

ΒΙΒΛΙΟΘΗΚΗ
ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΙΩΑΝΝΙΝΩΝ



026000265802



Αρ. εισ.:...216.....2001..

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
Τμήμα Πληροφορικής

40/01

219

ΜΠΛΕ



Μεταπτυχιακή εργασία ειδίκευσης με θέμα:
«Μέθοδοι ολικής ελαχιστοποίησης»

Φώτης Θέος

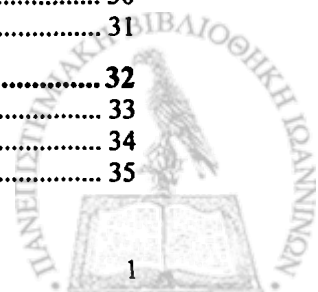
Επιβλέπων καθηγητής: Ισαάκ Η. Λαγαρής

☺ **Ιούνιος 2001**

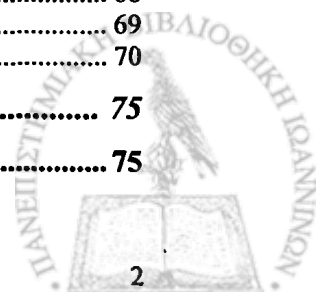


ΠΕΡΙΕΧΟΜΕΝΑ

Πρόλογος	4
Στόχος της εργασίας	5
Συνοπτική περιγραφή των περιεχομένων	6
Κεφάλαιο 1 – Γενικά στοιχεία ολικής ελαχιστοποίησης	8
1.1 Ορισμοί τοπικού κι ολικού ελαχίστου	8
1.2 Κριτήρια εύρεσης ολικού ελαχίστου	8
1.3 Γενικά χαρακτηριστικά μεθόδων ολικής ελαχιστοποίησης	9
1.4 Κατηγορίες μεθόδων ολικής ελαχιστοποίησης	11
1.4.1 Αιτιοκρατικές (Deterministic) μέθοδοι.....	11
1.4.2 Στοχαστικές μέθοδοι	11
1.4.3 Σύγκριση των μεθόδων των δύο κατηγοριών	12
Κεφάλαιο 2 – Στοχαστικές μέθοδοι και ολική ελαχιστοποίηση	13
2.1 Τυχαία αναζήτηση (Random search)	13
2.2 Πολλαπλή εκκίνηση (Multistart)	13
2.3 Μέθοδοι Ομαδοποίησης (Clustering)	14
2.4 Simulated annealing	15
2.5 Γενετικοί αλγόριθμοι	15
2.6 Αναζήτηση με απαγορεύσεις (Tabu search)	15
Κεφάλαιο 3 - Ελεγχόμενη τυχαία αναζήτηση (controlled random search)	17
3.1 Διάφοροι αλγόριθμοι τυχαίας αναζήτησης	17
3.2 Ο αλγόριθμος τυχαίας αναζήτησης του Price	18
3.3 Τα βήματα του αλγορίθμου	19
3.4 Βελτιωμένη έκδοση του αλγορίθμου του Price	21
3.5 Περιγραφή παραμέτρων του αλγορίθμου	21
3.5.1 Η χρήση ενός ζυγισμένου κεντροειδούς	21
3.5.2 Υπολογισμός του νέου σημείου με χρήση βαρών	22
3.5.3 Περιγραφή του αλγορίθμου	23
3.6 Πειραματικά αποτελέσματα	24
Κεφάλαιο 4 - Simulated annealing	27
4.1 Εισαγωγή	27
4.2 Θεωρητική μελέτη	28
4.2.1 Παρουσίαση του αλγορίθμου	28
4.2.2 Παρουσίαση διαφόρων αλγορίθμων μείωσης της θερμοκρασίας	29
4.2.3 Κλασσικό Simulated Annealing – Θεωρητική μελέτη	29
4.2.4 Θεωρητική μελέτη του Fast Simulated Annealing	30
4.2.5 Θεωρητική μελέτη του Very Fast Simulated Annealing	31
4.3 Περιγραφή του υλοποιημένου αλγορίθμου	32
4.3.1 Περιγραφή αλγορίθμου.....	33
4.3.2 Σχόλια - παρατηρήσεις.....	34
4.3.3 Παρουσίαση ενός νέου αλγορίθμου.....	35



4.4 Συνδυάζοντας τον αλγόριθμο του Simulated annealing με μια μέθοδο τοπικής ελαχιστοποίησης	38
4.4.1 Σύγκριση Simulated Annealing με μια μέθοδο τοπικής ελαχιστοποίησης	38
4.4.2 Περιγραφή του αλγορίθμου SALT	38
4.5 Πειραματικά αποτελέσματα.....	40
Κεφάλαιο 5 – Γενετικοί αλγόριθμοι.....	41
5.1 Εισαγωγή	41
5.2 Δομή του γενετικού αλγορίθμου	41
5.3 Αναπαράσταση μελών της γενιάς	42
5.4 Περιγραφή αλγορίθμου	43
5.5 Επιλογή (Selection)	43
5.6 Διασταύρωση (crossover) - Real-valued recombination	45
5.6.1 Διακριτή διασταύρωση	45
5.6.2 Ενδιάμεση διασταύρωση	46
5.7 Μετάλλαξη (mutation).....	47
5.8 Υβριδική μέθοδος.....	49
5.8.1 Γενική ιδέα.....	49
5.8.2 Θέματα προς συζήτηση για τις υβριδικές μεθόδους	50
5.8.3 Η Δαρβινική προσέγγιση	50
5.8.4 Η Λαμαρκιανή προσέγγιση.....	51
5.8.5 Σύγκριση των δύο παραπάνω προσεγγίσεων.....	51
5.8.6 Περιγραφή υλοποιημένου αλγορίθμου	52
5.9 Πειραματικά αποτελέσματα.....	53
5.10 Γενικά σχόλια και συμπεράσματα	53
Κεφάλαιο 6 - Μέθοδοι βασισμένες σε ομαδοποίηση	55
6.1 Γενικά.....	55
6.2 Περιγραφή μεθόδου ομαδοποίησης.....	56
6.3 Σχηματική αναπαράσταση του αλγορίθμου	58
6.4 Πώς γίνεται η ομαδοποίηση.....	59
6.5 Πειραματικά αποτελέσματα.....	60
Συγκεντρωτικά συγκριτικά αποτελέσματα στοχαστικών μεθόδων	61
Κεφάλαιο 7 - Γενικά στοιχεία για τις αιτιοκρατικές (deterministic) μεθόδους.....	63
7.1 Εισαγωγή	63
7.2 Μέθοδοι εύρεσης των ορίων των τιμών μιας συνάρτησης.....	63
7.2.1 Τεχνικές για τον υπολογισμό ορίων των διαστημάτων τιμών της συνάρτησης.....	64
7.2.2 Ένας απλός αλγόριθμος εύρεσης των ορίων μιας συνάρτησης.....	64
7.2.3 Ένας γενικός αλγόριθμος εύρεσης ορίων	65
7.3 Γενικά στοιχεία της αριθμητικής διαστημάτων	67
7.3.1 Διάφοροι ορισμοί.....	68
7.3.2 Βασικές πράξεις της αριθμητικής διαστημάτων	68
7.3.3 Το πρόβλημα της εξάρτησης	69
7.3.4 Συναρτήσεις διαστημάτων - Ιδιότητες.....	70
Κεφάλαιο 8 – Περιγραφή αλγορίθμων διαστήματος.....	75
8.1 Εισαγωγή	75



8.2 Περιγραφή του αλγορίθμου	75
8.2.1 Γενικό σχήμα	75
8.2.2 Περιγραφή κριτηρίου του μέσου	76
8.2.3 Έλεγχος κριτηρίου παραγωγού	77
8.2.4 Κριτήρια τερματισμού	77
8.2.5 Βελτιώσεις του παραπάνω αλγορίθμου	77
8.2 Αποτελέσματα υλοποίησης.....	79
Επίλογος	81
Παράρτημα	82
Αναφορές.....	84

Πρόλογος

Η ολική βελτιστοποίηση είναι ένα από τα σημαντικότερα θέματα έρευνας στον τομέα των εφαρμοσμένων μαθηματικών. Έχει πολύ σημαντικές εφαρμογές σε πολλά πεδία επιστημών, όπως οικονομία, χημεία, βιολογία και μηχανική. Η εύρεση μιας βέλτιστης τιμής έχει σε πολλές περιπτώσεις σημαντικές οικονομικές και κοινωνικές επιπτώσεις. Συνεπώς η επινόηση μεθόδων ολικής βελτιστοποίησης, που να μπορούν να επιλύουν όσο το δυνατόν μεγαλύτερο φάσμα προβλημάτων με μεγαλύτερη ακρίβεια κι αποδοτικότητα, είναι υψίστης σημασίας. Η εργασία αυτή αναφέρεται σε μια σειρά μεθόδων, που μπορούν να επιλύσουν δύσκολα προβλήματα ολικής ελαχιστοποίησης.

Για την περάτωση της εργασίας, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή κ. Ι.Η. Λαγαρή, για την συνεχή του καθοδήγηση κι αμέριστη συμπαράστασή του. Επίσης οφείλω ευχαριστίες στον κ. Δ. Παπαγεωργίου. Η βοήθειά του ήταν χρήσιμη κυρίως σε προβλήματα υλοποίησης, που αφορούν την γλώσσα MCL στο περιβάλλον βελτιστοποίησης MERLIN. Η βοήθεια του κ. Α. Λύκα ήταν επίσης καθοριστική, κυρίως στον τομέα των γενετικών αλγορίθμων και του simulated annealing.

Η συμπαράσταση των συμφοιτητών μου, ήταν σημαντική για την περάτωση αυτής της εργασίας. Οφείλω ιδιαίτερες ευχαριστίες στους μεταπτυχιακούς φοιτητές Κ. Βόγκλη, Ι. Τσούλο και Ι. Παπακωνσταντίνου. Τέλος αφιερώνω αυτή την εργασία στους γονείς μου, γιατί χωρίς την δική τους ηθική και υλική βοήθεια, η περάτωση αυτής της εργασίας δεν θα ήταν εφικτή.

Εισαγωγή

Πολλές εφαρμογές διαφόρων επιστημών, αντιμετωπίζονται ως προβλήματα βελτιστοποίησης. Τέτοιες εφαρμογές έχουμε κυρίως στην οικονομία, στη χημεία, στην ιατρική, στο μηχανικό σχεδιασμό και στην ψηφιακή επεξεργασία σήματος. Η εύρεση λοιπόν βέλτιστων λύσεων για τις παραπάνω εφαρμογές, έχει πολλές φορές σημαντικές οικονομικές και κοινωνικές συνέπειες. Για παράδειγμα η πραγματοποίηση του βέλτιστου σχεδιασμού ενός αεροσκάφους, έχει άμεσες συνέπειες στο χαμηλό κόστος κατασκευής του, όπως και η βελτιστοποίηση μιας συνάρτησης κόστους στην οικονομία μπορεί να επιφέρει σημαντικά ωφέλη.

Τρία είναι τα κύρια χαρακτηριστικά που συνθέτουν τα προβλήματα βελτιστοποίησης. Ένα σύνολο **αγνώστων ή μεταβλητών**, ο **τύπος της συνάρτησης** που θέλουμε να βελτιστοποιήσουμε κι ένα σύνολο **περιορισμών** που θα πρέπει να πληρεί η επιστρεφόμενη βέλτιστη λύση. Γενικά λοιπόν μπορούμε να ορίσουμε ένα πρόβλημα βελτιστοποίησης, ως την διαδικασία εύρεσης εκείνων των τιμών των παραμέτρων που ελαχιστοποιούν ή μεγιστοποιούν την συγκεκριμένη συνάρτηση.

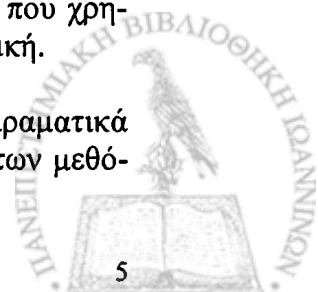
Υπάρχουν διάφοροι τύποι προβλημάτων βελτιστοποίησης. Οι μεταβλητές μπορεί να λαμβάνουν συνεχείς ή διακριτές τιμές. Η συνάρτηση βελτιστοποίησης μπορεί να είναι συνεχής ή διακριτή. Οι περιορισμοί στα διάφορα προβλήματα έχουν διαφορετική μορφή (γραμμικοί ή όχι), ή ακόμη μπορεί και να απουσιάζουν. Όλες αυτές οι παράμετροι συνθέτουν μια ποικιλία στα προβλήματα της 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.

Στόχος της εργασίας

Ο βασικός σκοπός αυτής της εργασίας είναι η μελέτη μεθόδων για την επίλυση του προβλήματος της ολικής ελαχιστοποίησης. Μελετήθηκαν στοχαστικές κι αιτιοκρατικές μέθοδοι. Δόθηκε μεγαλύτερη έμφαση στις στοχαστικές μεθόδους, γιατί είναι πιο αποτελεσματικές σε δύσκολα προβλήματα ολικής ελαχιστοποίησης, με πολλές παραμέτρους και πολλά τοπικά ελάχιστα. Οι αιτιοκρατικές μέθοδοι, αν κι εμφανίζουν το μεγάλο πλεονέκτημα, ότι τα αποτελέσματα που επιστρέφουν είναι εγγυημένα και σωστά, απαιτούν ισχυρούς υπολογιστικούς πόρους για την αντιμετώπιση δύσκολων προβλημάτων ελαχιστοποίησης.

Η υλοποίηση των στοχαστικών μεθόδων βασίζεται στο περιβάλλον βελτιστοποίησης MERLIN [1] και στην MCL [2]. Η MCL είναι μια γλώσσα προγραμματισμού του MERLIN, με την οποία μπορούμε να προγραμματίσουμε ειδικές στρατηγικές βελτιστοποίησης. Το βασικό πλεονέκτημα της υλοποίησης των μεθόδων ολικής ελαχιστοποίησης σε MCL είναι η χρησιμοποίηση διαφόρων ευκολιών που προσφέρει το περιβάλλον του MERLIN και η αξιοποίηση των συναρτήσεων τοπικής ελαχιστοποίησης. Κυρίως η υλοποίηση υβριδικών μεθόδων ολικής ελαχιστοποίησης σε MCL, που χρησιμοποιούν και τεχνικές τοπικής ελαχιστοποίησης, είναι πιο άμεση κι αποδοτική.

Περιγράφεται λοιπόν μια σειρά από μεθόδους, καθώς επίσης και διάφορα πειραματικά αποτελέσματα αυτών. Υπάρχουν αρκετά συγκριτικά αποτελέσματα μεταξύ των μεθόδων.



δων, από τα οποία μπορεί κανείς να εξάγει χρήσιμα συμπεράσματα. Οι υβριδικές μέθοδοι, είναι αυτές που έχουν δώσει τα καλύτερα αποτελέσματα. Τέλος, παρατίθενται και κάποια στοιχεία για τις αιτιοκρατικές μεθόδους διαστήματος (interval μέθοδοι). Η μελέτη των μεθόδων διαστήματος δεν είναι στις άμεσες προτεραιότητες της παρούσας διατριβής. Ωστόσο, για την πληρότητα της παρουσίασης του θέματος της ολικής ελαχιστοποίησης, κρίθηκε σκόπιμο να περιγραφούν έστω και συνοπτικά.

Συνοπτική περιγραφή των περιεχομένων

Η εργασία αυτή ασχολείται με το πρόβλημα της ολικής ελαχιστοποίησης. Χωρίζεται σε δύο μεγάλα τμήματα.

- Το πρώτο τμήμα περιλαμβάνει την περιγραφή *στοχαστικών μεθόδων* ολικής ελαχιστοποίησης. Υπάρχει μια συνοπτική περιγραφή των πιο διαδεδομένων μεθόδων, κι επίσης η αναλυτική περιγραφή των υλοποιημένων αλγορίθμων. Οι αλγόριθμοι που επιλέχθηκαν προς υλοποίηση, είναι εκείνοι που έχουν δώσει τα καλύτερα αποτελέσματα σε δύσκολα προβλήματα ολικής βελτιστοποίησης. Για την υλοποίηση χρησιμοποιήθηκε η γλώσσα **MCL (Merlin Control Language)**, που είναι γλώσσα προγραμματισμού στο περιβάλλον βελτιστοποίησης **MERLIN**. Το **MERLIN** [1] είναι ένα από τα πλέον αποδεκτά πακέτα λογισμικού στον τομέα της τοπικής ελαχιστοποίησης. Ο προγραμματισμός των αλγορίθμων σε **MCL** [2] εκμεταλλεύεται σε μεγάλο βαθμό την αξιοπιστία του **MERLIN** και καθιστά αρκετά εύκολη την υλοποίηση υβριδικών μεθόδων ολικής ελαχιστοποίησης.
- Το δεύτερο τμήμα περιλαμβάνει την περιγραφή *αιτιοκρατικών (deterministic)* μεθόδων. Περιγράφονται οι **μέθοδοι διαστήματος** που είναι και οι πιο δημοφιλείς εκπρόσωποι αυτής της κατηγορίας. Για την υλοποίηση αυτών των μεθόδων χρησιμοποιήθηκε η γλώσσα προγραμματισμού **FORTRAN 77** και μια βιβλιοθήκη με χρήσιμες συναρτήσεις της αριθμητικής διαστημάτων η **INTLIB**, που ο κώδικάς της είναι επίσης γραμμένος σε **FORTRAN77**.

Η παρούσα διατριβή έχει την παρακάτω δομή :

Κεφάλαιο 1 – Ορισμός ολικού και τοπικού ελαχίστου. Χαρακτηριστικά των αλγορίθμων ολικής ελαχιστοποίησης

Κεφάλαιο 2 – Στοχαστικές μέθοδοι. Συνοπτική περιγραφή των διαφόρων μεθόδων αυτής της κατηγορίας με αναφορές για πιο εκτενή μελέτη τους.

Κεφάλαια 3 – 6 - Αναλυτική περιγραφή των υλοποιημένων αλγορίθμων σε **MCL** καθώς επίσης κι αποτελέσματα αυτών σε συγκεκριμένα παραδείγματα

Κεφάλαιο 3 - Μέθοδος «ελεγχόμενης» τυχαίας αναζήτησης (Controlled Random Search)

Κεφάλαιο 4 – Simulated Annealing

Κεφάλαιο 5 – Γενετικοί αλγόριθμοι

Κεφάλαιο 6 – Μέθοδος βασισμένη στην ομαδοποίηση (clustering)

Κεφάλαιο 7 - Γενικά στοιχεία της αριθμητικής διαστημάτων



Κεφάλαιο 8 - Περιγραφή υλοποιημένων μεθόδων διαστήματος (interval μέθοδοι)

Κεφάλαιο 8 - 1 - 1 - γράφοι κτλ. για ολικούς ελαχίστους

1.1 Ορισμοί τοπικού κι ολικού ελαχίστου

Το σημείο x^* λέγεται τοπικό ελαχίστο μιας f εφόσον υπάρχει περιβάλλον δ τέτοιο ώστε $f(x) \geq f(x^*)$ για όλα τα σημεία x του περιβάλλοντος δ γύρω από το x^* .

Από τον ορισμό προκύπτει ότι x^* είναι τοπικό ελαχίστο

$$f(x) \geq f(x^*) \quad \forall x \in \delta$$

Αντιθέτως, είναι ολικό ελαχίστο η f για την \mathbb{R}^n εφόσον $f(x) \geq f(x^*)$ για όλα τα $x \in \mathbb{R}^n$. Στο σημείο αυτό γίνεται σκόδαμο να ερμηνεύσουμε την έννοια του τοπικού ελαχίστου, ώστε να διαφανεί η διαφορά του με το ολικό ελαχίστο.

Το ολικό ελαχίστο είναι ένα σημείο x^* και αν θεωρήσουμε K^* (όπου n η διάσταση του χώρου \mathbb{R}^n), τότε θα μπορούσαμε να υπάρχει μια περιοχή B του x^* με την ιδιότητα

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

Από τον παραπάνω ορισμό είναι προφανές η διαφορά μεταξύ ολικού ελαχίστου και τοπικού. Το τοπικό ελαχίστο αφορά μόνο το περιβάλλον δ του x^* . Αν υποθέσουμε ότι η περιοχή B του x^* είναι τοπικό ελαχίστο, τότε οι κενές συνθήκες για την έννοια του ολικού ελαχίστου είναι:

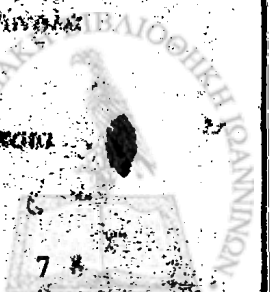
- Η f να μην έχει παραγώγο στο σημείο x^* να είναι μηδέν $(\nabla f(x^*) = 0)$
- Η f να είναι μηδέν στο σημείο x^* $(\nabla f(x^*) = 0)$, να είναι θετικό ημισφαιρικό

Οι παραπάνω συνθήκες για την έννοια του ολικού ελαχίστου δεν αρκούν για την επαβεβαιώνση ότι ένα σημείο αποτελεί κι ολικό ελαχίστο. Το πρόβλημα λοιπόν των ολικών ελαχίστων είναι το ερώτημα παραπάνω είναι άλυτο. Αν χρησιμοποιήσουμε μια μέθοδο με παραγόμενο αριθμό βήματων, μπορούμε όμως να λύσουμε πρόβλημα ολικής ελαχίστοποίησης. Η πιο καλή εφαρμογή του μεθόδου είναι στα προβλήματα ελαχίστου για όλο το σύνολο των πραγματικών αριθμών.

1.2 Κριτήρια εφικτής ολικού ελαχίστου

Οι μέθοδοι εφικτής ολικού ελαχίστου, αν χρησιμοποιήσουμε να μας δώσουν αριθμητικές λύσεις. Έτσι μπορούμε να ερμηνεύσουμε την έννοια του ολικού ελαχίστου της f ως λύση, αν υπάρχει, του προβλήματος $f(x) \geq f(x^*)$ για όλα τα $x \in \mathbb{R}^n$.

Από τον ορισμό προκύπτει ότι x^* είναι τοπικό ελαχίστο f αν και μόνο αν υπάρχει $\delta > 0$ τέτοιο ώστε $f(x) \geq f(x^*)$ για όλα τα $x \in \delta$.



Κεφάλαιο 1 – Γενικά στοιχεία ολικής ελαχιστοποίησης

1.1 Ορισμοί τοπικού κι ολικού ελαχίστου

Το πρόβλημα της ολικής βελτιστοποίησης είναι η εύρεση εκείνου του σημείου, έστω x^* , μιας συνάρτησης f με πραγματικές τιμές, στο οποίο η συνάρτηση παρουσιάζει την ελάχιστη τιμή.

Δηλαδή η εύρεση ενός σημείου x^* τέτοιου ώστε

$$f(x^*) \leq f(x) \quad \forall x \in \mathbb{R}^n.$$

Δυστυχώς, μόνο λίγες μέθοδοι έχουν μελετηθεί για την εύρεση ενός ελαχίστου σε σύγκριση με την πληθώρα των μεθόδων, που υπάρχουν για την εύρεση τοπικών ελαχίστων. Στο σημείο αυτό κρίνεται σκόπιμο να ορίσουμε την έννοια του τοπικού ελαχίστου, ώστε να διαφανεί η διαφορά του με το ολικό ελάχιστο.

