(x^*) = -1.8013$

για n=5, $f(x^*) = -4.687658$

για n=10, $f(x^*) = -9.66015$



Σχήμα 7.9: Michalevic function

6.Schwefel

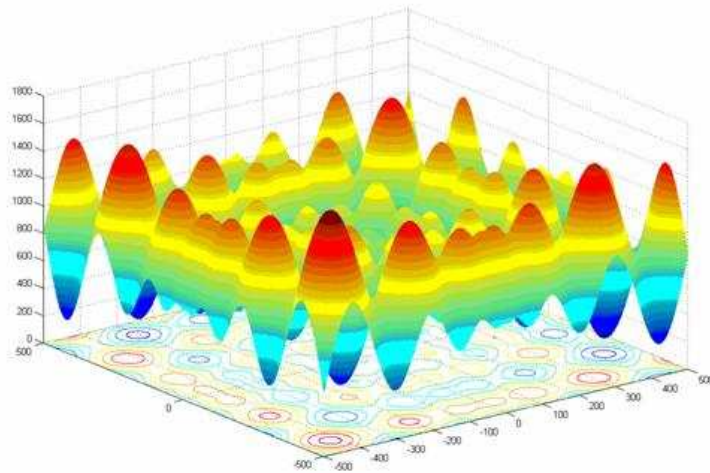
Αριθμός μεταβλητών: n

Συνάρτηση:

$$f(x) = 418.9820n - \sum_{i=1}^n (x_i \sin \sqrt{|x_i|})$$

Πεδίο τιμών: $x_i \in [-500, 500], i = 1, 2, \dots, n$

Ολικό ελάχιστο: $x^* = (1, \dots), f(x^*) = 0$



Σχήμα 7.10: Schwefel function

7. Goldstein & Price Function

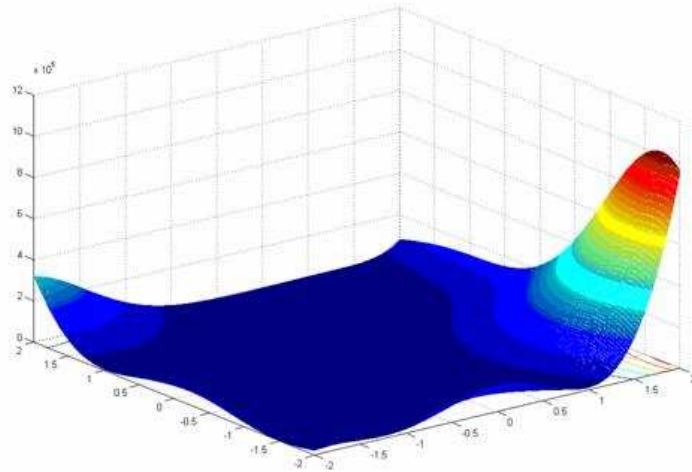
Αριθμός μεταβλητών: 2

Συνάρτηση:

$$f(x) = (1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 13x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2))(30 + (2x_1 - 3x_2)^2)(18 - 32x_1 + 12x_1^2 - 48x_2 - 36x_1x_2 + 27x_2^2)$$

Πεδίο τιμών: $x_i \in [-2, 2], i = 1, 2$

Ολικό ελάχιστο: $x^* = (0, 1), f(x^*) = 3$



Σχήμα 7.11: *Goldstein function*