Το ολικό ελάχιστο είναι ένα σημείο x^* που ανήκει στο \mathbb{R}^n (όπου n η διάσταση του σημείου x^*), τέτοιο ώστε να υπάρχει μια γειτονιά B του x^* με την ιδιότητα

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

Από τον παραπάνω ορισμό είναι φανερό η δυσκολία εύρεσης ολικού 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 ελαχίστων είναι :

- Η τιμή της παραγώγου στο σημείο x^* να είναι μηδέν ($\nabla f(x^*) = 0$)
- Η δεύτερη παράγωγος στο σημείο x^* ($\nabla^2 f(x^*)$), να είναι θετικά ορισμένη

Οι παραπάνω έλεγχοι για την εύρεση τοπικών ελαχίστων δεν αρκούν για την επιβεβαίωση ότι ένα σημείο αποτελεί κι ολικό ελάχιστο. Το πρόβλημα λοιπόν της ολικής ελαχιστοποίησης, όπως το ορίσαμε παραπάνω είναι άλυτο, αν χρησιμοποιήσουμε μια μέθοδο με πεπερασμένο αριθμό βημάτων. Μπορούμε όμως να λύσουμε προβλήματα ολικής ελαχιστοποίησης, όπου το πεδίο ορισμού των μεταβλητών είναι πεπερασμένο (π.χ δεν είναι όλο το σύνολο των πραγματικών αριθμών).

1.2 Κριτήρια εύρεσης ολικού ελαχίστου

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

$A_{x^*(\epsilon)} = \{ x \in S \mid \|x - x^*\| < \epsilon \}$ όπου $S \subset \mathbb{R}^n$ (το S είναι ένα σύνολο κυρτό με κάποια όρια)



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

Ένα άλλο μέτρο για την αξιοπιστία των αποτελεσμάτων των μεθόδων ολικής ελαχιστοποίησης δίνεται από τον παρακάτω τύπο :

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

Όπου m είναι το μέτρο Lebesgue (Lebesgue measure) και προκύπτει το σύνολο

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

Σημειώνουμε ότι το παραπάνω σύνολο μπορεί να περιέχει και σημεία, των οποίων η τιμή της συνάρτησης διαφέρει αρκετά από αυτή του y^* .

1.3 Γενικά χαρακτηριστικά μεθόδων ολικής ελαχιστοποίησης

Οι μέθοδοι ολικής ελαχιστοποίησης επιχειρούν ολικές και τοπικές αναζητήσεις στις διάφορες περιοχές ενδιαφέροντος, προσπαθώντας να διερευνήσουν το μεγαλύτερο μέρος του πεδίου ορισμού των μεταβλητών. Μια μέθοδος ολικής ελαχιστοποίησης έχει την ικανότητα να απεγκλωβίζεται από τοπικά ελάχιστα με διάφορους τρόπους, όπως θα δούμε 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 ολικής ελαχιστοποίησης συνδυάζουν τα πλεονεκτήματα της τοπικής ελαχιστοποίησης με την σφαιρική διερεύνηση του πεδίου ορισμού των παραμέτρων. Πολλοί αλγόριθμοι ολικής ελαχιστοποίησης απαιτούν μεγάλο υπολογιστικό χρόνο. Έτσι, πολλές φορές η υπολογιστική πολυπλοκότητα κάποιων προβλημάτων φτάνει σε επίπεδα απαγορευτικά. Ωστόσο στις μέρες μας, οι εφαρμογές *παράλληλων αλγορίθμων*, έχουν μειώσει σημαντικά τον απαιτούμενο χρόνο για την εύρεση ικανοποιητικής λύσης, με αντίτιμο τις αυξημένες ανάγκες σε εξοπλισμό (μηχανήματα μεγάλης υπολογιστικής ισχύος).

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

- **Αναπαράσταση του πεδίου αναζήτησης**

Το πεδίο αναζήτησης ορίζεται ως το πεδίο ορισμού των παραμέτρων του προβλήματος. Το πεδίο αναζήτησης μπορεί σε κάποια προβλήματα να είναι πεπερασμένο 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 κι αποτελείται από το πλήθος των πόλεων. Σε προβλήματα βελτιστοποίησης συνεχών συναρτήσεων, το πεδίο αναζήτησης μπορεί να είναι είτε άπειρο είτε πεπερασμένο. Η εργασία αυτή ασχολείται μόνο με προβλήματα συνεχών συναρτήσεων, όπου τα διαστήματα ορισμού των παραμέτρων είναι πεπερασμένα.

- **Στρατηγικές διαίρεσης (decomposition strategies)**



Πολλοί αλγόριθμοι εφαρμόζονται σε όλο το πεδίο ορισμού των παραμέτρων, ενώ άλλοι διαμερίζουν το σύνολο αυτό σε μικρότερα υποσύνολα, όπου κι εφαρμόζονται ξεχωριστά σε καθένα από αυτά. Η γνωστή τεχνική διακλάδωσης και φράγματος (**branch and bound**) χρησιμοποιεί την παραπάνω φιλοσοφία, απορρίπτοντας υποσύνολα για περαιτέρω αναζήτηση και υποδιαιρώντας κάποια άλλα, που ενδεχομένως περιέχουν την λύση του προβλήματος.

- **Πρόβλεψη κατεύθυνσης διερεύνησης στο πεδίο ορισμού των παραμέτρων**

Το θέμα της επιλογής κατάλληλης κατεύθυνσης αναζήτησης με τελικό στόχο το πραγματικό ολικό ελάχιστο είναι πολύ σημαντικό, αφού η εξαντλητική αναζήτηση του πεδίου ορισμού σε προβλήματα συνεχών συναρτήσεων κρίνεται απαγορευτική. Σε κάποιες μεθόδους για παράδειγμα, υπολογίζεται το σύνολο τιμών της συνάρτησης για συγκεκριμένο πεδίο ορισμού και προτιμούνται εκείνες οι περιοχές, που το κάτω άκρο του διαστήματος τιμών είναι μικρότερο. Το διάστημα αυτό υποδιαιρείται, ελέγχεται πάλι το κάτω άκρο του διαστήματος τιμών κι έτσι η μέθοδος κατευθύνεται με αυτόν τον τρόπο. Σε άλλες περιπτώσεις η τιμή της συνάρτησης του γειτονικού σημείου παρέχει πληροφορία για την κατεύθυνση αναζήτησης. Στους γενετικούς αλγορίθμους όπως θα δούμε στη συνέχεια, η τιμή καταλληλότητας κάθε μέλους 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.

- **Μηχανισμοί απεγκλωβισμού από τοπικά ελάχιστα**

Για την εύρεση ενός ολικού ελαχίστου, ο αλγόριθμος αναζήτησης θα πρέπει να έχει την ικανότητα απεγκλωβισμού από τα διάφορα τοπικά ελάχιστα. Για παράδειγμα η μέθοδος του *simulated annealing* δέχεται και σημεία που έχουν μεγαλύτερη τιμή συνάρτησης από την τρέχουσα ελάχιστη τιμή, ιδίως στα πρώτα βήματα του αλγορίθμου, ώστε να επιτύχει 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 καμιά σχέση με 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.

- **Κριτήρια τερματισμού**

Ορισμένα προβλήματα ελαχιστοποίησης μπορούν να επιλυθούν σε ικανοποιητικό χρόνο. Δυστυχώς όμως τα περισσότερα προβλήματα, κυρίως συναρτήσεων πολλών παραμέτρων, έχουν μεγάλη υπολογιστική πολυπλοκότητα, κάτι που κάνει την εύρεση βέλτιστων λύσεων αρκετά επίπονη και δύσκολη διαδικασία. Κρίνεται επομένως απαραίτητο οι αλγόριθμοι εύρεσης ολικού ελαχίστου να τερματίζουν μετά από κάποιο συγκεκριμένο χρονικό διάστημα, δίνοντας λύσεις ικανοποιητικής προσέγγισης. Ο βαθμός προσέγγισης της πραγματικής λύσης του ολικού ελαχίστου είναι τις περισσότερες φορές ανάλογος με τον χρόνο εκτέλεσης του αλγορίθμου. Τα κριτήρια τερματισμού είναι κάποιες φορές εξειδικευμένα και κατάλληλα μόνο για τον συγκεκριμένο αλγόριθμο, όπου χρησιμοποιούνται. Ωστόσο τα γενικότερα κριτήρια τερματισμού είναι: τέλος του αλγορίθμου μετά από ένα προκαθορισμένο πλήθος αποτιμήσεων της συνάρτησης (*function's evaluations*) ή τερματισμός όταν ο

βαθμός βελτίωσης της τρέχουσας ελάχιστης τιμής παραμένει σταθερός για κάποιο προκαθορισμένο αριθμό βημάτων του αλγορίθμου.

1.4 Κατηγορίες μεθόδων ολικής ελαχιστοποίησης

Οι μέθοδοι που έχουν αναπτυχθεί για την εύρεση στοιχείων των παραπάνω συνόλων (δηλαδή για την εύρεση ολικών ελαχίστων) διακρίνονται σε δύο μεγάλες κατηγορίες, ανάλογα με το αν χρησιμοποιούν ή όχι στοχαστικά στοιχεία. Έχουμε λοιπόν τις εξής δύο μεγάλες κατηγορίες μεθόδων:

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

1.4.1 Αιτιοκρατικές (Deterministic) μέθοδοι

Οι αιτιοκρατικές μέθοδοι [24], οι οποίες δεν έχουν καθόλου στοχαστικά στοιχεία, βρίσκουν το ολικό ελάχιστο, εφαρμόζοντας μια εξαντλητική αναζήτηση σε όλο το πεδίο ορισμού των μεταβλητών της συνάρτησης. Τα αποτελέσματα αυτών των μεθόδων είναι εγγυημένα, αλλά για να εφαρμοστούν πρέπει η συνάρτηση f να πληρεί κάποιες προϋποθέσεις. Η πιο δημοφιλής από αυτές τις προϋποθέσεις, που πρέπει να ικανοποιείται, είναι η εξής :

Δεδομένης κάποιας σταθεράς **Lipschitz L** για όλα τα x και x' πρέπει να ισχύει

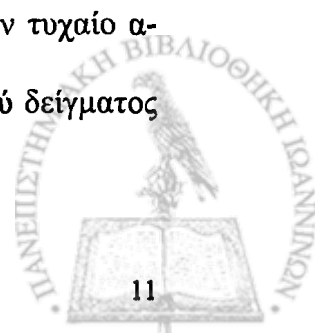
$$|f(x) - f(x')| \leq L \|x - x'\|$$

Η τιμή του άνω ορίου μεταβάλλεται (σε μικρότερες τιμές), ώστε να περιοριστεί η αναζήτηση σε συγκεκριμένες υποπεριοχές (χωρίς να υπάρχει κίνδυνος να χαθεί η υποπεριοχή, όπου υπάρχει η πραγματική λύση). Δυστυχώς όμως, η εξέταση αυτού του κριτηρίου σε επίπεδο υλοποίησης δεν είναι πάντα εύκολη κι εφικτή. 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 διαστήματος (*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.

1.4.2 Στοχαστικές μέθοδοι [4]

Η επόμενη αρκετά δημοφιλής κατηγορία μεθόδων κυρίως για προβλήματα πολλών μεταβλητών είναι οι στοχαστικές μέθοδοι. Οι στοχαστικές μέθοδοι, περιλαμβάνουν συνήθως δύο μεγάλες φάσεις.

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



1.4.3 Σύγκριση των μεθόδων των δύο κατηγοριών

Γενικά το μεγάλο μειονέκτημα των στοχαστικών μεθόδων σε σύγκριση με τις αιτιοκρατικές είναι ότι τα επιστρεφόμενα αποτελέσματα δεν είναι εγγυημένα. Παρόλα αυτά υπό κάποιες προϋποθέσεις, που αφορούν την κατανομή των σημείων του αρχικού δείγματος και την συνάρτηση f , η πιθανότητα να έχουμε εγγυημένες λύσεις για το πρόβλημα του ολικού ελαχίστου πλησιάζει την μονάδα, όσο αυξάνεται το πλήθος των σημείων του αρχικού δείγματος. Επιπλέον, πρέπει στο σημείο αυτό να παρατηρήσουμε, 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 την μονάδα. Υπάρχουν βέβαια και ειδικά κριτήρια τερματισμού, που σχετίζονται με την ιδιαίτερη φύση των μεθόδων και είναι κατάλληλα μόνο για αυτές, όπως αναφέρθηκε νωρίτερα.



Κεφάλαιο 2 – Στοχαστικές μέθοδοι και ολική ελαχιστοποίηση

Παρακάτω θα αναφερθούμε με συντομία στις πιο γνωστές από τις στοχαστικές μεθόδους και θα επιχειρήσουμε μια θεωρητική μελέτη αυτών, ώστε να διαφανεί η ποιότητα, ορθότητα και αξιοπιστία των αποτελεσμάτων τους. Στη συνέχεια θα δοθεί μεγαλύτερη έμφαση σε εκείνες από τις μεθόδους, οι οποίες έχουν δώσει και τα καλύτερα αποτελέσματα στην πλειοψηφία των συναρτήσεων για τις οποίες έγιναν δοκιμές. Οι συναρτήσεις που επιλέχθηκαν, είναι κυρίως δύσκολα προβλήματα ολικής ελαχιστοποίησης, με ορισμένες ιδιαιτερότητες, όπως θα εξηγήσουμε εκτενώς στη συνέχεια.

2.1 Τυχαία αναζήτηση (Random search)

Είναι ίσως η πιο απλή από τις μεθόδους ολικής ελαχιστοποίησης. Λαμβάνεται ένα αρχικό δείγμα σημείων και αποτιμάται η τιμή της συνάρτησης στα σημεία αυτά. Τα σημεία ικανοποιούν συνήθως την ομοιόμορφη κατανομή και υπάρχουν ισχυρές θεωρητικές αποδείξεις, ότι όσο αυξάνεται το πλήθος των σημείων του δείγματος, τόσο αυξάνεται και η πιθανότητα να βρούμε το ολικό ελάχιστο. Ωστόσο, γίνεται φανερό ότι ο χρόνος που χρειάζεται αυτή η μέθοδος για να μας δώσει ικανοποιητικό αποτέλεσμα, είναι πολύ αυξημένος. Σήμερα, μέθοδοι που ανήκουν σε αυτή την κατηγορία παρουσιάζουν ενδιαφέρον, μόνο στην περίπτωση συναρτήσεων με λίγα τοπικά ελάχιστα. Ενδιαφέρον παρουσιάζουν οι μέθοδοι της *ελεγχόμενης τυχαίας αναζήτησης (controlled random search)*, όπου το νέο σημείο στο οποίο αποτιμάται η συνάρτηση, προκύπτει μελετώντας την γνώση που έχουμε για τις τιμές της συνάρτησης από σημεία, που έχουμε ήδη εξετάσει. Ο πιο δημοφιλής εκπρόσωπος αυτής της κατηγορίας, με ικανοποιητικά αποτελέσματα ακόμη και σε δύσκολα πολυδιάστατα προβλήματα ολικής ελαχιστοποίησης, είναι ο *αλγόριθμος του Price* [9]. Ο αλγόριθμος αυτός θα περιγραφεί διεξοδικά στη συνέχεια.

2.2 Πολλαπλή εκκίνηση (Multistart)

Λόγω των σημαντικών μειονεκτημάτων που παρουσιάζει η προηγούμενη μέθοδος έχουν προταθεί διάφορες επεκτάσεις. Η μέθοδος λοιπόν που θα περιγράψουμε, αρχίζει επίσης από ένα τυχαίο αρχικό δείγμα σημείων, περιλαμβάνει όμως και διάφορες τοπικές αναζητήσεις. Πιο συγκεκριμένα, μερικά ή όλα τα σημεία του δείγματος θεωρούνται ως αρχικά σημεία για κάποιες μεθόδους τοπικής ελαχιστοποίησης. Αυτό λοιπόν προϋποθέτει την ύπαρξη κάποιας μεθόδου P τοπικής ελαχιστοποίησης, η οποία αρχίζοντας από ένα $x \in S$, επιστρέφει ένα τοπικό ελάχιστο x^* στην γειτονιά του x . Ανάλογα με τον τύπο της συνάρτησης, επιλέγουμε εκείνη την μέθοδο τοπικής ελαχιστοποίησης, που υποθέτουμε ότι θα δώσει τα πιο ακριβή αποτελέσματα. Για παράδειγμα στην περίπτωση συναρτήσεων που ο τύπος τους είναι άθροισμα τετραγώνων, κρίνεται πιο κατάλληλη η μέθοδος *Levenberg - Marquardt*.

Η πιο απλή μέθοδος ολικής ελαχιστοποίησης που συνδυάζει και τις δύο φάσεις που περιγράψαμε παραπάνω (*global and local phase*) είναι η μέθοδος *πολλαπλής εκκίνησης (Multistart)*. Εδώ, η μέθοδος P εφαρμόζεται σε όλα τα σημεία του δείγματος, και *το μικρότερο από τα τοπικά ελάχιστα* που επιστρέφονται μετά από διαδοχικές κλήσεις της P , αποτελεί την τιμή του ολικού ελαχίστου.

Ακόμη κι αν αυτή η μέθοδος φαίνεται να είναι πιο αποδοτική από την **τυχαία αναζήτηση**, υπάρχουν αρκετά μειονεκτήματα. Ας προσπαθήσουμε να μελετήσουμε αρχικά την ύπαρξη κατάλληλου κριτηρίου τερματισμού. Το πρόβλημα αυτό αντιμετωπίζεται μέσω μιας στοχαστικής προσέγγισης. Η προσέγγιση αυτή βασίζεται σε μια *Μπευζιανή μελέτη* που αφορά τον αριθμό των τοπικών ελαχίστων και το μέγεθος κάθε διαφορετικής **περιοχής προσέλκυσης (region of attraction)**. Η περιοχή προσέλκυσης $R(x^*)$ ορίζεται ως το σύνολο όλων των σημείων του S , τα οποία αν αποτελέσουν τα αρχικά σημεία για μια μέθοδο τοπικής ελαχιστοποίησης P , θα οδηγήσουν στο ίδιο **τοπικό ελάχιστο x^*** .

Αν γνωρίζουμε το πλήθος των τοπικών ελαχίστων της συγκεκριμένης συνάρτησης καθώς επίσης και το μέγεθος κάθε περιοχής ενδιαφέροντος (π.χ το μέτρο της ακτίνας αυτής της περιοχής, με κέντρο το αντίστοιχο τοπικό ελάχιστο), μπορούμε να αναλύσουμε ποιοτικά τα επιστρεφόμενα αποτελέσματα της μεθόδου πολλαπλής εκκίνησης. Βέβαια η γνώση αυτών των παραμέτρων δεν είναι δυνατό 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 εκκίνησης, αναπτύσσονται από τον Rinnooy Kann [3], [4].

Παρά το γεγονός όμως ότι η θεωρητική μελέτη της μεθόδου **πολλαπλής εκκίνησης** είναι εφικτή, η παραπάνω μέθοδος δεν μπορεί να θεωρηθεί αποδοτική. Η κύρια αιτία για αυτό είναι ότι το ίδιο τοπικό ελάχιστο αναπόφευκτα θα βρεθεί παραπάνω από μία φορές. Για την αποφυγή αυτής της κατάστασης θα έπρεπε να ξεκινάμε μια μέθοδο τοπικής ελαχιστοποίησης, όχι παραπάνω από μια φορά, σε κάθε *περιοχή προσέλκυσης (region of attraction)*.

2.3 Μέθοδοι Ομαδοποίησης (Clustering)

Οι μέθοδοι *ομαδοποίησης* [30] είναι από τους πιο αποτελεσματικούς κι αποδοτικούς αλγορίθμους, που έχουν προταθεί για ολική ελαχιστοποίηση. Η βασική ιδέα πίσω από αυτές τις μεθόδους είναι, ότι αρχίζοντας από ένα δείγμα ομοιόμορφα κατανεμημένων σημείων, δημιουργούμε *ομάδες* από σημεία, που ανήκουν στην ίδια περιοχή προσέλκυσης (region of attraction), έτσι ώστε να εφαρμόζεται η μέθοδος τοπικής ελαχιστοποίησης *μια μόνο φορά* σε μια τέτοια περιοχή.

Η διαδικασία της ομαδοποίησης γίνεται με διάφορους τρόπους. Η βασική ιδέα όμως παραμένει η ίδια. Σε κάθε ομάδα (cluster) υπάρχει πάντα ένα σημείο *πυρήνας (seed point)*. Ως πυρήνας μπορεί να θεωρηθεί το σημείο εκείνο του δείγματος που έχει την χαμηλότερη τιμή συνάρτησης, ή καλύτερα το τοπικό ελάχιστο που λάβαμε αρχίζοντας τοπική ελαχιστοποίηση από εκείνο το σημείο. Τα υπόλοιπα σημεία προστίθενται στις διάφορες *ομάδες (cluster)* με βάση κάποιον κανόνα για την *ομαδοποίηση*, μέχρι να ικανοποιηθεί κάποιο κριτήριο τερματισμού. Οι περισσότερες μέθοδοι διαφέρουν στην επιλογή αυτού του κανόνα.

Οι **Kan** και **Timmer** [29] περιγράφουν μια μέθοδο *ομαδοποίησης* και τις συνθήκες για τις οποίες όλα τα τοπικά ελάχιστα θα βρεθούν μετά από ένα πεπερασμένο αριθμό επαναλήψεων, με πιθανότητα μονάδα. 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 [11], [13], [15] έχει εμπνευστεί από την γνωστή διαδικασία του *annealing* στην φυσική για την απόκτηση καταστάσεων χαμηλής ενέργειας κι από τις διαδικασίες εύρεσης ελαχίστων λύσεων σε προβλήματα διακριτής βελτιστοποίησης. Αυτό που γίνεται σε γενικές γραμμές είναι: υπολογίζεται σε κάποιο σημείο x' η τιμή της συνάρτησης. Θεωρούμε ένα γειτονικό σημείο x'' και κάνουμε τον εξής έλεγχο :

Αν $f(x'') - f(x') < 0$, τότε δεχόμαστε το νέο σημείο. Αλλιώς το νέο σημείο το δεχόμαστε με πιθανότητα $\exp(- (f(x'') - f(x')) / T)$

Η θετική παράμετρος T είναι η *θερμοκρασία* η οποία μειώνεται με σταθερό ρυθμό, όπως γίνεται στην διαδικασία του *annealing*. Αρχικά δηλαδή 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 μεθόδου ολικής ελαχιστοποίησης. Όσο η θερμοκρασία μειώνεται, η αναζήτησή μας περιορίζεται σε πολύ μικρές περιοχές του πεδίου ορισμού των μεταβλητών. Σε πολύ χαμηλές θερμοκρασίες η αναζήτηση περιορίζεται σε μια πολύ μικρή περιοχή και είναι πλέον αδύνατο να απεγκλωβιστούμε από κάποιο τοπικό ελάχιστο. Είναι φανερό ότι όσο μειώνεται η θερμοκρασία, τόσο καλύτερα προσεγγίζουμε το ολικό ελάχιστο, οπότε ο εγκλωβισμός στην τελική φάση της μεθόδου, είναι επιθυμητός.

2.5 Γενετικοί αλγόριθμοι

Οι γενετικοί αλγόριθμοι [22] προέρχονται από την 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 καινούριων 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 ολικής ελαχιστοποίησης (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 μιας *ακολουθίας κινήσεων*. Η κίνηση που επιλέγεται από κάποιο σημείο σε κάποιο άλλο, είναι εκείνη που έχει αποδειχτεί από προηγούμενα βήματα, ότι μπορεί να οδηγήσει σε χαμηλότερες τιμές της συνάρτησης. Επίσης κατά την επιλογή αυτών των κινήσεων, κάποιες κινήσεις θεωρούνται *απαγορευμένες (tabu)*, ώστε να αποφεύγονται κύκλοι κι εγκλωβισμός σε τοπικά ελάχιστα. Κατά την εξέλιξη του αλγο-

ρίθμου κρατείται ένα ιστορικό με τις έως τώρα κινήσεις και τις λύσεις στις οποίες οδηγήσαμε με βάση αυτές. Σύμφωνα λοιπόν με αυτό το ιστορικό, μια κίνηση μπορεί να χαρακτηριστεί ως *απαγορευμένη (tabu)*, αν έχει γίνει αρκετές φορές ή σχετικά πρόσφατα. Επίσης για να εξασφαλίσουμε ότι δεν περιορίζομαστε σε ένα περιορισμένο αριθμό κινήσεων, σε κάποια στάδια του αλγορίθμου δεχόμαστε και κινήσεις, που έχουν χαρακτηριστεί ως *απαγορευμένες* [5], [6].

Κεφάλαιο 3 - Ελεγχόμενη τυχαία αναζήτηση (controlled random search)

Αν θεωρήσουμε ότι έχουμε επαρκή γνώση της συνάρτησης που θέλουμε να ελαχιστοποιήσουμε (για παράδειγμα γνωρίζουμε το πλήθος των τοπικών ελαχίστων της συνάρτησης), τότε μπορούμε να κάνουμε διαδοχικές κλήσεις μιας τοπικής μεθόδου γύρω από τα σημεία που περιμένουμε κατά προσέγγιση να έχουμε ελάχιστα. Έτσι, κρατάμε από τα ελάχιστα εκείνο που έχει την μικρότερη τιμή, με μεγάλη πιθανότητα να έχουμε βρει σωστή λύση για το ολικό ελάχιστο. Στην πιο συνηθισμένη περίπτωση όμως, όπου δεν υπάρχει επαρκής εκ των προτέρων γνώση της συνάρτησης, η αναζήτηση σε ολόκληρο το πεδίο ορισμού των παραμέτρων της, κρίνεται απαραίτητη. Έχοντας λοιπόν κάνει μια επαρκή αναζήτηση στο μεγαλύτερο μέρος του πεδίου ορισμού των παραμέτρων, μπορούμε στη συνέχεια να ξεκινήσουμε από το καλύτερο σημείο (αυτό με την μικρότερη τιμή συνάρτησης) μια τοπική μέθοδο, για να προσεγγίσουμε καλύτερα το ολικό ελάχιστο. Ωστόσο, μια εξαντλητική αναζήτηση στο πεδίο ορισμού των παραμέτρων, είναι κάτι ανέφικτο και θα απαιτούσε πολλές αποτιμήσεις της συνάρτησης, απαγορευτικές εξαιτίας του υπολογιστικού κόστους. Πρέπει λοιπόν η αναζήτηση να περιοριστεί με τέτοιο τρόπο, ώστε να καλύπτει ένα ευρύ φάσμα των περιοχών ενδιαφέροντος. Οι πιο διαδεδομένες κι απλές μέθοδοι, που δεν απαιτούν τον υπολογισμό παραγώγων της συνάρτησης, είναι οι μέθοδοι τυχαίας αναζήτησης.

3.1 Διάφοροι αλγόριθμοι τυχαίας αναζήτησης

Ο **Brooks** (1955) [7] έχει προτείνει πολλές τεχνικές τυχαίας αναζήτησης. Η πιο απλή μέθοδος τυχαίας αναζήτησης προτείνει την αποτίμηση της συνάρτησης σε ένα πλήθος σημείων τυχαία επιλεγμένων κι επιστρέφει ως πιθανό ελάχιστο το σημείο με την μικρότερη τιμή. Μια βελτίωση της παραπάνω τεχνικής είναι η *διαμέριση του συνόλου ορισμού των παραμέτρων* της συνάρτησης σε υποσύνολα ίδιου περίπου εύρους και επιλογή ενός τυχαίου σημείου σε καθένα από τα υποσύνολα αυτά. Η μέθοδος αυτή είναι αποδοτική στην περίπτωση συναρτήσεων, που δεν υπάρχει συσχέτιση των τιμών μεταξύ γειτονικών σημείων. Παρόλα αυτά όμως στις περισσότερες περιπτώσεις υπάρχει μια τέτοια συσχέτιση. Έτσι κρίνεται λογικό να εστιάσουμε την αναζήτησή μας στα υποσύνολα εκείνα, που έχουν τις χαμηλότερες τιμές συνάρτησης, ξεκινώντας από εκεί την τυχαία επιλογή σημείων για το αρχικό δείγμα.

Μια βελτιωμένη έκδοση των προηγούμενων μεθόδων προτείνουν οι **Becker** και **Lago** (1970) [8]. Η προτεινόμενη διαδικασία ξεκινά με τυχαία δειγματοληψία σημείων σε όλο το πεδίο ορισμού των παραμέτρων της συνάρτησης. Αντί όμως να κρατάμε μόνο το σημείο με την χαμηλότερη τιμή, κρατάμε έναν προκαθορισμένο αριθμό σημείων (εκείνα με τις χαμηλότερες τιμές συνάρτησης). Στη συνέχεια χρησιμοποιείται ένας αλγόριθμος (**mode-seeking algorithm**) για την κατάταξη των παραπάνω σημείων σε ομάδες (cluster). Τα όρια κάθε ομάδας αντιστοιχούν στα άκρα υποδιαστήματος του πεδίου ορισμού των παραμέτρων της συνάρτησης. Οι ομάδες αυτές ταξινομούνται με βάση την χαμηλότερη τιμή της συνάρτησης ανάμεσα στα σημεία της. Η διαδικασία επαναλαμβάνεται παίρνοντας νέα τυχαία σημεία από εκείνο το υποδιάστημα του πεδίου ορισμού των παραμέτρων, που αντιστοιχεί στην ομάδα με την χαμηλότερη τιμή συνάρτησης. Επίσης μπορούμε να επαναλάβουμε αυτή την διαδικασία και για τα σημεία των υπολοίπων ομάδων, επιλέγοντας ως δεύτερη ομάδα εκείνη, που περιέχει το σημείο με την αμέσως μεγαλύτερη τιμή συνάρτησης.

3.2 Ο αλγόριθμος τυχαίας αναζήτησης του Price

Ο αλγόριθμος του Price [9] είναι παρόμοιος με τον αλγόριθμο των *Becker και Lago*. Δεν απαιτείται η συνάρτηση να είναι παραγωγίσιμη κι επίσης γίνεται τυχαία αναζήτηση σε όλο το σύνολο ορισμού (και όχι σε υποσύνολα) των παραμέτρων. Διαφέρει στο ότι δεν χρησιμοποιεί ένα ξεχωριστό αλγόριθμο (*mode seeking algorithm*) για την δημιουργία ομάδων, αλλά συνδυάζει την τυχαία επιλογή σημείων και την ομαδοποίησή τους σε μια και μόνο διαδικασία.

Αρχικά ορίζεται το πεδίο ορισμού D των παραμέτρων της συνάρτησης κι επιλέγονται m τυχαία σημεία στο πεδίο ορισμού D . Ο προτεινόμενος αριθμός σημείων του δείγματος είναι $25 N$ (όπου N η διάσταση της συνάρτησης). Η συνάρτηση αποτιμάται σε καθένα από τα σημεία αυτά και οι θέσεις καθώς και οι αντίστοιχες τιμές της, αποθηκεύονται σε πίνακα. Σε κάθε επανάληψη του αλγορίθμου επιλέγεται τυχαία ένα νέο σημείο, από ένα σύνολο σημείων (υποσύνολο του αρχικού δείγματος σημείων), του οποίου η θέση είναι σχετική με την θέση των m σημείων του δείγματος. Αν η τιμή της συνάρτησης στο νέο σημείο είναι μικρότερη από την τρέχουσα μέγιστη τιμή, τότε το νέο σημείο αντικαθιστά το σημείο με την μέγιστη τιμή στο αρχικό σύνολο. Αν όμως είναι μεγαλύτερη, τότε το σημείο αυτό απορρίπτεται και επιλέγεται ένα σημείο από τυχαίο υποσύνολο του αρχικού δείγματος. Με την πάροδο του χρόνου, τα εναπομείναντα σημεία τείνουν να συγκεντρωθούν γύρω από το σημείο του ολικού ελαχίστου. Η πιθανότητα να συγκλίνει ο παραπάνω αλγόριθμος στο πραγματικό ολικό ελάχιστο, εξαρτάται από το πλήθος των σημείων του αρχικού δείγματος και την πολυπλοκότητα της συνάρτησης.

Επειδή ο σκοπός του παραπάνω αλγορίθμου είναι η εύρεση του ολικού ελαχίστου, η αναζήτηση στα διάφορα υποδιαστήματα του πεδίου ορισμού των παραμέτρων, είναι πιο σημαντική από την ταχύτητα σύγκλισης σε κάποιο ελάχιστο. Ο αλγόριθμος αυτός προσπαθεί να διερευνήσει όσο το δυνατόν μεγαλύτερο κομμάτι του πεδίου ορισμού των παραμέτρων, επιλέγοντας τυχαία τα υποσύνολα των πιθανών νέων σημείων από το αρχικό δείγμα. Σε κάθε βήμα επιλέγεται ένα υποσύνολο $N+1$ σημείων από το αρχικό δείγμα. Το πρώτο σημείο του υποσυνόλου επιλέγεται τυχαία ως σημείο *πυρήνας*. Το νέο σημείο που αποτιμάται η συνάρτηση, προκύπτει ως *συμμετρικό του πυρήνα ως προς το κεντροειδές των N σημείων του υποσυνόλου και δίνεται από τον παρακάτω τύπο:*

$$x^k = c^k - (x_{i_0}^k - c^k) \quad \begin{array}{l} \text{όπου } x^k \text{ το νέο σημείο,} \\ c^k \text{ το κεντροειδές των υπολοίπων σημείων του υποσυνόλου,} \\ x_{i_0} \text{ το σημείο πυρήνας του υποσυνόλου.} \end{array}$$

Το πλήθος των διαφορετικών τρόπων με τους οποίους μπορούμε να επιλέξουμε $N+1$ σημεία από τα αρχικά M σημεία είναι ο *συνδυασμός των M ανά $N+1$* . Επίσης πρέπει να λάβουμε υπόψη μας ότι ο πυρήνας του κάθε υποσυνόλου επιλέγεται τυχαία και επομένως το πλήθος των διαφορετικών σημείων είναι $(N + 1)$ επί το προηγούμενο πλήθος.

Ας υποθέσουμε ένα απλό παράδειγμα, όπου το αρχικό πλήθος σημείων είναι 6 και η διάσταση της συνάρτησης είναι 2. Το πλήθος των διαφορετικών νέων σημείων, που προκύπτουν από τους παραπάνω τύπους είναι 60.

Η θέση αυτών των σημείων καλύπτει ένα μεγάλο μέρος του πεδίου ορισμού των παραμέτρων της συνάρτησης και η επιλογή οποιουδήποτε από τα σημεία αυτά ως το νέο σημείο, είναι πιθανό να έχει ως αποτέλεσμα καλύτερη αναζήτηση, από μια εντελώς τυχαία αναζήτηση στο πεδίο ορισμού D των παραμέτρων. Γίνεται φανερό ότι όσο μεγαλώνει το

πλήθος των σημείων του αρχικού δείγματος, αυξάνεται περισσότερο το ποσοστό του πεδίου αναζήτησης σε σχέση με το αρχικό πεδίο ορισμού D .

Επίσης ο αλγόριθμος αυτός προσπαθεί να κρατήσει το *ρυθμό επιτυχίας* σταθερό γύρω στο 50%. Κάθε προσπάθεια θεωρείται επιτυχής, αν η τιμή της συνάρτησης για το νέο σημείο που επιλέγεται με τον παραπάνω τρόπο, είναι μικρότερη από την τρέχουσα μέγιστη τιμή της συνάρτησης. Ο ρυθμός επιτυχίας δίνεται λοιπόν από το *πηλίκο των επιτυχιών προς το πλήθος των συνολικών προσπαθειών*. Όταν ο ρυθμός αυτός πέσει κάτω από το 50% τότε το νέο σημείο επιλέγεται σύμφωνα με τον παρακάτω τύπο :

$$x^k = (c^k + x_{i_0}^k) / 2$$

Η θέση αυτών των σημείων καλύπτει ένα πολύ μικρό τμήμα του συνολικού διαστήματος ορισμού των παραμέτρων. Στα τελευταία στάδια του αλγορίθμου, τα σημεία αυτά συσσωρεύονται στην περιοχή του ολικού ελαχίστου.

3.3 Τα βήματα του αλγορίθμου

- **Βήμα 0**

Αρχικοποίηση του μετρητή επαναλήψεων ($k = 1$).

Δημιουργία αρχικού συνόλου σημείων $S^k = \{x_1^k, \dots, x_m^k\}$ (τα σημεία x_i^k , $i = 1, \dots, m$ επιλέγονται τυχαία στο πεδίο ορισμού D).

Υπολογισμός της τιμής της συνάρτησης στα σημεία του συνόλου S^k .

- **Βήμα 1**

Υπολογισμός των τιμών x_{\max}^k, x_{\min}^k και f_{\max}^k, f_{\min}^k (όπου f_{\max}^k και f_{\min}^k η μέγιστη και η ελάχιστη τιμή της συνάρτησης για τα σημεία του συνόλου S^k).

Εξέταση κριτηρίου τερματισμού

Αν ικανοποιείται

τότε, τέλος αλγορίθμου

αλλιώς, συνέχεια στο βήμα 2

Τέλος Αν

- **Βήμα 2**

Επιλογή $n+1$ τυχαίων σημείων (όπου n η διάσταση της συνάρτησης) από το δείγμα σημείων S^k . Τα σημεία αυτά συμβολίζονται $x_{i_0}, x_{i_1}, \dots, x_{i_n}$ για το σύνολο S^k , (όπου το i ισούται με τον αύξοντα αριθμό επανάληψης).

Υπολογισμός του κεντροειδούς των παραπάνω n σημείων

$$c^k = \frac{1}{n} \sum_{j=1}^n x_{ij}^k$$

Επιλογή ενός νέου σημείου x^k , σύμφωνα με τον ακόλουθο τύπο :

$$x^k = c^k - (x_{i_0}^k - c^k)$$

Αν το νέο σημείο είναι εκτός του πεδίου ορισμού της συνάρτησης,



τότε επανάληψη του βήματος 2,
αλλιώς συνέχεια στο βήμα 3
Τέλος Αν

• Βήμα 3

Αν $f(x^k) \geq f(x_{\max}^k)$ τότε

Αν ο ρυθμός επιτυχίας > 0.5

$$S^{k+1} = S^k$$

Αύξηση του μετρητή επανάληψης ($k=k+1$)

Επιστροφή στο βήμα 2

Αλλιώς

Επιλογή ενός νέου σημείου σύμφωνα με τον τύπο

$$x^k = \frac{c^k + x_{i0}^k}{2}$$

Αν το νέο σημείο είναι εκτός του πεδίου ορισμού της συνάρτησης,
τότε επανάληψη του βήματος 2,

αλλιώς,

Αν $f(x^k) < f(x_{\max}^k)$ τότε

$$S^{k+1} = S^k \cup \{x^k\} - \{x_{\max}^k\}$$

Αύξηση μετρητή επαναλήψεων ($k=k+1$).

Επιστροφή στο βήμα 1.

Αλλιώς, $S^{k+1} = S^k$

Αύξηση του μετρητή επανάληψης ($k=k+1$)

Επιστροφή στο βήμα 2

Τέλος Αν

Τέλος Αν

Τέλος Αν (Αν ο ρυθμός επιτυχίας > 0.5)

Τέλος Αν

• Βήμα 4

Αν $f(x^k) < f(x_{\max}^k)$ τότε

$$S^{k+1} = S^k \cup \{x^k\} - \{x_{\max}^k\}$$

Αύξηση μετρητή επαναλήψεων ($k=k+1$).

Επιστροφή στο βήμα 1.

Τέλος Αν

Σχόλια :

Το κριτήριο τερματισμού που χρησιμοποιείται είναι $f_{\max}^k - f_{\min}^k < \varepsilon$, όπου $\varepsilon = 10^{-6}$

Ο ρυθμός επιτυχίας (success rate) ορίζεται ως:

επιτυχίες (αν $f(x^k) \leq f(x_{\max}^k)$) / σύνολο προσπαθειών

3.4 Βελτιωμένη έκδοση του αλγορίθμου του Price

Η βελτιωμένη μορφή του αλγορίθμου [10] παρουσιάζει τις εξής καινοτομίες σε σχέση με τον προηγούμενο αλγόριθμο

- Το κεντροειδές δεν προκύπτει από τον μέσο όρο των σημείων του τυχαίου δείγματος, αλλά είναι ένα ζυγισμένο κεντροειδές. Στις πρώτες επαναλήψεις το κεντροειδές είναι περίπου ίσο με τον μέσο όρο των σημείων, όπως συνέβαινε και στον προηγούμενο αλγόριθμο. Μετά όμως από κάποιες επαναλήψεις στον υπολογισμό του κεντροειδούς, δίνεται μεγαλύτερη βαρύτητα σε εκείνα τα σημεία που βρίσκονται πιο κοντά στην τρέχουσα ελάχιστη τιμή της συνάρτησης.
- Το νέο σημείο επιλέγεται πιο κοντά στην γειτονιά του πυρήνα (του τυχαίου υποσυνόλου $N+1$ σημείων από το αρχικό δείγμα), απ'ότι στην γειτονιά του κεντροειδούς, αν η τιμή της συνάρτησης στο σημείο του πυρήνα είναι μικρότερη από την τιμή της συνάρτησης στο κεντροειδές κι αντίστροφα.

Παρακάτω θα δώσουμε τους μαθηματικούς τύπους που επιστρέφουν το ζυγισμένο κεντροειδές και το νέο σημείο εξηγώντας τη φυσική σημασία τους.

3.5 Περιγραφή παραμέτρων του αλγορίθμου

3.5.1 Η χρήση ενός ζυγισμένου κεντροειδούς (weighted centroid)

Το κεντροειδές αυτό δίνεται από τον παρακάτω τύπο

$$c_w^k = \sum_{j=1}^n w_j^k x_{ij}^k \quad (3.5.1.1)$$

$$\text{όπου } w_j^k = n_j^k / \sum_{j=1}^n n_j^k \quad (3.5.1.2)$$

$$n_j^k = 1 / (f(x_{ij}^k) - f_{\min}^k + \varphi^k) \quad (3.5.1.3)$$

$$f_{\min}^k = \min f(x) \quad (x \in S)$$

όπου x_{ij}^k είναι η τιμή της παραμέτρου x_j του i -οστού στοιχείου του δείγματος σημείων στην k -οστή επανάληψη του αλγορίθμου (όπως θα δούμε παρακάτω),

όπου $f(x_{ij}^k)$ η τιμή της συνάρτησης στο παραπάνω σημείο.

- Η ακολουθία φ^k είναι μια ακολουθία από θετικούς αριθμούς τέτοια ώστε $\varphi^k \gg f(x_{ij}^k) - f_{\min}^k$ ($j = 1, \dots, n$) στις αρχικές επαναλήψεις. Το φ τείνει στο μηδέν όσο αυξάνεται το πλήθος των επαναλήψεων.

Μια καλή επιλογή για την συνάρτηση φ είναι

$$\varphi^k = \omega (f_{\max}^k - f_{\min}^k)^2 / (f_{\max}^0 - f_{\min}^0)$$

όπου ω είναι ένας θετικός αριθμός αρκετά μεγάλος (θέτουμε $\omega=1000$) και



$$f_{\max}^k = f(x_{\max}^k).$$

Στις αρχικές επαναλήψεις το ζυγισμένο κεντροειδές (weighted centroid) είναι περίπου το ίδιο με το κεντροειδές που προκύπτει από έναν απλό μέσο όρο των τιμών, δηλαδή

$$c^k = (1/n) \sum_{j=1}^n x_{ij}^k \quad \text{όπου } n \text{ το πλήθος των σημείων του δείγματος}$$

Όταν το πλήθος των επαναλήψεων είναι αρκετά μεγάλο οι τιμές των φ^k είναι αρκετά μικρές, και οι συντεταγμένες των βαρών w_j^k έχουν τέτοιες τιμές, ώστε τα σημεία εκείνα που η τιμή τους είναι πιο κοντά στην έως τώρα μικρότερη τιμή, να έχουν μεγαλύτερη συνεισφορά στον υπολογισμό του κεντροειδούς. Η λογική στην παραπάνω ιδέα είναι ότι όσο αυξάνεται ο αριθμός των επαναλήψεων, τα σημεία του συνόλου S^k είναι καλύτερες προσεγγίσεις σε σημεία τοπικών ελαχίστων. Η χρήση λοιπόν των βαρών στον υπολογισμό του κεντροειδούς βοηθά περισσότερο στο να εστιάσουμε τον έλεγχο σε περιοχές, που είναι στην γειτονιά του ολικού ελαχίστου.

3.5.2 Υπολογισμός του νέου σημείου με χρήση βαρών (weighted reflection)

Το νέο σημείο υπολογίζεται σύμφωνα με την παρακάτω διαδικασία :

- Αρχικά υπολογίζουμε την τιμή της συνάρτησης στο κεντροειδές σύμφωνα με την παρακάτω σχέση

$$f_w^k = \sum_{j=1}^n w_j^k f(x_{ij}^k) \quad (3.5.2.1)$$

Το w_j^k δίνεται από τον προηγούμενο τύπο (3.5.1.2)

- Ο υπολογισμός λοιπόν του νέου σημείου δίνεται από τις παρακάτω σχέσεις

$$x^k = c_w^k - a^k (x_{i0}^k - c_w^k) \quad \text{αν } f_w^k < f(x_{i0}^k) \quad (3.5.2.2)$$

$$x^k = x_{i0}^k - a^k (c_w^k - x_{i0}^k) \quad \text{αν } f_w^k \geq f(x_{i0}^k) \quad (3.5.2.3)$$

όπου

$$a^k = 1 - ((f(x_{i0}^k) - f_w^k) / (f_{\max}^k - f_{\min}^k + y^k)) \quad \text{αν } f_w^k < f(x_{i0}^k) \quad (3.5.2.4)$$

$$a^k = 1 - ((f_w^k - f(x_{i0}^k)) / (f_{\max}^k - f_{\min}^k + y^k)) \quad \text{αν } f_w^k \geq f(x_{i0}^k) \quad (3.5.2.5)$$

Για τον ορισμό του y^k , ισχύει ότι και στην περίπτωση της φ^k .

Στον υπολογισμό του νέου σημείου λαμβάνουμε υπόψη τις τιμές της συνάρτησης που έχουμε ήδη υπολογίσει. Επίσης υπολογίζουμε την καταλληλότητα του κεντροειδούς σημείου. Για τον υπολογισμό της συνάρτησης σε αυτό το σημείο χρησιμοποιούμε τον τύπο (3.5.2.1), ώστε να αποφύγουμε μια επιπρόσθετη αποτίμηση της συνάρτησης. Η τιμή που υπολογίζεται με αυτόν τον τύπο είναι ένας σταθμισμένος μέσος όρος, όπου τα βάρη υπολογίζονται ακριβώς με τον ίδιο τρόπο, όπως και στην περίπτωση του κεντροειδούς

(σχέση (3.5.1.2)). Συγκρίνοντας την τιμή της συνάρτησης στο κεντροειδές με την τιμή $f(x_{i0}^k)$ προσδιορίζουμε την κατεύθυνση που θα πάρουμε το νέο σημείο. Αν $f_w^k \leq f(x_{i0}^k)$ το νέο σημείο είναι πιο κοντά στην γειτονιά του κεντροειδούς, ενώ στην αντίθετη περίπτωση βρίσκεται πιο κοντά στο σημείο x_{i0}^k . Οι τιμές του a^k είναι περίπου μονάδα στις πρώτες επαναλήψεις, ενώ όσο αυξάνεται ο αριθμός των επαναλήψεων η τιμή αυτή γίνεται μικρότερη. Αυτό σημαίνει ότι αρχικά επιτρέπουμε μεγάλα βήματα, ώστε να καλύψουμε όλο και μεγαλύτερο εύρος στο διάστημα αναζήτησης. Όσο η αναζήτηση προχωράει, το βήμα περιορίζεται, θεωρώντας ότι έχουμε εγκλωβιστεί στην γειτονιά του ολικού ελαχίστου.

3.5.3 Περιγραφή του αλγορίθμου

- **Βήμα 0**

Αρχικοποίηση του αριθμού επαναλήψεων ($k = 1$).

Δημιουργία αρχικού συνόλου σημείων $S^k = \{x_1^k, \dots, x_m^k\}$ (τα σημεία x_i^k , $i = 1, \dots, m$ επιλέγονται τυχαία στο πεδίο ορισμού D).

Υπολογισμός της τιμής της συνάρτησης στα σημεία του συνόλου S^k .

- **Βήμα 1**

Υπολογισμός των τιμών x_{\max}^k , x_{\min}^k και f_{\max}^k , f_{\min}^k (όπου f_{\max}^k και f_{\min}^k η μέγιστη και η ελάχιστη τιμή της συνάρτησης για τα σημεία του συνόλου S^k).

Εξέταση κριτηρίου τερματισμού

Αν ικανοποιείται

τότε, τέλος αλγορίθμου

αλλιώς, συνέχεια στο βήμα 2

Τέλος Αν

- **Βήμα 2**

Επιλογή $n+1$ τυχαίων σημείων (όπου n η διάσταση της συνάρτησης) από το παραπάνω δείγμα m σημείων S^k . Τα σημεία αυτά συμβολίζονται $x_{i0}, x_{i1}, \dots, x_{in}$ για το σύνολο S^k , (όπου το i ισούται με τον αύξοντα αριθμό επανάληψης).

Υπολογισμός του κεντροειδούς με βάση τον τύπο (3.5.1.1).

Επιλογή ενός νέου σημείου x^k σύμφωνα με τους τύπους (3.5.2.4) και (3.5.2.5)

Αν το νέο σημείο είναι εκτός του πεδίου ορισμού της συνάρτησης,

τότε επανάληψη βήματος 2,

αλλιώς συνέχεια στο βήμα 3

Τέλος Αν

- **Βήμα 3**

Αν $f(x^k) \geq f(x_{\max}^k)$ τότε

Αν ο ρυθμός επιτυχίας > 0.5 , τότε

$S^{k+1} = S^k$

Αύξηση μετρητή επαναλήψεων ($k = k+1$)

Επιστροφή στο βήμα 2

αλλιώς



Επιλογή ενός νέου σημείου σύμφωνα με τον τύπο
 $x^k = (c^k + x_{io}^k) / 2$

Αν το νέο σημείο είναι εκτός του πεδίου ορισμού της συνάρτησης,
τότε επανάληψη του βήματος 2,

αλλιώς,

Αν $f(x^k) < f(x_{max}^k)$ τότε

$$S^{k+1} = S^k \cup \{x^k\} - \{x_{max}^k\}$$

Αύξηση μετρητή επαναλήψεων ($k=k+1$).

Επιστροφή στο βήμα 1.

αλλιώς, $S^{k+1} = S^k$

Αύξηση του μετρητή επανάληψης ($k=k+1$)

Επιστροφή στο βήμα 2

Τέλος Αν

Τέλος Αν

Τέλος Αν (Αν ο ρυθμός επιτυχίας > 0.5)

Τέλος Αν

• Βήμα 4

Αν $f(x^k) < f(x_{max}^k)$ τότε

$$S^{k+1} = S^k \cup \{x^k\} - \{x_{max}^k\}.$$

Αύξηση μετρητή επαναλήψεων ($k = k+1$).

Επιστροφή στο βήμα 1.

Τέλος Αν

Για τους ορισμούς του κριτηρίου τερματισμού και του ρυθμού επιτυχίας, ισχύει ότι και στην περιγραφή του απλού αλγορίθμου του Price.

3.6 Πειραματικά αποτελέσματα

Οι δύο αλγόριθμοι τυχαίας αναζήτησης (απλός αλγόριθμος του Price και βελτιωμένη έκδοση) εφαρμόστηκαν για διάφορες συναρτήσεις ελέγχου (δες παράρτημα) κι έδωσαν τα παρακάτω αποτελέσματα. Τα κριτήρια τερματισμού που χρησιμοποιήθηκαν είναι τα ακόλουθα :

- Τερματισμός όταν ο αλγόριθμος καταφέρνει να φτάσει στην επιθυμητή τιμή για το ολικό ελάχιστο (target value).
- Ικανοποίηση του απλού κριτηρίου τερματισμού του αλγορίθμου, δηλαδή τερματισμός όταν $f_{max} - f_{min} < 10^{-6}$
- Μέγιστο πλήθος αποτιμήσεων της συνάρτησης (function evaluations) = 150000

Απλός αλγόριθμος του Price

	Διάσταση	Ελαχ. τιμή	Ολ. ελάχιστο	Αποτιμήσεις
Sixhump	2	-1.0316284	-1.0316285	2203
Goldstein	2	3.0000026	3	2856
Rastrigin	2	-1.9999	-2	3521
Griewank2	2	0.00000008	0	3224
Neural	15	0.00043	0	150000
Griewank10	10	0.000123	0	86123

Στον πίνακα αυτό παρατίθενται κατά σειρά στήλης από τα αριστερά προς τα δεξιά οι ακόλουθες πληροφορίες.

- Το όνομα της συνάρτησης
- Το πλήθος παραμέτρων της
- Η ελάχιστη επιστρεφόμενη τιμή του αλγορίθμου
- Η πραγματική τιμή του ολικού ελαχίστου για κάθε συνάρτηση ελέγχου (*test function*)
- Το πλήθος των αποτιμήσεων(κλήσεων) της συνάρτησης

Βελτιωμένη έκδοση του αλγορίθμου του Price

	Διάσταση	Ελαχ. τιμή	Ολ. ελάχιστο	Αποτιμήσεις
Sixhump	2	-1.0316284	-1.0316285	1878
Goldstein	2	3	3	2345
Rastrigin	2	-1.9999	-2	2209
Griewank2	2	0.000000011	0	1932
Neural	15	0.000061345	0	150000
Griewank10	10	0.0000083	0	81778

Σχόλια- Συμπεράσματα

Από τα πειραματικά αποτελέσματα, διαφαίνεται η αναμφισβήτητη ανωτερότητα της βελτιωμένης έκδοσης του αλγορίθμου του Price, σε σχέση με τον απλό αλγόριθμο. Το αποτέλεσμα είναι βέβαιο αναμενόμενο, αφού ο βελτιωμένος αλγόριθμος αξιοποιεί καλύτερα την γνώση για τις τιμές της συνάρτησης σε προηγούμενα σημεία. Κρίνεται αποτελεσματικός και στην περίπτωση συναρτήσεων με πολλές παραμέτρους (Griewank10), αφού καταφέρνει να προσεγγίσει ικανοποιητικά την τιμή του πραγματικού ολικού ελαχίστου. Η συνάρτηση Griewank10 είναι μάλιστα ένα αρκετά δύσκολο πρόβλημα, αφού έχει χιλιάδες τοπικά ελάχιστα, γύρω από την περιοχή του πραγματικού ολικού ελαχίστου. Η επιλογή ενός πιο αυστηρού κριτηρίου τερματισμού θα είχε ως αποτέλεσμα μεγαλύτερη ακρίβεια της επιστρεφόμενης λύσης. Πάντως θα πρέπει να σημειώσουμε εδώ, ότι ο αλγόριθμος αυτός σε συναρτήσεις με πολλές παραμέτρους επιστρέφει ως λύση, σημεία στην γειτονιά του πραγματικού ολικού ελαχίστου. Για την επίτευξη λοιπόν μεγαλύτερης

υπολογιστικής ακρίβειας συνίσταται η εφαρμογή κάποιας τοπικής μεθόδου, ξεκινώντας από το επιστρεφόμενο σημείο του αλγορίθμου.

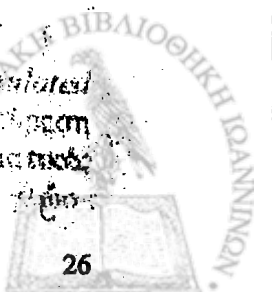
4.1 Επιστροφή

Η μέθοδος που αναφέρεται απλοϊκώς [11], [12] είναι μια απλή μέθοδος που ξεκινάει με κάποιο σημείο εκκίνησης και ακολουθεί την ακολουθία των σημείων που προκύπτει από την επανάληψη της διαδικασίας που περιγράφεται παραπάνω. Η διαδικασία αυτή ονομάζεται *Απλοστολισμός*, και είναι η πιο απλή και εύκολη και στο ίδιο το κείμενο αναφέρεται ως *συνήθης μέθοδος*. Εφαρμόζοντας αυτή τη μέθοδο, ο υπολογιστής βρίσκει το σημείο που πλησιάζει περισσότερο στο σημείο που αναζητούμε. Η διαδικασία αυτή είναι πολύ απλή και εύκολη, αλλά απαιτείται να έχουμε κάποιο σημείο εκκίνησης. Ο υπολογιστής με τη βοήθεια αυτής της διαδικασίας, βρίσκει το σημείο που πλησιάζει περισσότερο στο σημείο που αναζητούμε. Για να γίνει διαφορετική η διαδικασία αυτή, απαιτείται να έχουμε κάποιο σημείο εκκίνησης. Η διαδικασία αυτή ονομάζεται *Απλοστολισμός*, και είναι η πιο απλή και εύκολη και στο ίδιο το κείμενο αναφέρεται ως *συνήθης μέθοδος*.

Η μέθοδος αυτή συνδέεται με την κατοχή της πληροφορίας της εφαρμογής που μπορεί να πάρει μόνο με ορισμένες πηγές. Η θερμότητα που παράγει την αθροιστική με την ταχεία διακύμανση των σημείων, όπου η πηγή της πληροφορίας συνδέεται με κάποια σχέση με την πληροφορία που παράγει. Αρχικά σε αυτήν η πληροφορία η πληροφορία αυτή είναι πολύ μεγάλη με αποτέλεσμα η αναζήτηση να είναι σε όσο το δυνατόν μεγαλύτερη ποσότητα των σημείων. Η απλοστολιστική μέθοδος είναι η πιο απλή και εύκολη και στο ίδιο το κείμενο αναφέρεται ως *συνήθης μέθοδος*. Η διαδικασία αυτή ονομάζεται *Απλοστολισμός*, και είναι η πιο απλή και εύκολη και στο ίδιο το κείμενο αναφέρεται ως *συνήθης μέθοδος*.

Η απλοστολιστική μέθοδος είναι η πιο απλή και εύκολη και στο ίδιο το κείμενο αναφέρεται ως *συνήθης μέθοδος*. Η διαδικασία αυτή ονομάζεται *Απλοστολισμός*, και είναι η πιο απλή και εύκολη και στο ίδιο το κείμενο αναφέρεται ως *συνήθης μέθοδος*.

Στο ίδιο το κείμενο αναφέρεται ως *συνήθης μέθοδος*. Η διαδικασία αυτή ονομάζεται *Απλοστολισμός*, και είναι η πιο απλή και εύκολη και στο ίδιο το κείμενο αναφέρεται ως *συνήθης μέθοδος*.



Κεφάλαιο 4 - Simulated annealing

4.1 Εισαγωγή

Η μέθοδος του *simulated annealing* [11], [13] είναι μια από τις πλέον διαδεδομένες μεθόδους ολικής ελαχιστοποίησης. Είναι εμπνευσμένη από την διαδικασία του *annealing* στην φυσική για την εύρεση καταστάσεων χαμηλής ενέργειας. Αρχικά χρησιμοποιήθηκε για την επίλυση προβλημάτων *διακριτής βελτιστοποίησης* και στη συνέχεια επεκτάθηκε και στην βελτιστοποίηση συνεχών συναρτήσεων. Εφαρμόζεται με επιτυχία και στην περίπτωση συναρτήσεων με πολλές παραμέτρους κι έχει ως πλεονέκτημα ότι δεν είναι απαραίτητο η συνάρτηση στην οποία εφαρμόζεται να είναι παραγωγίσιμη ή απλά συνεχής. Συγκριτικά με άλλους αλγορίθμους ελαχιστοποίησης, υλοποιείται μάλλον εύκολα [12]. Το μεγαλύτερο όμως πλεονέκτημα αυτής της μεθόδου είναι ότι *εγγυάται στατιστικά* την εύρεση του ολικού ελαχίστου. Υπάρχουν διάφορες μαθηματικές αποδείξεις, που αναφέρουν ότι αν χρησιμοποιηθούν συγκεκριμένοι τύποι για μείωση της θερμοκρασίας και συγκεκριμένες κατανομές (για την παραγωγή διαφορετικών καταστάσεων), η μέθοδος θα συγκλίνει στο πραγματικό ολικό ελάχιστο.

Η μέθοδος αυτή συνδέεται με την ύπαρξη της *παραμέτρου της θερμοκρασίας* που μπορεί να παίρνει μόνο μη αρνητικές τιμές. Η θερμοκρασία καθορίζει την πιθανότητα με την οποία δεχόμαστε και σημεία, όπου η τιμή της αντικειμενικής συνάρτησης αυξάνεται, σε σχέση με προηγούμενες τιμές. Αρχικά σε υψηλές θερμοκρασίες η πιθανότητα αυτή είναι πολύ μεγάλη, με αποτέλεσμα η αναζήτηση να γίνεται σε όσο το δυνατόν μεγαλύτερες περιοχές των παραμέτρων. Η αποδοχή καταστάσεων με μεγαλύτερες τιμές συνάρτησης από προηγούμενες, προσδίδει στη μέθοδο χαρακτήρα μεθόδου ολικής ελαχιστοποίησης, αφού με αυτόν τον τρόπο επιτρέπει τον απεγκλωβισμό από περιοχές πιθανών τοπικών ελαχίστων. Σε πολύ χαμηλές θερμοκρασίες, η πιθανότητα αποδοχής γίνεται πολύ μικρή κι έτσι φτάνουμε σε μια κατάσταση ισορροπίας κατά την οποία η μετάβαση από μια κατάσταση σε κάποια άλλη επιφέρει σχεδόν ασήμαντες μεταβολές. Ο παράγοντας της θερμοκρασίας καθορίζει σε μεγάλο βαθμό και το μήκος του βήματος, με το οποίο γίνεται η αναζήτηση στο πεδίο ορισμού των παραμέτρων.

Η επιτυχία αυτής της μεθόδου και η σύγκλιση της στο πραγματικό ολικό ελάχιστο, εξαρτάται από τον ρυθμό μείωσης της θερμοκρασίας. Γρήγορη μείωση της θερμοκρασίας, έχει ως αποτέλεσμα την άμεση σύγκλιση της μεθόδου σε σημείο που πολύ πιθανόν να μην είναι το πραγματικό ολικό ελάχιστο. Επίσης πολύ αργός ρυθμός για την μείωση της θερμοκρασίας μπορεί να επιφέρει καλύτερα αποτελέσματα, αλλά έχοντας ως αντίτιμο τον τεράστιο υπολογιστικό(απαγορευτικό σε κάποιες περιπτώσεις) χρόνο για την σύγκλιση της μεθόδου. Το θέμα λοιπόν της επιλογής κατάλληλου σχεδίου για την μείωση της θερμοκρασίας (*annealing schedule*) είναι πολύ κρίσιμο για την επιτυχία της μεθόδου. Έχουν προταθεί διάφοροι τρόποι μείωσης της θερμοκρασίας, οι οποίοι εγγυώνται στατιστικά την εύρεση του πραγματικού ολικού ελαχίστου. Πολλές φορές όμως η αυστηρή εφαρμογή αυτών των αλγορίθμων δεν είναι εφικτή κι έχουν γίνει προτάσεις για διαφορετικούς τρόπους μείωσης της θερμοκρασίας. Ο *Ingber* ονόμασε αυτές τις μεθόδους με τον όρο *Simulated Quenching* [15].

Στο κεφάλαιο αυτό επιχειρείται αρχικά μια περιγραφή του αλγορίθμου του *Simulated annealing* με αναφορά σε διάφορες εκδόσεις του αλγορίθμου, που εγγυώνται την εύρεση του πραγματικού ολικού ελαχίστου. Επίσης παρατίθενται και οι αντίστοιχες μαθηματικές αποδείξεις. Έπειτα, παρουσιάζεται αναλυτικά ο αλγόριθμος που υλοποιήθηκε στη γλώσ-

σα MCL του περιβάλλοντος MERLIN και αποτελέσματα από την εφαρμογή του. Τέλος προτείνεται ένας νέος αλγόριθμος, που εντάσσεται στην κατηγορία των υβριδικών μεθόδων, αφού συνδυάζει μεθόδους τοπικής ελαχιστοποίησης με τον αλγόριθμο του simulated annealing. Ο νέος αυτός αλγόριθμος επιτυγχάνει όπως θα δούμε από κάποια πειραματικά αποτελέσματα, μεγαλύτερη ακρίβεια λύσης και πιο γρήγορη σύγκλιση σε κάποιες περιπτώσεις.

4.2 Θεωρητική μελέτη

4.2.1 Παρουσίαση του αλγορίθμου [19]

Βήμα 1

- Επιλογή ενός αρχικού σημείου $x^0 \in \Xi$, όπου Ξ το πεδίο ορισμού των παραμέτρων
- Προσδιορισμός συνάρτησης πυκνότητας πιθανότητας για την επιλογή μετάβασης από μια κατάσταση σε κάποια άλλη (generation probability density function). Η συνάρτηση αυτή εξαρτάται από την θερμοκρασία T
- Προσδιορισμός μιας συνάρτησης που ενημερώνει την θερμοκρασία
- Υπολογισμός της τιμής της συνάρτησης στο συγκεκριμένο σημείο $f(x_0)$
Ανάθεση των τιμών $X^k = x^0$ και $k=0$

Βήμα 2

- Προσδιορισμός τυχαίου διανύσματος Z^k με βάση την συνάρτηση πυκνότητας πιθανότητας (generation probability density function), που ορίστηκε στο προηγούμενο βήμα
- Παραγωγή ενός σημείου Y^k (γειτονικό στο X^k)
 $Y^k = X^k + Z^k$
- Αν το Y^k δεν ανήκει στο πεδίο ορισμού Ξ ,
Τότε, επανάληψη του βήματος 2
Αλλιώς, υπολογισμός της τιμής της συνάρτησης $f(Y^k)$ στο νέο αυτό σημείο

Βήμα 3

- Έλεγχος του Metropolis κριτηρίου, για την αποδοχή ή όχι του νέου σημείου Y^k
- Επιλογή ενός τυχαίου αριθμού n μεταξύ 0 και 1 από μια ομοιόμορφη κατανομή. Υπολογισμός της πιθανότητας $P(Y^k | X^k, T_k)$ αποδοχής του νέου σημείου X^{k+1} δοθέντων των X^k και T_k

$$P(Y^k | X^k, T_k) = \min \{ 1, \exp \{ [f(X^k) - f(Y^k)] / T_k \} \}$$

- Αν το $n \leq P(Y^k | X^k, T_k)$
Τότε, $X^{k+1} = Y^k$ και $f(X^{k+1}) = f(Y^k)$
Αλλιώς, $X^{k+1} = X^k$ και $f(X^{k+1}) = f(X^k)$

Βήμα 4

- Αν ικανοποιείται το κριτήριο τερματισμού
Τότε, τερματισμός αλγορίθμου



Αλλιώς, ενημέρωση της θερμοκρασίας σύμφωνα με τη συνάρτηση ενημέρωσης της θερμοκρασίας. Επιστροφή στο βήμα 2

4.2.2 Παρουσίαση διαφόρων αλγορίθμων μείωσης της θερμοκρασίας (annealing schedules) [15]

Το κυριότερο μειονέκτημα του SA είναι ο αργός ρυθμός σύγκλισης. Έχει αποδειχτεί από τους Geman και Geman, ότι για την σύγκλιση του αλγορίθμου, απαιτείται ο ρυθμός μείωσης της θερμοκρασίας να είναι αντιστρόφως ανάλογος με μια λογαριθμική συνάρτηση του χρόνου

$$T(k) = T_0 / \ln(1+k) \quad (4.2.2.1)$$

όπου T_0 είναι η αρχική θερμοκρασία

Είναι φανερό από την εξίσωση (4.2.2.1) ότι χρειάζεται αρκετός χρόνος για την θερμοκρασία $T(k)$ να πλησιάσει το μηδέν. Πολλές έρευνες έχουν γίνει με στόχο την αύξηση του ρυθμού σύγκλισης του αλγορίθμου του SA, προτείνοντας διαφορετικά σχήματα μείωσης της θερμοκρασίας, διαφορετικά από την παραπάνω λογαριθμική προσέγγιση. Οι Szu και Hartley έχουν προτείνει τον αλγόριθμο του **Fast Simulated Annealing** [16] κατά τον οποίον ο ρυθμός μείωσης της θερμοκρασίας είναι αντιστρόφως ανάλογος, γραμμικά με τον χρόνο. Έτσι λοιπόν για $k=1,2,\dots$, έχουμε

$$T(k) = T_0 / (1+k) \quad (4.2.2.2)$$

Αυτός ο ρυθμός μείωσης είναι εκθετικά πιο γρήγορος από τον προηγούμενο ρυθμό μείωσης της εξίσωσης (4.2.2.1).

Στη συνέχεια ο Ingber βελτιώνει το πιο πάνω αποτέλεσμα και προτείνει το **Very Fast Simulated Annealing** [17], κατά το οποίο ο ρυθμός μείωσης της θερμοκρασίας είναι αντιστρόφως ανάλογος με μια εκθετική συνάρτηση του χρόνου. Έτσι λοιπόν για $k=1,2,\dots$, έχουμε :

$$T(k) = T_0 / e^k \quad (4.2.2.3)$$

4.2.3 Κλασικό Simulated Annealing – Θεωρητική μελέτη [14]

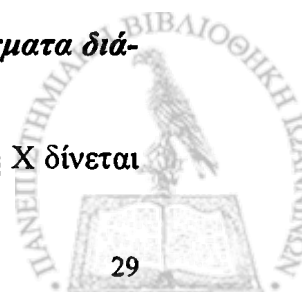
Έστω $g_k(\Delta X)$, όπου $\Delta X = Y - X$, η συνάρτηση πυκνότητας πιθανότητας της $G_{XY}(T_k)$. Δηλαδή η πιθανότητα να φτάσουμε από την κατάσταση X στην κατάσταση Y σε θερμοκρασία T_k

Η συνάρτηση αυτή θα έχει την παρακάτω μορφή

$$g_k(\Delta X) = (2\pi T_k)^{-D/2} \exp(-\|\Delta X\|^2 / (2T_k^2)) \quad (4.2.3.1)$$

όπου D είναι η διάσταση του χώρου καταστάσεων και $X, Y, \Delta X$ είναι διανύσματα διάστασης D

Η πιθανότητα αποδοχής της μετάβασης στην κατάσταση Y από την κατάσταση X δίνεται από την παρακάτω σχέση:



$$A_{XY}(T_k) = \min \left\{ 1, \exp \left(- \frac{c(Y) - c(X)}{T_k} \right) \right\} \quad (4.2.3.2)$$

όπου c η συνάρτηση κόστους για την δεδομένη κατάσταση

Έχει αποδειχτεί ότι ο αλγόριθμος του SA που ικανοποιεί τις εξισώσεις (4.2.3.1) και (4.2.3.2) συγκλίνει σε καταστάσεις που είναι ολικά βέλτιστες.

Απόδειξη

Αρκεί να δείξουμε ότι κάθε κατάσταση στον χώρο καταστάσεων, μπορεί να προκύψει σε πεπερασμένο χρόνο. Πρέπει λοιπόν να δείξουμε ότι η πιθανότητα να μην μεταβούμε σε μια συγκεκριμένη κατάσταση μετά την πάροδο πεπερασμένου χρόνου είναι μηδέν, δηλαδή

$$\prod_{k=k_0}^{\infty} (1 - g_k) = 0 \quad (4.2.3.3) \text{ που είναι ισοδύναμο με}$$

$$\sum_{k=k_0}^{\infty} g_k = \infty \quad (4.2.3.4)$$

Όπου g_k είναι συντόμευση για την συνάρτηση $G_k(\Delta X)$

Η σχέση (4.2.3.4) από τις σχέσεις (4.2.2.1) και (4.2.3.1) είναι ισοδύναμη με:

$$\sum_{k=k_0}^{\infty} g_k = \sum_{k=k_0}^{\infty} \left(\frac{2\pi T_0}{\ln(1+k)} \right)^{-D/2} \exp \left(- \frac{(\Delta X \ln(1+k))^2}{2T_0^2} \right)$$

$$\geq \sum_{k=k_0}^{\infty} \exp(-\ln(1+k)) = \sum_{k=k_0}^{\infty} \frac{1}{1+k} = \infty$$

όπου k_0 είναι μια αρκετά μεγάλη σταθερά. Έτσι λοιπόν ο κλασικός αλγόριθμος του **Simulated Annealing** μπορεί να επιφέρει οποιαδήποτε κατάσταση του χώρου καταστάσεων και κατά συνέπεια να συγκλίνει και στο ολικό ελάχιστο.

4.2.4 Θεωρητική μελέτη του Fast Simulated Annealing [17]

Η συνάρτηση πυκνότητας πιθανότητας της πιθανότητας $G_{XY}(T_k)$ δίνεται από τον ακόλουθο τύπο :

$$g_k(\Delta X) = \frac{T_k}{((\Delta X)^2 + T_k^2)^{(D+1)/2}} \quad (4.2.4.1)$$

Με παρόμοιο τρόπο όπως και προηγουμένως αποδεικνύεται από τις σχέσεις (4.2.2.2) και (4.2.4.1) ότι

$$\sum_{k=k_0}^{\infty} g_k = \sum_{k=k_0}^{\infty} \frac{T_0 / (1+k)}{(\|\Delta X\|^2 + T_0^2 / (1+k)^2)^{(D+1)/2}}$$

$$\geq \frac{T_0}{(2(\Delta X))^{D+1}} \sum_{k=k_0}^{\infty} \frac{1}{1+k} = \infty$$

Ο αλγόριθμος αυτός χρησιμοποιεί ημι-τοπικές αναζητήσεις δηλαδή επιτρέπει περιστασιακά κάποια μεγάλα βήματα, ώστε να απομακρυνθούμε αρκετά από τα διάφορα τοπικά ελάχιστα.

4.2.5 Θεωρητική μελέτη του Very Fast Simulated Annealing [17]

Ο Ingber ακολουθεί την ίδια προσέγγιση με αυτή των **Szu και Hartley** στην υλοποίηση του αλγορίθμου **Very Fast Simulated Annealing**. Η συνάρτηση πυκνότητας πιθανότητας παραγωγής των διαφόρων καταστάσεων (**generation probability density function**) είναι το γινόμενο μονοδιάστατων συναρτήσεων. Κάθε τέτοια συνάρτηση αναφέρεται μόνο σε μια παράμετρο. Επιπρόσθετα, προτείνει η θερμοκρασία να έχει διαφορετικό ρυθμό μείωσης για τον μονοδιάστατο χώρο κάθε παραμέτρου.

Έστω, λοιπόν $\mathbf{x}(k) = (x_1(k), x_2(k), \dots, x_D(k))$ η κατάσταση στον χρόνο k . Το $x_i(k)$ με $1 \leq i \leq D$ (D το πλήθος των παραμέτρων) ανήκει στο διάστημα $[A_i, B_i]$.

Το επόμενο σημείο $x_i(k+1)$ λαμβάνεται σύμφωνα με τον τύπο

$$x_i(k+1) = x_i(k) + z_i(B_i - A_i)$$

όπου z_i τυχαίος αριθμός στο διάστημα $[-1, 1]$

Η συνάρτηση πυκνότητας πιθανότητας για τα z_i είναι

$$g_k(z_i) = \frac{1}{2(|z_i| + T_i(k)) \ln(1 + 1/T_i(k))} \quad (4.2.5.1)$$

$T_i(k)$: είναι η θερμοκρασία για την παράμετρο i την χρονική στιγμή k .

Η συνάρτηση πυκνότητας πιθανότητας δημιουργίας νέων καταστάσεων (**generation density function**) για το VFSA ορίζεται ως εξής:

$$g_k(\mathbf{Z}) = \prod_{i=1}^D \frac{1}{2(|z_i| + T_i(k)) \ln(1 + 1/T_i(k))} \quad (4.2.5.2)$$

Ο αλγόριθμος του Very Fast Simulated Annealing χρησιμοποιεί τον παρακάτω ρυθμό για την μείωση της θερμοκρασίας

$$T_i(k) = T_{oi} \exp(-b_i k^{1/D}) \quad (4.2.5.3)$$

όπου b_i είναι ένας θετικός σταθερός αριθμός.



Θα πρέπει λοιπόν να αποδείξουμε ότι ισχύει η σχέση

$$\sum_{k=k_0}^{\infty} g_k = \infty \quad \text{όπως και προηγουμένως}$$

Σύμφωνα λοιπόν με τις εξισώσεις (4.2.5.2) και (4.2.5.3) έχουμε

$$\begin{aligned} \sum_{k=k_0}^{\infty} g_k &= \sum_{k=k_0}^{\infty} \prod_{i=1}^D g_k(z_i) = \sum_{k=k_0}^{\infty} \prod_{i=1}^D \frac{1}{2(|z_i| + T_{oi} \exp(-b_i k^{1/D})) \ln(1 + \exp(b_i k^{1/D}) / T_{oi})} \\ &\approx \prod_{i=1}^D \frac{1}{2|z_i| b_{ik}} \sum_{k=k_0}^{\infty} \frac{1}{k} = \infty \end{aligned}$$

όπου k_0 είναι μια αρκετά μεγάλη σταθερά

4.3 Περιγραφή του υλοποιημένου αλγορίθμου

Ο αλγόριθμος που υλοποιήθηκε είναι σύμφωνος με το άρθρο του Goffe [16] και τον προτεινόμενο κώδικά του σε Fortran. Ο αλγόριθμος αυτός σύμφωνα με τον Ingber ονομάζεται **Simulated Quenching**, αφού ο ρυθμός μείωσης της θερμοκρασίας που χρησιμοποιείται είναι αρκετά γρήγορος, και δεν γίνεται **εργοδική αναζήτηση**. Σύμφωνα με τον Ingber [18] η αναζήτηση θεωρείται εργοδική, μόνο όταν διέρχεται από όλες τις δυνατές καταστάσεις ενέργειας του συστήματος. Ωστόσο, η ευκολία στην υλοποίηση αυτού του αλγορίθμου και ο συνδυασμός της μεγάλης ταχύτητας σύγκλισης με την ακρίβεια του αποτελέσματος, μάς επιτρέπει να ξεφύγουμε από τον αργό αλλά ισχυρό θεωρητικά κλασικό αλγόριθμο του Boltzman Simulated Annealing. Ακολουθεί λοιπόν η περιγραφή του υλοποιημένου αλγορίθμου.

4.3.1 Περιγραφή αλγορίθμου

Δεδομένα εισόδου

- Αρχική θερμοκρασία T , αρχικό σημείο $x = \{x_i, i = 1, 2, \dots, N\}$ (όπου N η διάσταση της συνάρτησης)
- Αρχικό βήμα μετατόπισης σε κάθε διάστημα παραμέτρου $h = \{h_i, i = 1, 2, \dots, N\}$
- Όρια του πεδίου ορισμού για κάθε παράμετρο a_i, b_i , τέτοια ώστε το $x_i \in [a_i, b_i]$
- Παράγοντας μείωσης της θερμοκρασίας α
- Επαναλήψεις N_s - για N_s επαναλήψεις λαμβάνονται σημεία με σταθερό βήμα μετατόπισης
- Επαναλήψεις N_T - μετά από $N_T * N_s$ επαναλήψεις προσαρμόζεται το μήκος του διανύσματος μετατόπισης
- Επαναλήψεις N_A - μετά από $N_s * N_T * N_A$ επαναλήψεις γίνεται μείωση της θερμοκρασίας

Βήμα 0

- ◆ Υπολόγισε την τιμή της συνάρτησης στο αρχικό σημείο
 $f_c = f(x)$
- ◆ Επανάλαβε N_A φορές (μείωση θερμοκρασίας)
Επανάλαβε N_T φορές (προσαρμογή μήκους διανύσματος μετατόπισης)
Επανάλαβε N_s φορές (βασική κίνηση αλγορίθμου)
Για κάθε μεταβλητή x_i
 $x^t = x$ (ολόκληρο το διάνυσμα)
 $x_i^t = x_i + (2\xi - 1) h_i$, ξ τυχαίος αριθμός στο διάστημα $[0, 1]$
Αν το νέο σημείο είναι εκτός ορίων του επιτρεπτού διαστήματος $[a_i, b_i]$ τότε
 $x_i^t = a_i + (b_i - a_i) * \xi$, όπου ξ ένας τυχαίος αριθμός στο διάστημα $[0, 1]$
 $f_t = f(x^t)$
Αν $f_t \leq f_{best}$ (όπου f_{best} η καλύτερη (ελάχιστη) έως τώρα τιμή της συνάρτησης) τότε
 $x_{best} = x^t$
Αν $f_t < f_c$ τότε
 $x = x^t$ και $f_c = f_t$
Αλλιώς
Υπολόγισε την πιθανότητα $p = e^{-(f_t - f_c) / T}$
Αν το $p > \xi$ (όπου ξ τυχαίος αριθμός στο διάστημα $[0, 1]$) τότε
 $x = x^t$ και $f_c = f_t$
Τέλος Αν ($f_t < f_c$)
Τέλος N_s επαναλήψεων
Τέλος N_T επαναλήψεων
Προσαρμογή διανύσματος μετατόπισης (ο τρόπος με τον οποίο γίνεται, περιγράφεται αναλυτικά πιο κάτω)
Τέλος N_A επαναλήψεων
Μείωση θερμοκρασίας: $T = \alpha T$
 $x = x_{best}$, $f_c = f_{best}$
Αν ικανοποιείται το κριτήριο τερματισμού, τότε τέλος αλγορίθμου
Αλλιώς πήγαινε στο βήμα 0

Παρατηρήσεις



- Η προσαρμογή του βήματος μετατόπισης γίνεται με τέτοιο τρόπο, ώστε να επιτύχουμε ρυθμό αποδοχής κινήσεων γύρω στο 50 %. Η προσαρμογή του βήματος γίνεται λοιπόν με τον ακόλουθο αλγόριθμο :

Για κάθε παράμετρο x_i

Υπολόγισε τον λόγο $r_i = M_i / N_s$, όπου M_i το πλήθος των αποδεκτών κινήσεων στην κατεύθυνση της παραμέτρου x_i κατά τις τελευταίες N_s προσπάθειες.

$$\text{Αν } r_i > 0.6 \quad h_i = h_i \left(1 + 2 \frac{r_i - 3/5}{2/5} \right)$$

$$\text{Αν } r_i < 0.4 \quad h_i = \frac{h_i}{1 + 2 \frac{2/5 - r_i}{2/5}}$$

$$\text{Αν } h_i > b_i - a_i \quad h_i = b_i - a_i$$

- **Κριτήριο τερματισμού:** αν η διαφορά των τιμών της συνάρτησης για τα σημεία που επέστρεψε ο αλγόριθμος στις τελευταίες (*Last* : προκαθορισμένες από τον χρήστη) θερμοκρασίες είναι περίπου μηδέν και η τιμή για το τελευταίο σημείο ισούται με την τρέχουσα καλύτερη τιμή, τότε ο αλγόριθμος τερματίζει.

4.3.2 Σχόλια - παρατηρήσεις

Ο αλγόριθμος του **Simulated Annealing** που περιγράφηκε παραπάνω προσπαθεί να λύσει το πρόβλημα εύρεσης του ολικού ελαχίστου μιας συνάρτησης διάστασης N . Δέχεται οπωσδήποτε τις νέες τιμές των μεταβλητών, όταν αυτές οδηγούν σε μικρότερη τιμή της συνάρτησης από την τρέχουσα ελάχιστη τιμή, αλλά είναι *ανεκτικός* και στην αποδοχή κινήσεων που οδηγούν σε μεγαλύτερη τιμή της συνάρτησης, όπως θα δούμε παρακάτω.

- Αρχικά από κάποιο τυχαίο αρχικό σημείο επιλέγεται κάποιο γειτονικό του που βρίσκεται σε απόσταση, που καθορίζεται από ένα προσαρμοζόμενο διάνυσμα μετατόπισης (**διάνυσμα h**). Ελέγχεται βέβαια αν το καινούριο σημείο δεν ξεπερνά τα όρια του πεδίου ορισμού. Υπολογίζεται η τιμή της συνάρτησης στο νέο αυτό σημείο και συγκρίνεται με την τιμή της συνάρτησης στο αρχικό σημείο.
- Είναι αποδεκτές όλες εκείνες οι κινήσεις, που οδηγούν σε μικρότερη τιμή για την συνάρτηση, ενώ για κινήσεις που μας επιφέρουν μεγαλύτερη τιμή συνάρτησης ελέγχεται το **Metropolis criterion**. Σύμφωνα με αυτό το κριτήριο έστω x' το αρχικό σημείο κι έστω x'' ένα γειτονικό σημείο. Αν $f(x'') - f(x') < 0$, τότε δεχόμαστε το νέο σημείο. Αλλιώς, το νέο σημείο το δεχόμαστε με πιθανότητα : $\exp(- (f(x'') - f(x')) / T)$.
- Η παραπάνω πιθανότητα αποδοχής του νέου σημείου εξαρτάται από την παράμετρο της θερμοκρασίας και το μέγεθος του βήματος μετατόπισης. Όσο μεγαλύτερη είναι η τιμή της θερμοκρασίας και όσο μικρότερη τιμή έχουμε για το βήμα μετατόπισης, τόσο αυξάνεται και η πιθανότητα να αποδεχτούμε το νέο σημείο. Αν το νέο σημείο γίνει αποδεκτό, τότε ο αλγόριθμος του **simulated annealing** συνεχίζει από αυτό το σημείο, αλλιώς επιλέγεται κάποιο άλλο σημείο.

- Κάθε στοιχείο του διανύσματος μετατόπισης VM προσαρμόζεται περιοδικά με τέτοιο τρόπο, ώστε περίπου μισές από τις τιμές της συνάρτησης προς την κατεύθυνση κάθε παραμέτρου να είναι αποδεκτές.
- Η μείωση της θερμοκρασίας γίνεται σύμφωνα με τον τύπο $T(i+1) = \alpha T$, όπου i είναι η i -οστή επανάληψη και α ο παράγοντας μείωσης της θερμοκρασίας. Όσο η θερμοκρασία μειώνεται, μειώνεται και η πιθανότητα να αποδεχτούμε σημεία με μεγαλύτερη τιμή από την έως τώρα ελάχιστη τιμή της συνάρτησης. Μελετώντας το διάνυσμα μετατόπισης h παρατηρούμε ότι το μήκος του διανύσματος μειώνεται, καθώς μειώνεται και η θερμοκρασία. Η λογική για την παραπάνω μείωση της μετατόπισης είναι ότι με την πάροδο του χρόνου θα πρέπει να επικεντρωθούμε στην περιοχή του ολικού ελαχίστου. Μεγάλα λοιπόν βήματα αυτή την χρονική στιγμή, θα είχαν ως αποτέλεσμα να απομακρυνθούμε από την γειτονιά του ολικού ελαχίστου.
- **Σημασία της παραμέτρου της θερμοκρασίας**

Η επιλογή κάποιας αρχικής τιμής για την θερμοκρασία καθώς επίσης και του παράγοντα μείωσής της, είναι κρίσιμες για την επιτυχία του αλγορίθμου. Για μια πολύ μικρή αρχική θερμοκρασία, το μήκος του διανύσματος h θα είναι επίσης πολύ μικρό. Έτσι θα επικεντρωθούμε σε ένα πολύ μικρό υποσύνολο του πεδίου ορισμού της συνάρτησης, με μεγάλη πιθανότητα αποτυχίας για την σωστή απόφαση του ολικού ελαχίστου. Είναι λοιπόν άμεση η εξάρτηση του μήκους του διανύσματος h και της αρχικής θερμοκρασίας T . Θα είναι καλύτερο να τρέξουμε μια φορά τον αλγόριθμο δοκιμαστικά, ώστε να αποφασίσουμε την κατάλληλη τιμή για την αρχική θερμοκρασία T και τον ρυθμό μείωσής της.

4.3.3 Παρουσίαση ενός νέου αλγορίθμου

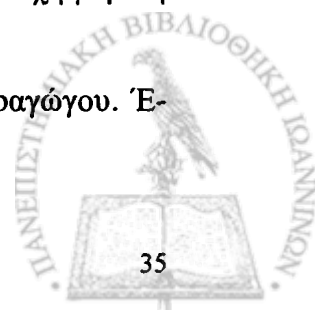
Ο αλγόριθμος που θα παρουσιάσουμε στη συνέχεια, ακολουθεί την ίδια λογική με τον προηγούμενο. Διαφέρει στο ότι το νέο σημείο που προκύπτει είναι ενημερωμένο ως προς όλες τις παραμέτρους. Το βήμα μετατόπισης δίνεται σύμφωνα με τον τύπο :

$$\Delta x_i = h * [a_i - x_i + (b_i - a_i) * \xi_i] * S_i \quad (4.3.3.1)$$

όπου x_i η τιμή της παραμέτρου i για το σημείο x , a_i , b_i το αριστερό και δεξιό άκρο του πεδίου ορισμού για την παράμετρο i και h ο συντελεστής που προσαρμόζεται με τέτοιο τρόπο, ώστε μισά περίπου από τα σημεία να είναι αποδεκτά

Το S_i είναι ένα διάνυσμα που κάθε συνιστώσα του, εκφράζει την ευαισθησία της αντίστοιχης παραμέτρου σε πολύ μικρές μεταβολές της τιμής της. Όσο λοιπόν μεγαλύτερη είναι η τιμή της παραγώγου ως προς μια παράμετρο, τόσο περισσότερο επηρεάζονται οι τιμές της συνάρτησης από πολύ μικρές μεταβολές της. Πρέπει λοιπόν κατά τον υπολογισμό του μήκους του διανύσματος μετατόπισης να λαμβάνεται υπόψη το παραπάνω γεγονός. Έτσι το μήκος αυτού του διανύσματος, θα πρέπει να είναι αισθητά μειωμένο ως προς την κατεύθυνση εκείνης της συνιστώσας, που έχει μεγάλη τιμή η αντίστοιχη μερική παράγωγος. Το διάνυσμα S_i δίνεται λοιπόν από τους παρακάτω τύπους.

Αρχικά υπολογίζουμε την μέγιστη και την ελάχιστη απόλυτη τιμή της παραγώγου. Έχουμε λοιπόν τις παρακάτω σχέσεις :



$$|\partial f|_{\max} = \max \left\{ \left| \frac{\partial f}{\partial x_1} \right|, \dots, \left| \frac{\partial f}{\partial x_n} \right| \right\}$$

$$|\partial f|_{\min} = \min \left\{ \left| \frac{\partial f}{\partial x_1} \right|, \dots, \left| \frac{\partial f}{\partial x_n} \right| \right\}$$

Αν $|\partial f|_{\max} \neq |\partial f|_{\min}$ τότε

$$S_i = \frac{1}{1 + (k-1) \frac{|\partial f| - |\partial f|_{\min}}{|\partial f|_{\max} - |\partial f|_{\min}}} \quad (4.3.3.2)$$

όπου k σταθερά, ίση με την διάσταση της συνάρτησης f

Αν $|\partial f|_{\max} = |\partial f|_{\min}$ τότε

$$S_i = 1 \quad (3) \quad (4.3.3.3)$$

Ο νέος λοιπόν αλγόριθμος διαφέρει από τον προηγούμενο, στον τρόπο με τον οποίο επιλέγονται τα νέα γειτονικά σημεία. Το σχήμα του αλγορίθμου θα είναι το εξής :

Δεδομένα εισόδου

Αρχική θερμοκρασία T , αρχικό σημείο $x = \{x_i \ i=1, 2, \dots, N\}$ (όπου N η διάσταση της συνάρτησης).

Συντελεστής μετατόπισης h .

Όρια του πεδίου ορισμού για κάθε παράμετρο a_i, b_i , τέτοια ώστε το $x_i \in [a_i, b_i]$

Παράγοντας μείωσης της θερμοκρασίας α .

Επαναλήψεις N_s - για τόσες επαναλήψεις λαμβάνονται σημεία με σταθερό βήμα μετατόπισης.

Επαναλήψεις N_T - μετά από $N_T * N_s$ επαναλήψεις προσαρμόζουμε το μήκος του διάνυσματος μετατόπισης.

Επαναλήψεις N_A - μετά από $N_s * N_T * N_A$ επαναλήψεις γίνεται μείωση της θερμοκρασίας.

Βήμα 0

Υπολόγισε την τιμή της συνάρτησης στο αρχικό σημείο

$$f_c = f(x)$$

Επανάλαβε N_A φορές (μείωση θερμοκρασίας)

Επανάλαβε N_T φορές (προσαρμογή συντελεστή μετατόπισης h)

Επανάλαβε N_s φορές (βασική κίνηση αλγορίθμου)

Για κάθε μεταβλητή x_i

$$x^t = x \text{ (ολόκληρο το διάνυσμα)}$$

$$x_i^t = x_i + \Delta x_i$$

(Το Δx_i δίνεται από τον τύπο (4.3.3.1))

$$f_t = f(x^t)$$

Αν $f_t \leq f_{best}$, όπου f_{best} η καλύτερη (ελάχιστη) έως τώρα τιμή της συνάρτησης τότε $x_{best} = x^t$

Αν $f_t < f_c$ τότε

$$x = x^t \text{ και } f_c = f_t$$

Αλλιώς

Υπολόγισε την πιθανότητα $p = e^{-(f_t - f_c)/T}$

Αν το $p > \xi$ (όπου ξ τυχαίος αριθμός στο διάστημα $[0, 1]$) τότε

$$x = x^t \text{ και } f_c = f_t$$

Τέλος Αν ($f_t < f_c$)

Τέλος N_S επαναλήψεων

Υπολογισμός των παραμέτρων S_i σύμφωνα με τους τύπους (4.3.3.2) και (4.3.3.3)

Τέλος N_T επαναλήψεων

Προσαρμογή συντελεστή μετατόπισης h (ο τρόπος με τον οποίο γίνεται περιγράφεται αναλυτικά πιο κάτω)

Τέλος N_A επαναλήψεων

Μείωση θερμοκρασίας : $T = \alpha T$

$$x = x_{best}, f_c = f_{best}$$

Αν ικανοποιείται το κριτήριο τερματισμού, τότε τέλος αλγορίθμου
αλλιώς πήγαινε στο βήμα 0

Παρατηρήσεις

Προσαρμογή του συντελεστή h .

Υπολόγισε τον λόγο $r = M / N_S$, όπου M το πλήθος των αποδεκτών κινήσεων κατά τις τελευταίες N_S προσπάθειες

$$\text{Αν } r > 0.6 \text{ θέσε } h = h \left(1 + 2 \frac{r - 3/5}{2/5} \right)$$

$$\text{Αν } r < 0.4 \text{ θέσε } h = \frac{h}{1 + 2 \frac{2/5 - r}{2/5}}$$

$$\text{Αν } h > 1 \text{ θέσε } h = 1$$

- Κριτήριο τερματισμού : το ίδιο με τον προηγούμενο αλγόριθμο.

4.4 Συνδυάζοντας τον αλγόριθμο του Simulated annealing με μια μέθοδο τοπικής ελαχιστοποίησης [20]

4.4.1 Σύγκριση Simulated Annealing με μια μέθοδο τοπικής ελαχιστοποίησης

Το σημαντικότερο πλεονέκτημα του simulated annealing είναι η ικανότητα να απεγκλωβίζεται από περιοχές τοπικών ελαχίστων, με το να δέχεται με αυξημένη πιθανότητα (κυρίως στις αρχικές υψηλές θερμοκρασίες) και σημεία, στα οποία η τιμή της συνάρτησης αυξάνεται. Ωστόσο, σε περιπτώσεις συναρτήσεων με πολλές παραμέτρους στερείται ακρίβειας. Αντιθέτως, οι υπάρχουσες μέθοδοι τοπικής ελαχιστοποίησης έχουν την ικανότητα να προσεγγίζουν με μεγάλη ακρίβεια το τοπικό ελάχιστο, που βρίσκεται σε μικρές περιοχές ενδιαφέροντος. Έτσι, ο συνδυασμός των δύο παραπάνω μεθόδων, μπορεί να πετύχει την *χρυσή τομή*, αναιρώντας σε μεγάλο βαθμό τα μειονεκτήματα των δύο αυτών ξεχωριστών μεθόδων.

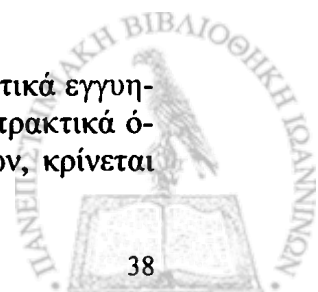
Μπορούμε συνεπώς με την χρήση του SA να εντοπίσουμε τις περιοχές εκείνες που παρουσιάζουν ενδιαφέρον (βρίσκονται κοντά στο ολικό ελάχιστο) και με την χρήση μιας μεθόδου τοπικής ελαχιστοποίησης να φτάσουμε με μεγάλη ακρίβεια στα διάφορα τοπικά ελάχιστα. Καλούμε αυτή την υβριδική μέθοδο **SALT (Simulated Annealing and Local Tuning)**. Μια άλλη υβριδική μέθοδος παρουσιάζεται στην αναφορά [21].

4.4.2 Περιγραφή του αλγορίθμου SALT

- Ξεκινάμε μια τοπική μέθοδο ελαχιστοποίησης από ένα τυχαίο αρχικό σημείο (θεωρούμε ότι ξέρουμε την παράγωγο). Αρχικά επιτρέπουμε στην μέθοδο τοπικής ελαχιστοποίησης να κάνει πολλές επαναλήψεις, ώστε να λάβουμε ένα καλό αρχικό σημείο για τον απλό αλγόριθμο του SA.
- Στη συνέχεια ο αλγόριθμος του απλού simulated annealing εξελίσσεται κανονικά.
- Μόνο σε πολύ χαμηλές θερμοκρασίες ξανακαλούμε μια τοπική μέθοδο, επιτρέποντας όμως λίγες επαναλήψεις.

Η λογική του παραπάνω αλγορίθμου στηρίζεται στο γεγονός ότι εκμεταλλευόμαστε την ακρίβεια των αποτελεσμάτων της τοπικής μεθόδου μόνο στις τελευταίες επαναλήψεις του αλγορίθμου, ώστε να αποφευχθεί σε σημαντικό βαθμό ο κίνδυνος εγκλωβισμού σε κάποιο “κακό” τοπικό ελάχιστο, από τα πρώτα κιόλας βήματα. Θεωρούμε ότι ο αλγόριθμος του simulated annealing έχει καταφέρει μετά από σημαντικό αριθμό επαναλήψεων να προσεγγίσει (τουλάχιστον) την γειτονιά του ολικού ελαχίστου. Επομένως η κλήση μιας τοπικής μεθόδου, βοηθάει να φτάσουμε με μεγάλη ακρίβεια το ολικό ελάχιστο, στην συγκεκριμένη περιοχή ενδιαφέροντος. Η σύγκλιση της μεθόδου γίνεται συνεπώς με πιο ταχύ ρυθμό, αφού η πιθανότητα αποδοχής σημείων με υψηλότερη τιμή συνάρτησης από την τρέχουσα ελάχιστη τιμή (*Metropolis criterion*) γίνεται ακόμη μικρότερη, συγκριτικά με το απλό SA. Η προσαύξηση λοιπόν των κλήσεων της συνάρτησης με την εφαρμογή της τοπικής μεθόδου, αντισταθμίζεται με την αύξηση στην ταχύτητα σύγκλισης. Βέβαια παραμένει και το σημαντικό πλεονέκτημα ότι πετυχαίνουμε να προσεγγίσουμε την πραγματική τιμή του ολικού ελαχίστου με πολύ μεγάλη ακρίβεια.

Στο σημείο αυτό πρέπει να σημειώσουμε ότι η μέθοδος αυτή δεν είναι στατιστικά εγγυημένο, ότι θα καταφέρει να βρει την πραγματική τιμή του ολικού ελαχίστου, πρακτικά όμως όπως θα δούμε και παρακάτω στα αποτελέσματα διαφόρων πειραμάτων, κρίνεται πολύ αποδοτική και ακριβής.



Το γενικό σχήμα της υβριδικής αυτής μεθόδου παρατίθεται στη συνέχεια. Τονίζεται ότι χρησιμοποιείται η δεύτερη έκδοση του αλγορίθμου SA (4.3.3.3).

Ο αλγόριθμος του SALT

Εφάρμοσε μια μέθοδο τοπικής ελαχιστοποίησης με αρχικό σημείο το x
 $x = \text{τοπικό ελάχιστο } x^*$

Βήμα 1

Επανάλαβε N_A φορές (μείωση θερμοκρασίας)

Επανάλαβε N_T φορές (προσαρμογή συντελεστή μετατόπισης h)

Επανάλαβε N_S φορές (βασική κίνηση αλγορίθμου)

$x' = x$ (ολόκληρο το διάνυσμα)

Αν $T = 0.001 * T_{\text{αρχ}}$ (όπου $T_{\text{αρχ}}$ η αρχική θερμοκρασία) τότε εφάρμοσε μια τοπική μέθοδο (με κάποια πιθανότητα), για λίγες επαναλήψεις

$x = x^*$ (όπου x^* το σημείο που προκύπτει από τοπ. ελαχιστοποίηση)

Πάρε ένα νέο γειτονικό σημείο, (ενημερωμένο ως προς όλες τις παραμέτρους)

$x'_i = x_i + \Delta x_i$

(Το Δx_i δίνεται από τον τύπο (4.3.3.1))

$f_i = f(x')$ (και $f_c = f(x)$)

Αν $f_i \leq f_{\text{best}}$ τότε

$x_{\text{best}} = x'$ και $f_{\text{best}} = f_i$

Αν $f_i < f_c$ τότε

$x = x'$ και $f_c = f_i$

Αλλιώς

Υπολόγισε την πιθανότητα $p = e^{-(f_i - f_c)/T}$

Αν το $p > \xi$ (όπου ξ τυχαίος αριθμός στο διάστημα $[0, 1]$) τότε

$x = x'$ και $f_c = f_i$

Τέλος Αν ($f_i < f_c$)

Τέλος N_S επαναλήψεων

Υπολογισμός των παραμέτρων S_i σύμφωνα με τους τύπους (4.3.3.2) και (4.3.3.3)

Τέλος N_T επαναλήψεων

Προσαρμογή συντελεστή μετατόπισης h (όπως ακριβώς και στον προηγούμενο αλγόριθμο, (4.3.3))

Τέλος N_A επαναλήψεων

Μείωση θερμοκρασίας : $T = \alpha T$

$x = x_{\text{best}}$, $f_c = f_{\text{best}}$

Αν ικανοποιείται το κριτήριο τερματισμού,

Τότε, τέλος αλγορίθμου

Αλλιώς, επιστροφή στο βήμα 1

4.5 Πειραματικά αποτελέσματα

	Ολ. ελάχιστο	SA1 Ελάχ. τιμή	SA1 Αποτιμήσεις	SA2 Ελάχ.τιμή	SA2 Αποτιμήσεις	SALT Ελάχ. τιμή	SALT Αποτιμήσεις
Sixhump	-1.0316285	-1.0316284	5329	-1.0316284	6083	-1.0316284	2429
Goldstein	3	3.000423	5123	3.000672	6123	3	2678
Rastrigin	-2	-1.9888888	4804	-1.9999998	5503	-2	5901
Griewank2	0	0.0000001	12404	0.00000014	14671	0	6103
Neural	0	0.0000198	150000	0.0000532	150000	0.00000076	127855
Griewank10	0	0.0002238	150000	0.0002145	150000	0.00000013	103751

Τα παραπάνω πειράματα (δες παράρτημα) έγιναν χρησιμοποιώντας τα ακόλουθα κριτήρια τερματισμού

- Τερματισμός όταν προσεγγίζεται η πραγματική τιμή του ολικού ελαχίστου
- Τερματισμός μετά από 150000 αποτιμήσεις της συνάρτησης (function evaluations)
- Ικανοποίηση του κριτηρίου τερματισμού, όπως αυτό περιγράφηκε νωρίτερα

Συμβολισμοί

SA1 : ο αλγόριθμος του Goffe (4.3.1)

SA2 : δεύτερη έκδοση αλγορίθμου simulated annealing (4.3.3)

SALT : υβριδική μέθοδος (4.4.2)

Σχόλια – Συμπεράσματα

Η μέθοδος του SALT κρίνεται η πιο αποδοτική από τις παραπάνω μεθόδους. Πετυχαίνει μεγάλη ακρίβεια αποτελέσματος. Η κλήση λοιπόν της τοπικής μεθόδου στα τελευταία στάδια του αλγορίθμου είναι πολύ επιτυχής, αφού ξεκινάει από πολύ καλά αρχικά σημεία και προσεγγίζει σε πολύ λίγα βήματα με μεγάλη ακρίβεια, το πραγματικό ολικό ελάχιστο.

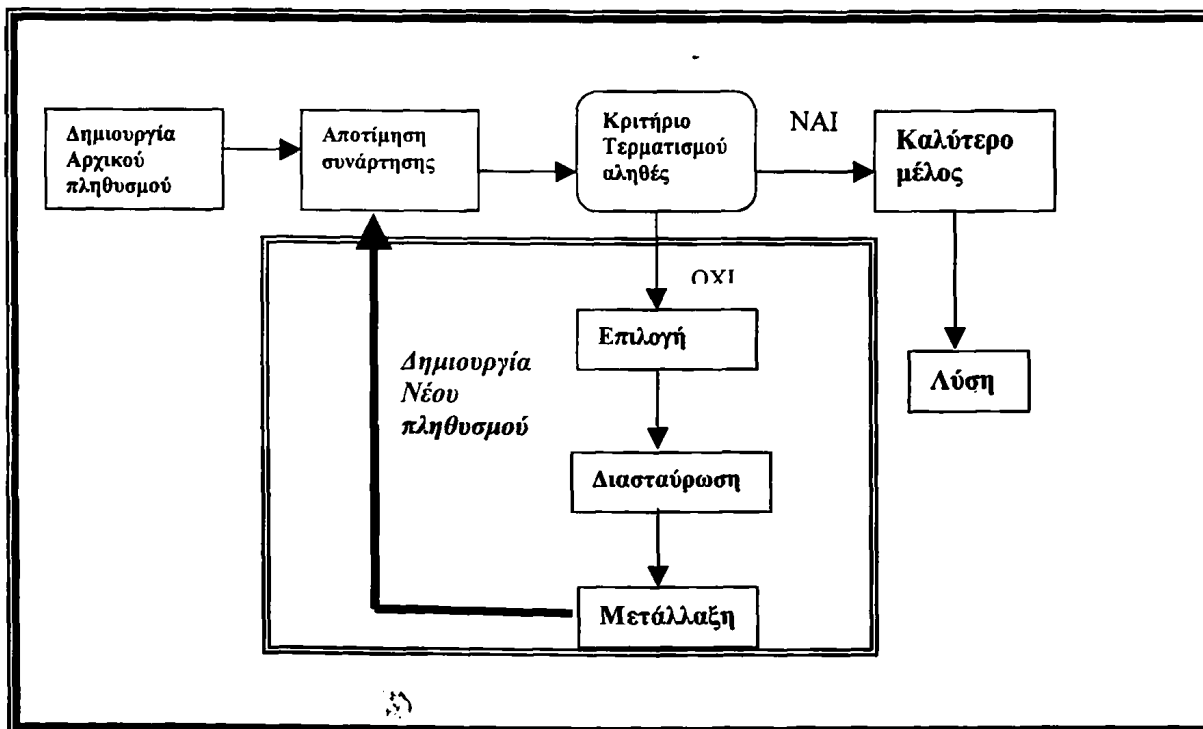
Κεφάλαιο 5 – Γενετικοί αλγόριθμοι

5.1 Εισαγωγή

Εμφανίζουν αναλογία με τα μοντέλα βιολογικής εξέλιξης, στα οποία από τον πληθυσμό της αρχικής γενιάς παίρνουμε νέους πληθυσμούς με την χρήση των γνωστών γενετικών τελεστών της διασταύρωσης και της μετάλλαξης. Ο πληθυσμός αποτελείται από ένα σύνολο σημείων, στα οποία ορίζεται η συνάρτηση. Πιο κατάλληλα θεωρούνται τα σημεία στα οποία η τιμή της συνάρτησης είναι μικρή. Αυτά έχουν και μεγάλη πιθανότητα να επιβιώσουν στις επόμενες γενιές. Τα μέλη της επόμενης γενιάς, προκύπτουν από την εφαρμογή των γενετικών τελεστών. Οι μεταλλάξεις επιστρέφουν σημεία (μέλη πληθυσμού), που είναι κοντά σε σχέση με τα προηγούμενα, ενώ οι διασταυρώσεις επιφέρουν ποικιλόμορφους πληθυσμούς. Η αποδοτικότητα αυτών των αλγορίθμων, εξαρτάται από την επιλογή των γενετικών τελεστών. Έχει παρατηρηθεί ότι οι γενετικοί αλγόριθμοι έχουν καλά αποτελέσματα για πολλά προβλήματα, έστω κι αν απαιτείται αρκετός χρόνος για να δώσουν μια αξιόπιστη λύση. Ο συνδυασμός τους με κάποια μέθοδο τοπικής ελαχιστοποίησης, έχει αποδειχτεί ότι μπορεί να επιφέρει καλύτερα αποτελέσματα, όπως θα περιγράψουμε στη συνέχεια.

5.2 Δομή του γενετικού αλγορίθμου

Η δομή ενός γενετικού αλγορίθμου [22], [32] στην εφαρμογή προβλημάτων εύρεσης ολικού ελαχίστου δίνεται από το παρακάτω σχήμα



Σχήμα 5.1 Γενικό σχήμα γενετικού αλγορίθμου για ολική ελαχιστοποίηση

5.3 Αναπαράσταση μελών της γενιάς

Πολλές υλοποιήσεις σε προβλήματα ολικής ελαχιστοποίησης, χρησιμοποιούν δυαδική αναπαράσταση των παραμέτρων, όπως συνηθίζεται στις περισσότερες εφαρμογές των γενετικών αλγορίθμων. Αυτό έχει ως αποτέλεσμα να αντιμετωπίζονται συνεχή προβλήματα σαν διακριτά. Επίσης θα πρέπει το αποτέλεσμα που επιστρέφεται σε δυαδική μορφή να μετατρέπεται σε μια αναπαράσταση πραγματικών αριθμών, ώστε να είναι σε συμβατή μορφή για τον χρήστη. Έτσι απαιτούνται αλγόριθμοι μετατροπής, τόσο για την μετατροπή των πραγματικών αριθμών σε δυαδική μορφή, όσο και για το αντίστροφο.

Σύμφωνα λοιπόν με τα παραπάνω γίνεται φανερό η ανάγκη εισαγωγής ενός νέου μοντέλου αναπαράστασης των μελών του πληθυσμού. Στα περισσότερα λοιπόν προβλήματα ελαχιστοποίησης, επιλέγεται η αναπαράσταση με χρήση πραγματικών αριθμών [33] (**floating point representation**). Σε αυτή την περίπτωση κάθε πραγματικός αριθμός αναπαριστά και το αντίστοιχο μέλος του πληθυσμού, χωρίς κάποιο ενδιάμεσο βήμα μετατροπής σε άλλη μορφή. Η αναπαράσταση αυτή έχει πολλά πλεονεκτήματα [23] όπως :

- Καλύτερη προσαρμογή σε συνεχή προβλήματα
- Αύξηση της ταχύτητας αναζήτησης, αφού δεν απαιτείται μετατροπή πραγματικών αριθμών σε δυαδικούς κι αντίστροφα
- Συνδυάζονται καλύτερα οι γενετικοί αλγόριθμοι με άλλες μεθόδους (με μεθόδους για τοπική ελαχιστοποίηση)

Τα κυριότερα μειονεκτήματά τους είναι:

- Ανάγκη επινόησης νέων γενετικών τελεστών
- Δεν υπάρχουν ισχυρές μαθηματικές αποδείξεις, όπως συμβαίνει στην περίπτωση των γενετικών αλγορίθμων με δυαδική αναπαράσταση

Κριτήρια τερματισμού

Το πιο απλό κριτήριο τερματισμού είναι, όταν τα μέλη του πληθυσμού στην τρέχουσα γενιά έχουν παρόμοιες τιμές. Ένα άλλο κριτήριο, είναι να ελέγχουμε την βελτίωση στην συνάρτηση καταλληλότητας. Όταν η βελτίωση της είναι κάτω από ένα όριο κατωφλίου, τότε τερματίζεται ο αλγόριθμος. Επίσης, κάτι άλλο που είναι ευρέως διαδεδομένο, είναι ο τερματισμός μετά από ένα συγκεκριμένο πλήθος γενεών.

5.4 Περιγραφή αλγορίθμου

- **Βήμα 1** : Στην αρχή του αλγορίθμου δημιουργούνται τυχαία μέλη για τον αρχικό πληθυσμό. Η τιμή της συνάρτησης αποτιμάται γι' αυτά τα μέλη του πληθυσμού. Έτσι παράγεται η πρώτη γενεά.
- **Βήμα 2** : Εφαρμόζεται η διαδικασία της επιλογής, κατά την οποία επιλέγονται με μεγαλύτερη πιθανότητα τα μέλη με μεγάλη τιμή καταλληλότητας (εδώ όσο πιο μικρή τιμή είναι η τιμή της συνάρτησης, τόσο πιο μεγάλη είναι η τιμή για την συνάρτηση της καταλληλότητας).
- **Βήμα 3** : Με κάποια πιθανότητα διασταυρώνονται μέλη της γενιάς για να δώσουν καινούρια μέλη στην επόμενη γενιά.
- **Βήμα 4** : Με κάποια πιθανότητα εφαρμόζεται για τα διάφορα μέλη ο τελεστής της **μετάλλαξης (mutation)** - συνοπτικά η μετάλλαξη δεν είναι τίποτε άλλο από μια μικρή αλλαγή στο μέλος του πληθυσμού (πχ προσθήκη θορύβου)
- **Βήμα 5** : Γίνεται αποτίμηση της συνάρτησης καταλληλότητας για τα καινούρια μέλη του πληθυσμού κι ελέγχεται το κριτήριο τερματισμού του αλγορίθμου.
 - **Βήμα 5.1** Αν αυτό ικανοποιείται, τότε επιστρέφεται σαν λύση η τιμή εκείνου του μέλους του πληθυσμού, που έχει την μεγαλύτερη τιμή για την συνάρτηση καταλληλότητας.
 - **Βήμα 5.2** Αν δεν ικανοποιείται το κριτήριο, τότε πηγαίνει στο βήμα 2 κι επανέλαβε την διαδικασία από εκεί και κάτω.

Το παραπάνω γενικό σχήμα, έχει δώσει λύσεις σε πολλά και πολύπλοκα προβλήματα ελαχιστοποίησης. Πέρα από αυτό το μοντέλο, υπάρχουν και διάφορα άλλα μοντέλα γενετικών αλγορίθμων πιο πολύπλοκα, με αρκετά καλή απόδοση στην επίλυση συγκεκριμένων δύσκολων προβλημάτων.

5.5 Επιλογή (Selection)

Με τον γενετικό τελεστή της επιλογής αποφασίζουμε ποια μέλη θα συνεχίσουν να υφίστανται αυτούσια, ώστε να εφαρμοστούν στη συνέχεια σε αυτά οι τελεστές της διασταύρωσης και της μετάλλαξης. Η επιλογή γίνεται με βάση την τιμή της συνάρτησης καταλληλότητας των διαφόρων μελών. Όσο πιο μεγάλη είναι αυτή η τιμή σε σχέση με την τιμή των άλλων μελών, τόσο αυξάνεται η πιθανότητα να επιλεγεί το συγκεκριμένο μέλος. Αν κάποιο μέλος έχει μεγάλη τιμή καταλληλότητας σε σχέση με τα υπόλοιπα μέλη, μπορεί να επιλεγεί παραπάνω από μια φορές. Όταν οι γενετικοί αλγόριθμοι εφαρμόζονται σε προβλήματα ολικής ελαχιστοποίησης η τιμή καταλληλότητας για κάθε μέλος του πληθυσμού δίνεται από τον τύπο :

Τιμή καταλληλότητας για το μέλος $i = (F_{MAX} - F(X) + \epsilon) / (F_{MAX} - F_{MIN} + \epsilon)$

αν $F_{MAX} \neq F_{MIN}$

F_{MAX} : η μέγιστη τιμή της συνάρτησης των μελών του πληθυσμού

F_{MIN} : η ελάχιστη τιμή της συνάρτησης των μελών του πληθυσμού

$F(X)$: η τιμή της συνάρτησης για το μέλος X του πληθυσμού



ε : θετικός αριθμός (με μικρή τιμή), ώστε να επιλέγεται κι εκείνο το μέλος του πληθυσμού, που η τιμή της συνάρτησης είναι μέγιστη

Σύμφωνα με τον παραπάνω τύπο τα μέλη εκείνα για τα οποία η τιμή της συνάρτησης είναι πολύ μικρή, έχουν μεγάλη τιμή καταλληλότητας κι αντίστροφα. Επίσης, όταν η μέγιστη και η ελάχιστη τιμή της συνάρτησης είναι ίσες, όλα τα μέλη λαμβάνουν την ίδια τιμή καταλληλότητας. Παρακάτω θα περιγράψουμε μια από τις πιο σημαντικές και διαδεδομένες μεθόδους επιλογής. Αυτή είναι η μέθοδος *roulette-wheel selection* (η μέθοδος του τροχού της τύχης).

Μέθοδος του τροχού της τύχης (Roulette wheel selection)

Αυτό είναι το πιο απλό σχήμα επιλογής κι ονομάζεται επίσης και **στοχαστική δειγματοληψία με αντικατάσταση** (stochastic sampling with replacement). Είναι ένας στοχαστικός αλγόριθμος που χρησιμοποιεί την ακόλουθη τεχνική :

Θεωρούμε ότι οι *τιμές καταλληλότητας* των διαφόρων μελών του πληθυσμού είναι κανονικοποιημένες στο διάστημα $[0,1]$. Οι τιμές καταλληλότητας των μελών του πληθυσμού παριστάνονται με ισάριθμα ευθύγραμμα τμήματα, που το μήκος τους είναι ανάλογο με την τιμή καταλληλότητας. Αυτά τα ευθύγραμμα τμήματα είναι διαδοχικά, οπότε το συνολικό ευθύγραμμο τμήμα που προκύπτει έχει μήκος 1 (αφού οι τιμές καταλληλότητας των διαφόρων μελών είναι κανονικοποιημένες). Επιλέγουμε τυχαία έναν αριθμό, ομοιόμορφα στο διάστημα $[0,1]$. Ο αριθμός αυτός θα ανήκει σε ένα από τα διαδοχικά ευθύγραμμα τμήματα. Επιλέγεται λοιπόν εκείνο το μέλος του πληθυσμού, που αντιστοιχεί στο ευθύγραμμο τμήμα, που περιέχει τον τυχαίο αριθμό. Η διαδικασία αυτή επαναλαμβάνεται μέχρι την επιλογή του επιθυμητού πλήθους μελών του πληθυσμού για την επόμενη γενιά. Είναι φανερό ότι μεγάλη τιμή καταλληλότητας για κάποιο μέλος, συνεπάγεται ότι θα καταλαμβάνει και μεγάλο μήκος στο συνολικό ευθύγραμμο τμήμα. Έτσι η πιθανότητα να επιλεγεί είναι μεγάλη.

Ο πίνακας 1 μας δείχνει την πιθανότητα επιλογής 11 μελών του πληθυσμού. Το πρώτο μέλος έχει την μεγαλύτερη τιμή καταλληλότητας και το διάστημά του πάνω στην ευθεία διαστημάτων, καταλαμβάνει το μεγαλύτερο μήκος. Το δέκατο μέλος του πληθυσμού είναι εκείνο με την χαμηλότερη τιμή καταλληλότητας και συνεπώς το διάστημά του έχει και το μικρότερο μήκος. Το εντέκατο μέλος του πληθυσμού έχει τιμή καταλληλότητας 0 κι επομένως καμιά πιθανότητα επιλογής για αναπαραγωγή.

Πίνακας 1

Αριθμός μέλους	1	2	3	4	5	6	7	8	9	10	11
Τιμή καταλληλότητας	2.0	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2	0.0
Πιθανότητα επιλογής	0.18	0.16	0.15	0.13	0.11	0.09	0.07	0.06	0.03	0.02	0.0

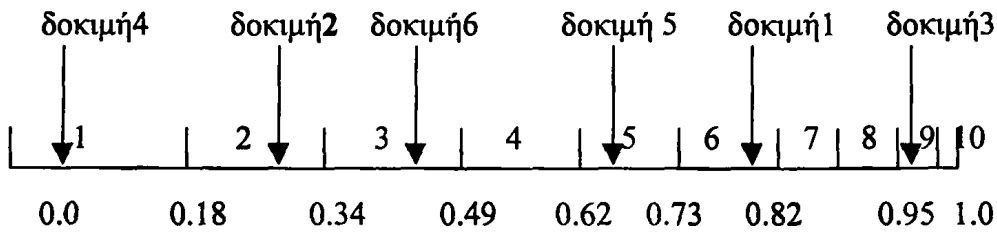
Ανάλογα με το πλήθος των μελών που θέλουμε να επιλέξουμε, τραβάμε τόσους τυχαίους αριθμούς. Οι αριθμοί αυτοί θα είναι μεταξύ 0 και 1. Εστω λοιπόν ότι έχουμε το εξής δείγμα από τυχαίους αριθμούς

Δείγμα τυχαίων αριθμών

0.81 , 0.32 , 0.96 , 0.01 , 0.65 , 0.42



Η επόμενη εικόνα δείχνει τα επιλεγόμενα μέλη μετά την εφαρμογή των προηγούμενων προσπαθειών. Η επιλογή βασίζεται στην τιμή καταλληλότητας, όπως αυτή δίνεται από τον προηγούμενο πίνακα.



Σχήμα 7.2 Σχηματική αναπαράσταση της μεθόδου του τροχού της τύχης για την επιλογή των μελών του πληθυσμού.

Μετά την επιλογή προκύπτουν τα εξής μέλη σύμφωνα με το παραπάνω σχήμα

1, 2, 3, 5, 6, 9

5.6 Διασταύρωση (crossover) - Real-valued recombination

5.6.1 Διακριτή διασταύρωση

Εδώ εφαρμόζεται ανταλλαγή των τιμών των μεταβλητών, ανάμεσα στα διάφορα μέλη που διασταυρώνονται. Ας θεωρήσουμε το ακόλουθο παράδειγμα, όπου έχουμε δύο μέλη με τρεις μεταβλητές το καθένα.

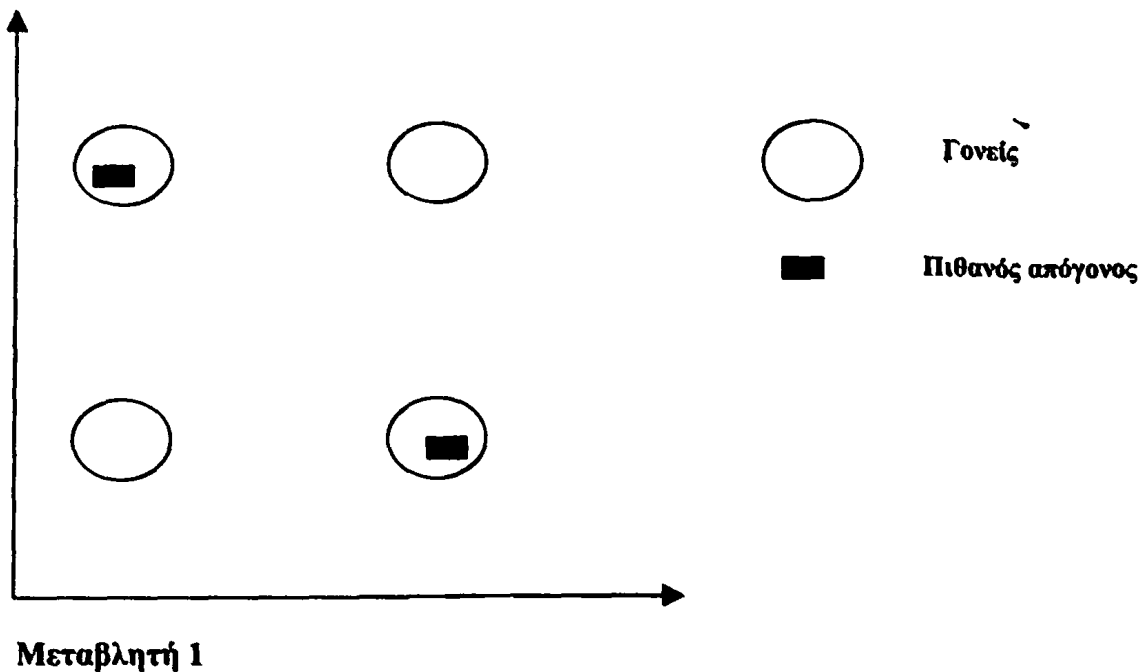
Μέλος 1	12	25	5
Μέλος 2	123	4	34

Για κάθε μεταβλητή επιλέγεται με ίση πιθανότητα, το μέλος εκείνο που προσφέρει την τιμή του, για την αντίστοιχη μεταβλητή, μετά την αναπαραγωγή. Η επιλογή αυτού του μέλους γίνεται με ίση πιθανότητα

Δείγμα 1	2	2	1
Δείγμα 2	1	2	1

Μετά την αναπαραγωγή προκύπτουν τα εξής μέλη, που έχουν τις παρακάτω τιμές στις θέσεις των διαφόρων μεταβλητών

Απόγονος 1	123	4	5
Απόγονος 2	12	4	5



Σχήμα 7.3 Πιθανές θέσεις των απογόνων μετά την εφαρμογή της διακριτής διασταύρωσης.

Το παραπάνω σχήμα δείχνει το γεωμετρικό αποτέλεσμα της διακριτής διασταύρωσης (discrete recombination)

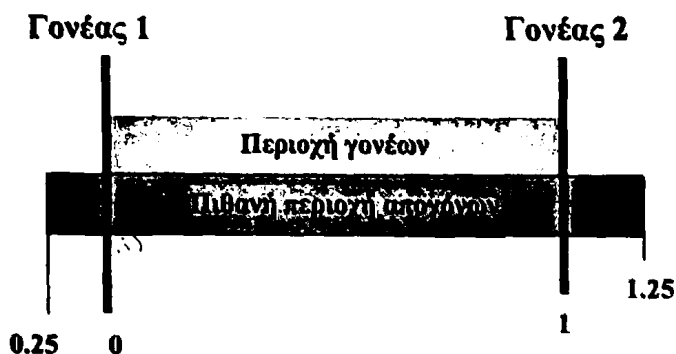
5.6.2 Ενδιάμεση διασταύρωση - Intermediate recombination

Οι τιμές των μεταβλητών που προκύπτουν μετά από αυτό το είδος της διασταύρωσης, είναι κοντά ή γύρω στην προηγούμενη περιοχή που ορίζονται οι μεταβλητές. Αυτό μπορούμε να το διαπιστώσουμε κι από το παρακάτω σχήμα

Τα μέλη που παράγονται μετά την διασταύρωση δίνονται από τον παρακάτω κανόνα :

$$\text{Απόγονος} = \text{γονέας 1} + \text{Alpha} (\text{γονέας 2} - \text{γονέας 1})$$

Η παράμετρος Alpha επιλέγεται τυχαία από ένα διάστημα $[-d, 1+d]$. Μια καλή τιμή για το d είναι 0.25. Η νέα τιμή για κάθε μεταβλητή παράγεται σύμφωνα με τον προηγούμενο τύπο, επιλέγοντας για κάθε μεταβλητή μια νέα τιμή για την παράμετρο Alpha.



Σχήμα 7.4 Σχηματική αναπαράσταση της πιθανής περιοχής απογόνων

Ας θεωρήσουμε τα ακόλουθα μέλη του πληθυσμού με 3 μεταβλητές το καθένα :

Μέλος 1	12	25	5
Μέλος 2	123	4	34

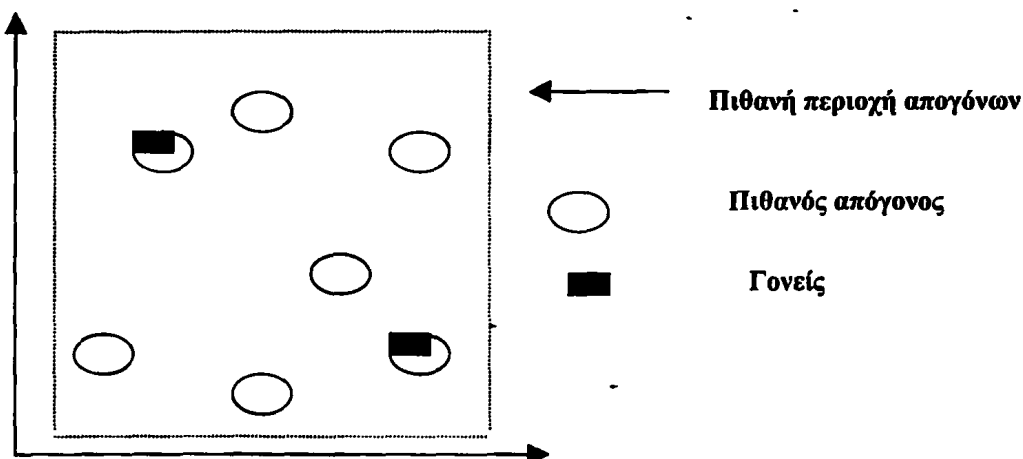
Οι επιλεγόμενες τιμές για την παράμετρο Alpha είναι :

Δείγμα 1	0.5	1.1	-0.1
Δείγμα 2	0.1	0.8	0.5

Τα νέα μέλη του πληθυσμού είναι τα ακόλουθα :

Απόγονος 1	67.5	1.9	2.1
Απόγονος 2	23.1	8.2	19.5

Το ακόλουθο σχήμα μας δείχνει την περιοχή που μπορούμε να έχουμε πιθανές τιμές για τα μέλη που θα προκύψουν μετά την διασταύρωση.



Σχήμα 7.5 Πιθανές θέσεις των απογόνων μετά την εφαρμογή της ενδιάμεσης διασταύρωσης

5.7 Μετάλλαξη (mutation)

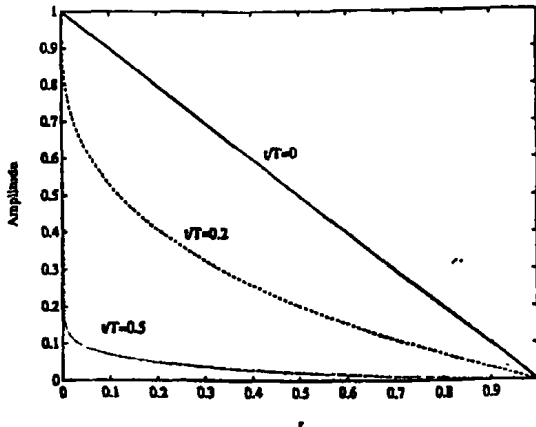
Έστω $x = (x_1, x_2, \dots, x_n)$ ένα μέλος του πληθυσμού κάποιας γενεάς. Κάθε μέλος του πληθυσμού στο οποίο εφαρμόζεται μετάλλαξη θα υποστεί μια τροποποίηση στην τιμή του, που για τις πρώτες γενιές θα είναι αρκετά μεγάλη, ενώ με την πάροδο των γενεών θα είναι πολύ μικρότερη, τείνοντας τελικά στο μηδέν.

Για μια συγκεκριμένη γενεά t επιλέγεται ένας τυχαίος αριθμός p , στο διάστημα $[0,1]$. Αν ο αριθμός αυτός είναι μεγαλύτερος από 0.5 τότε θα συμβεί κάποια θετική αλλαγή, αλλιώς έχουμε αρνητική αλλαγή. Επίσης επιλέγεται κι ένας άλλος τυχαίος αριθμός έστω r , που υποδηλώνει το μέτρο της αλλαγής στην τιμή της μεταβλητής και παίρνει τιμές στο διάστημα $[0,1]$. Έτσι λοιπόν οι νέες τιμές για τις μεταβλητές θα είναι :

$$x_k^{new} = x_k + (x_k^{max} - x_k) \{1 - r^{[1-(t/T)]^b}\} \quad \text{αν } p > 0.5$$

$$x_k^{new} = x_k + (x_k - x_{min}^k) \{1 - r^{[1-(t/T)]^b}\} \quad \text{αν } p \leq 0.5$$

Οι παράμετροι x_{max} και x_{min} είναι το δεξί κι αριστερό άκρο του πεδίου ορισμού των μεταβλητών. Το b είναι ο ρυθμός μείωσης της μετάλλαξης (έχει τιμή 5). Με t συμβολίζουμε τον αύξοντα αριθμό της γενιάς, για την οποία ο ρυθμός τροποποίησης τείνει στο μηδέν γεγονός, που σημαίνει ότι σχεδόν δεν γίνεται μετάλλαξη (ασήμαντη μεταβολή των τιμών των μεταβλητών του αντίστοιχου μέλους του πληθυσμού). Με T συμβολίζεται το συνολικό πλήθος των γενεών.



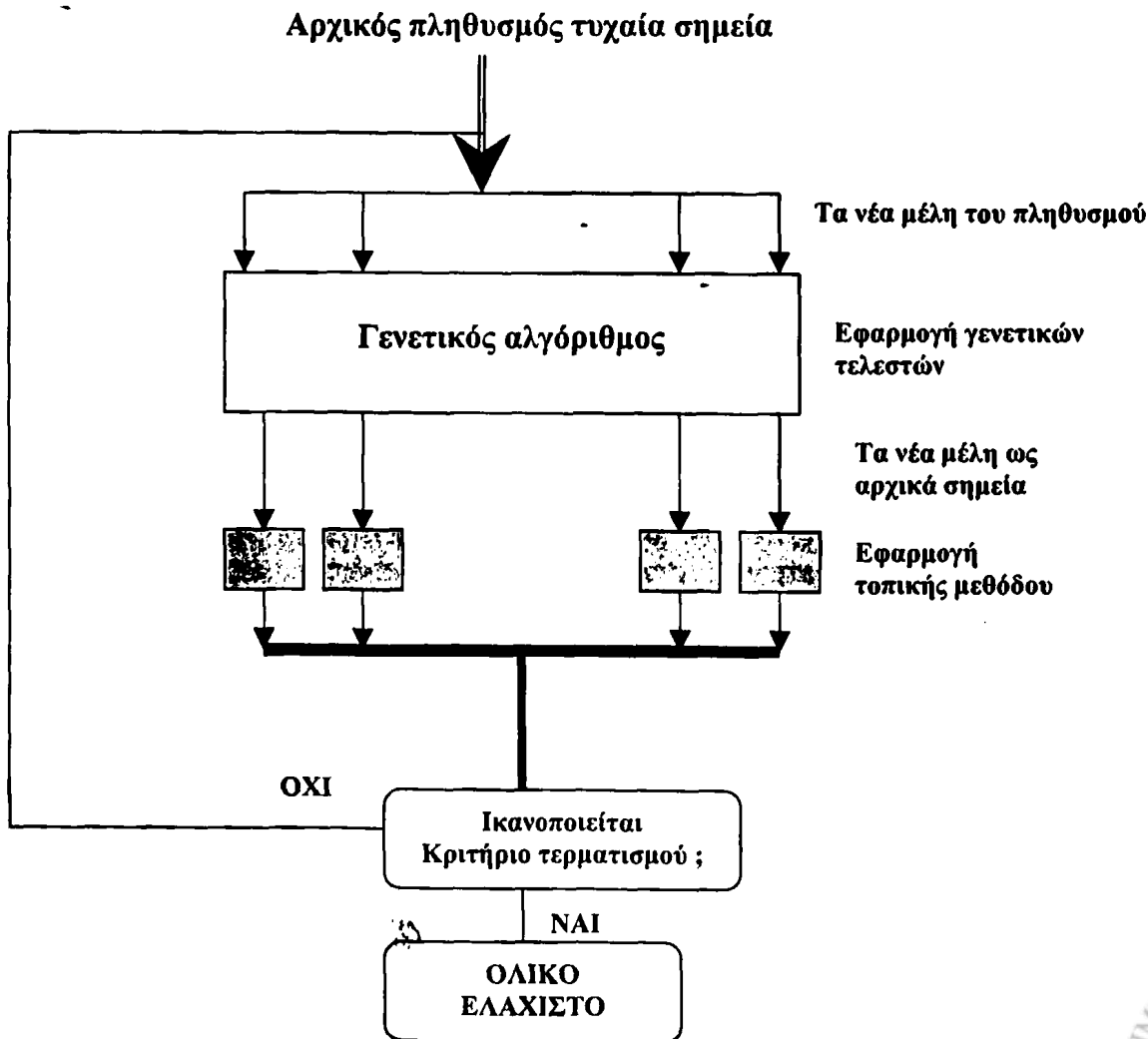
Σχήμα 7.6 Γραφική παράσταση του μέτρου της μετάλλαξης (amplitude), σε σχέση με τον τυχαίο αριθμό r .

Το μέτρο της μετάλλαξης είναι μεγάλο, όταν τα νέα μέλη που προκύπτουν είναι πολύ διαφορετικά από τα προηγούμενα. Παρατηρούμε ότι στις πρώτες γενιές, η μετάλλαξη είναι αξιοσημείωτη, ενώ με την πάροδο των γενεών είναι σχεδόν ανεπαίσθητη. Είναι κάτι αντίστοιχο με την θερμοκρασία στην μέθοδο του simulated annealing, όπου στις αρχικές υψηλές θερμοκρασίες έχουμε μεγάλο βήμα μετατόπισης, το οποίο συνεχώς φθίνει και σχεδόν μηδενίζεται σε πολύ χαμηλές θερμοκρασίες.

5.8 Υβριδική μέθοδος

5.8.1 Γενική ιδέα

Η γενική ιδέα είναι να χρησιμοποιήσει κανείς ένα συνδυασμό γενετικών αλγορίθμων με κάποια τεχνική τοπικής ελαχιστοποίησης, ώστε η μέθοδος που θα προκύψει να συνδυάζει τόσο τα πλεονεκτήματα της μιας όσο και της άλλης μεθόδου, απαλορίζοντας τα αντίστοιχα μειονεκτήματα [22]. Οι γενετικοί αλγόριθμοι έχουν την δυνατότητα να καλύπτουν όλη την περιοχή ενδιαφέροντος, εφαρμόζοντας τους γενετικούς τελεστές, ενώ η τοπική μέθοδος μάς βοηθάει να εντοπίσουμε εύκολα το τοπικό ελάχιστο σε συγκεκριμένες περιοχές. Η ποικιλία στα μέλη του πληθυσμού (σημεία συνάρτησης) που προσφέρουν οι γενετικοί τελεστές, βοηθά στο να απεγκλωβιστούμε από περιοχές, οι οποίες οδηγούν στο ίδιο τοπικό ελάχιστο. Αρχικά στις πρώτες γενιές ο πληθυσμός παρουσιάζει μεγάλη ανομοιομορφία, ώστε να καλυφθεί όσο το δυνατόν το μεγαλύτερο μέρος του πεδίου ορισμού των μεταβλητών. Η τοπική μέθοδος που εφαρμόζεται, βοηθάει στο να εντοπιστεί, το αντίστοιχο τοπικό ελάχιστο. Όσο αυξάνεται το πλήθος των γενεών, ο πληθυσμός γίνεται ολοένα και περισσότερο ομοιόμορφος και συγκλίνει στο ολικό ελάχιστο. Παρακάτω θα εξηγήσουμε τις δύο πιο διαδεδομένες προσεγγίσεις στην εφαρμογή μεθόδων με συνδυασμό γενετικών αλγορίθμων και τοπικής ελαχιστοποίησης. Το παρακάτω σχήμα περιγράφει την βασική φιλοσοφία στην οποία στηρίζονται οι υβριδικές μέθοδοι.



Σχήμα 7.7 Γενικό σχήμα υβριδικών μεθόδων (γενετικοί αλγόριθμοι και τοπική στοποίηση).

5.8.2 Θέματα προς συζήτηση για τις υβριδικές μεθόδους

Ο συνδυασμός ενός αλγορίθμου για τον υπολογισμό ενός ολικού ελαχίστου με χρήση γενετικών αλγορίθμων και μεθόδου τοπικής ελαχιστοποίησης, εμφανίζεται με διάφορες μορφές, που προσπαθούν κυρίως να αντιμετωπίσουν τα παρακάτω προβλήματα. Η γενική ιδέα, όπως αυτή περιγράφηκε στο πιο πάνω σχήμα, εξακολουθεί να παραμένει η ίδια. Τα κυριότερα από αυτά τα προβλήματα είναι τα εξής [23] :

Πόσο συχνά θα πρέπει να καλείται μια μέθοδος τοπικής ελαχιστοποίησης και για ποια μέλη του πληθυσμού

Σύμφωνα με την προηγούμενη περιγραφή της υβριδικής μεθόδου, κάθε μέλος του πληθυσμού αποτελεί κι ένα αρχικό σημείο για μια μέθοδο τοπικής ελαχιστοποίησης. Το κόστος που προκύπτει από την κλήση της μεθόδου τοπικής ελαχιστοποίησης, μας περιορίζει στο να χρησιμοποιούμε πληθυσμούς με λίγα μέλη και να επιτρέπουμε τον αλγόριθμο να τρέξει για συγκεκριμένο πλήθος γενεών. Έστω, ότι ο αλγόριθμος τρέχει για 1000 γενεές και ο πληθυσμός της κάθε γενεάς αποτελείται από 20 μέλη. Τότε θα γίνει κλήση μιας τοπικής μεθόδου για 20000 φορές. Επιπλέον, η εκκίνηση τοπικής μεθόδου από κάθε μέλος του πληθυσμού, δεν αυξάνει την αποδοτικότητα του αλγορίθμου, αφού τα τοπικά ελάχιστα που επιστρέφονται, είναι πολύ μεγαλύτερα από την πραγματική τιμή του ολικού ελαχίστου

Ποιος θα είναι ο επιτρεπόμενος αριθμός βημάτων για κάθε τοπική μέθοδο

Θεωρούμε ως περιοχή ενδιαφέροντος την περιοχή εκείνη, της οποίας τα σημεία οδηγούν στο ίδιο τοπικό ελάχιστο. Γίνεται φανερό, ότι στην περιοχή ενδιαφέροντος κοντά στο ολικό ελάχιστο θα πρέπει μια τοπική μέθοδος να εκτελεστεί για αρκετά βήματα, αφού η ακρίβεια της λύσης που θα προκύψει είναι σημαντική. Αν η περιοχή αυτή είναι μακριά από την περιοχή το ολικού ελαχίστου, μεγάλος αριθμός επαναλήψεων για την τοπική μέθοδο απλώς μειώνει την ταχύτητα του αλγορίθμου.

Ποια θα είναι η επιλεγόμενη τοπική μέθοδος

Η επιλογή της τοπικής μεθόδου είναι ένα θέμα αρκετά σημαντικό, αφού επηρεάζει την ακρίβεια της τελικής λύσης. Κατά την επιλογή της τοπικής μεθόδου, θα πρέπει να ληφθεί σοβαρά υπόψη τόσο το κόστος της μεθόδου, όσο και η απόδοσή της. Η μέθοδος εκείνη που συνδυάζει τους δύο παραπάνω παράγοντες με τον καλύτερο τρόπο, θα πρέπει να είναι και η επιλεγόμενη. Για παράδειγμα, για συναρτήσεις που είναι της μορφής αθροίσματος τετραγώνων, συνήθως πιο αποδοτική είναι η μέθοδος **Levenberg-Marquardt**.

5.8.3 Η Δαρβινική προσέγγιση (Darwin inspired method)

Επιλέγονται κάποια τυχαία αρχικά σημεία που αποτελούν τον αρχικό πληθυσμό. Το πλήθος των μελών για κάθε πληθυσμό είναι σταθερό για κάθε γενιά. Όσο μεγαλύτερος είναι ο αριθμός των μελών του πληθυσμού, τόσο αυξάνεται η πιθανότητα να είναι σωστά τα αποτελέσματα του παραπάνω αλγορίθμου, κάτι που είναι άλλωστε χαρακτηριστικό σε κάθε στοχαστική μέθοδο, που ξεκινά από ένα αρχικό δείγμα σημείων. Στη συνέχεια, θεωρούμε κάθε μέλος του πληθυσμού ως αρχικό σημείο, από το οποίο ξεκινάει μια μέθοδος τοπική αναζήτησης. Υπολογίζονται με βάση αυτά τα σημεία κάποια τοπικά ελάχιστα. Η τιμή της συνάρτησης καταλληλότητας για το αντίστοιχο μέλος του

πληθυσμού, είναι αντιστρόφως ανάλογη με την τιμή του τοπικού ελαχίστου που οδηγηθήκαμε. Δηλαδή, αν αυτό το μέλος του πληθυσμού οδήγησε σε ένα τοπικό ελάχιστο με χαμηλότερη τιμή από κάποιο άλλο, θα έχει μεγαλύτερη τιμή η συνάρτηση καταλληλότητας από κάποιο που οδήγησε σε μεγαλύτερη τιμή τοπικού ελαχίστου.

Στη συνέχεια εφαρμόζονται οι γενετικοί τελεστές στα αρχικά σημεία της μεθόδου τοπικής ελαχιστοποίησης κι όχι στα αποτελέσματα, που προκύπτουν μετά την εφαρμογή αυτών των μεθόδων. Οι γενετικοί τελεστές που εφαρμόζονται, είναι κατά σειρά επιλογή, διασταύρωση και μετάλλαξη.

5.8.4 Η Λαμαρκιανή προσέγγιση

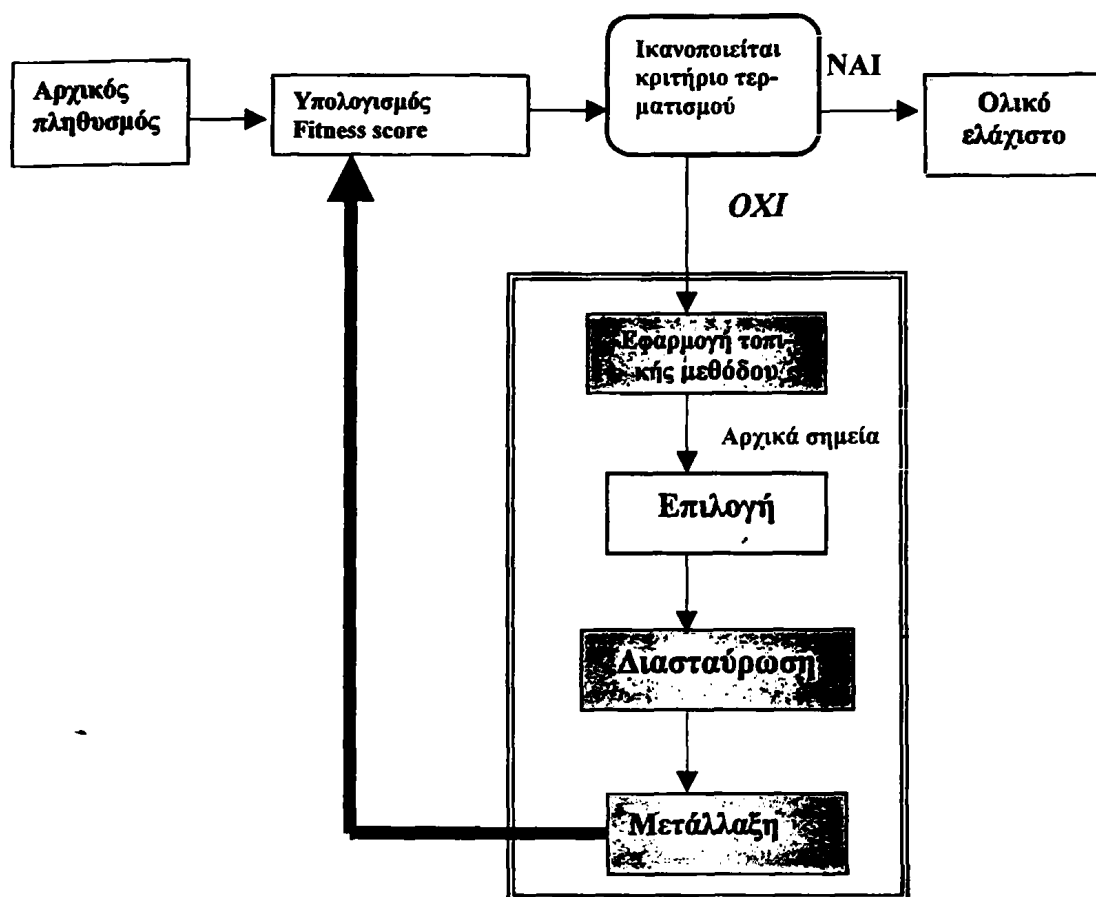
Η ειδοποιός διαφορά αυτής της μεθόδου από την προηγούμενη, είναι ότι η εφαρμογή των γενετικών τελεστών δεν γίνεται πλέον στα αρχικά σημεία της μεθόδου της τοπικής αναζήτησης, αλλά στα ίδια τα αποτελέσματα αυτής της μεθόδου. Τα υπόλοιπα βήματα του αλγορίθμου ακολουθούν την ίδια φιλοσοφία με την προηγούμενη προσέγγιση. Στο σημείο αυτό κρίνεται απαραίτητο να αναφέρουμε τις διαφορές ανάμεσα στις δυο μεθόδους και να τις συγκρίνουμε ως προς τον χρόνο σύγκλισης, την αξιοπιστία και την ακρίβειά τους.

5.8.5 Σύγκριση των δύο παραπάνω προσεγγίσεων

- Η Λαμαρκιανή προσέγγιση έχει μικρότερο χρόνο σύγκλισης αφού για συναρτήσεις με σχετικά ομαλή απεικόνιση και λίγα τοπικά ελάχιστα, φτάνουμε εύκολα και σύντομα σε ομοιόμορφο πληθυσμό.
- Η Δαρβινική προσέγγιση έχει πολύ μεγαλύτερο χρόνο σύγκλισης, αλλά κρίνεται πιο αποτελεσματική (efficient) για συναρτήσεις με πολλά τοπικά ελάχιστα, αφού καλύπτουν μεγαλύτερο εύρος του πεδίου ορισμού των μεταβλητών κι έχουν μεγάλη πιθανότητα απεγκλωβισμού από τοπικά ελάχιστα.

Με βάση την παραπάνω σύγκριση θέλοντας να κάνουμε την μέθοδο μας πιο αποδοτική και πιο γενική, που να καλύπτει δύσκολα προβλήματα στο πεδίο της ολικής ελαχιστοποίησης, υλοποιήθηκε η Δαρβινική προσέγγιση.

5.8.6 Περιγραφή υλοποιημένου αλγορίθμου



Σχήμα 7.8 Γενικό σχήμα υλοποιημένης υβριδικής μεθόδου. Υλοποιήθηκε η Δαρβινική προσέγγιση.

Λεπτομέρειες υλοποίησης

- Χρησιμοποιήθηκε η Δαρβινική προσέγγιση
- Εφαρμόστηκε η bfgs, ως μέθοδος τοπικής ελαχιστοποίησης
- Για την επιλογή, την διασταύρωση και την μετάλλαξη χρησιμοποιήθηκαν οι τα τελεστές που περιγράφηκαν στις 5.5 – 5.7
- **Κριτήριο τερματισμού** : ένα καθορισμένο πλήθος γενεών

5.9 Πειραματικά αποτελέσματα

	GEN	GEN	GENLOC	GENLOC	
	Ολ.Ελάχιστο	Ελ. τιμή	Αποτιμήσεις	Ελ. τιμή	Αποτιμήσεις
Sixhump	-1.0316285	-1.0316283	29830	-1.0316284	1690
Goldstein	3	3.0003456	25456	3	2365
Rastrigin	-2	-1.9999999	41813	-2	5206
Griewank2	0	0	33550	0	7828
Neural	0	0.000015	150000	0.00000004	27892
Griewank10	0	0.000067	81933	0.00000017	10197

Με **GEN** συμβολίζεται ο απλός γενετικός αλγόριθμος.

Με **GENLOC** συμβολίζεται η υβριδική μέθοδος, που συνδυάζει τους γενετικούς αλγόριθμους με κάποια τοπική μέθοδο.

(Για περισσότερες πληροφορίες, σχετικές με τις συναρτήσεις ελέγχου, δες παράρτημα).

Σχόλια - συμπεράσματα

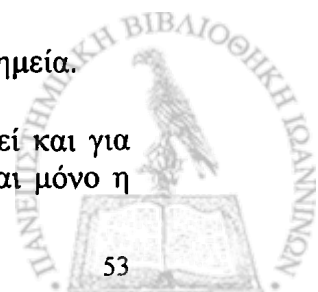
Οι απλοί γενετικοί αλγόριθμοι χρειάζονται πολλές αποτιμήσεις της συνάρτησης για να φτάσουν σε ικανοποιητική λύση. Μεγάλη δυσκολία παρουσιάζεται κυρίως σε συναρτήσεις με πολλές παραμέτρους. Επίσης, όταν το διάστημα ορισμού των παραμέτρων της είναι μεγάλο, απαιτούνται πολλές κλήσεις της συνάρτησης, ώστε να διερευνηθεί όσο το δυνατόν μεγαλύτερο κομμάτι του διαστήματος αυτού. Το μεγάλο τους πλεονέκτημα είναι ότι δεν χρειάζεται ο υπολογισμός παραγώγου.

Η υβριδική μέθοδος εμφανίζεται πολύ πιο αποδοτική συγκριτικά με τους απλούς γενετικούς αλγόριθμους. Συγκλίνει πολύ εύκολα από τις πρώτες γενιές, κυρίως όταν εφαρμόζεται σε συναρτήσεις με λίγα τοπικά ελάχιστα. Οι γενετικοί αλγόριθμοι προσφέρουν μεγάλη ποικιλία καλών αρχικών σημείων για τις τοπικές μεθόδους. Τα αποτελέσματα εμφανίζουν μεγαλύτερη ακρίβεια, όπως ήταν άλλωστε αναμενόμενο. Η ταχύτητα σύγκλισης της μεθόδου, μπορεί να αυξηθεί αν μειώσουμε τον αριθμό επαναλήψεων στην τοπική μέθοδο με την πάροδο των γενεών. Η λογική στην παραπάνω ιδέα, είναι ότι μετά από πολλές επαναλήψεις του αλγορίθμου, προκύπτουν σημεία που βρίσκονται στην γειτονιά του ολικού ελαχίστου, οπότε μια τοπική μέθοδος μπορεί να φτάσει στο πραγματικό ολικό ελάχιστο με λίγα βήματα.

5.10 Γενικά σχόλια και συμπεράσματα

Με βάση τα όσα αναφέραμε παραπάνω, μπορούμε να συμπεράνουμε ότι οι γενετικοί αλγόριθμοι παρουσιάζουν σημαντικές διαφορές σε σύγκριση με τις ήδη υπάρχουσες μεθόδους ολικής ελαχιστοποίησης. Οι πιο σημαντικές από αυτές είναι οι παρακάτω :

- Η αναζήτηση δεν γίνεται μόνο σε ένα σημείο, αλλά παράλληλα, σε πολλά σημεία.
- Δεν είναι απαραίτητη η γνώση της παραγώγου, άρα μπορεί να εφαρμοστεί και για την ελαχιστοποίηση συναρτήσεων που δεν είναι παραγωγίσιμες. Απαιτείται μόνο η τιμή της συνάρτησης για το συγκεκριμένο σημείο.



- Οι γενετικοί αλγόριθμοι χρησιμοποιούν στοχαστικούς κανόνες μετάβασης και όχι αιτιοκρατικούς.
- Είναι σχετικά εύκολη η υλοποίησή τους.
- Επίσης σαν αποτέλεσμα, παρέχουν ένα πλήθος τοπικών ελαχίστων, κάτι που είναι χρήσιμο για αρκετά προβλήματα.

Κεφάλαιο 6 - Μέθοδοι βασισμένες σε ομαδοποίηση (clustering)

6.1 Γενικά

Η βασική ιδέα της μεθόδου του clustering είναι ότι ένα αρχικό δείγμα σημείων οργανώνεται σε ομάδες από σημεία. Τα σημεία που ανήκουν στην ίδια ομάδα ανήκουν στην ίδια **περιοχή προσέλκυσης (region of attraction)**. Με τον όρο περιοχή προσέλκυσης με κάποιο σημείο x^* ως **πυρήνα**, εννοούμε το σύνολο εκείνων των σημείων που αν αποτελέσουν αρχικά σημεία για κάποια τοπική μέθοδο ελαχιστοποίησης, θα επιστρέψουν το ίδιο τοπικό ελάχιστο. Έχουν προταθεί δύο διαφορετικοί τρόποι για την δημιουργία αυτών των ομάδων [29].

Ο πρώτος τρόπος ονομάζεται **μείωση (reduction)** και βασίζεται στην απομάκρυνση ενός ποσοστού σημείων από το αρχικό δείγμα. Συνήθως, απομακρύνονται εκείνα τα σημεία που έχουν την μεγαλύτερη τιμή συνάρτησης. Ο δεύτερος τρόπος ονομάζεται **συγκέντρωση (concentration)** και βασίζεται στην εφαρμογή ενός ή περισσότερων **steepest descent** βημάτων στο αρχικό δείγμα σημείων. Η εφαρμογή ενός **steepest descent βήματος** στα αρχικά σημεία έχει ως αποτέλεσμα την συγκέντρωσή τους σε κάποιο σημείο συσσώρευσης. Πολλές μέθοδοι όπως αυτή που υλοποιήσαμε (η οποία θα περιγραφεί αναλυτικά παρακάτω) χρησιμοποιούν συνδυασμό και των δύο αυτών τεχνικών.

Υπάρχουν διάφοροι τρόποι για την οργάνωση αυτών των σημείων σε ομάδες (clustering). Όλοι όμως στηρίζονται στην ίδια γενική ιδέα. Κάθε ομάδα χαρακτηρίζεται από ένα σημείο που ονομάζεται **πυρήνας (seed point)**. Η ομάδα αυτή εμπλουτίζεται και με άλλα σημεία, μέχρι την ικανοποίηση ενός κατάλληλου κριτηρίου τερματισμού. Ο πυρήνας μιας ομάδας μπορεί να είναι εκείνο το σημείο από το αρχικό δείγμα που δεν έχει ομαδοποιηθεί, με την χαμηλότερη τιμή συνάρτησης ή ακόμη καλύτερα κάποιο τοπικό ελάχιστο. Οι μέθοδοι *βασισμένες σε ομαδοποίηση* [29], [30] διαφέρουν κυρίως στην επιλογή του κριτηρίου να ανήκει ένα σημείο σε κάποια ομάδα.

Η κατάταξη των σημείων στις διάφορες ομάδες γίνεται με βάση δύο κριτήρια της **πυκνότητας (density)** και της **απόστασης (distance)**. Το κριτήριο της πυκνότητας τοποθετεί στην ίδια ομάδα εκείνα τα σημεία που η τιμή της συνάρτησης είναι μέσα σε κάποια προκαθορισμένα όρια. Το κριτήριο της απόστασης κατατάσσει τα σημεία με τέτοιο τρόπο, ώστε γειτονικά σημεία (η ευκλείδεια απόστασή τους δεν υπερβαίνει κάποια προκαθορισμένη τιμή) να ανήκουν στην ίδια ομάδα.

6.2 Περιγραφή μεθόδου ολικής ελαχιστοποίησης με την διαδικασία της ομαδοποίησης

Παρακάτω θα περιγράψουμε αναλυτικά τον αλγόριθμο που υλοποιήσαμε, δίνοντας σχόλια για την επιλογή κάποιων συγκεκριμένων παραμέτρων. Η διαδικασία της ομαδοποίησης έχει ως βασικό κριτήριο την απόσταση και μάλιστα την ευκλείδεια απόσταση μεταξύ των σημείων. Η μέθοδος μας στηρίζεται σε δύο μεγάλες φάσεις :

Ολική φάση (global phase) : N σημεία επιλέγονται τυχαία κι αποτελούν το αρχικό δείγμα σημείων. Η συνάρτηση αποτιμάται σε καθένα από αυτά τα σημεία.

Τοπική φάση (local phase) : Επιλέγεται κάποιο από τα σημεία του δείγματος και ξεκινάμε από αυτό το σημείο μια μέθοδο τοπικής ελαχιστοποίησης.

Περιγραφή αλγορίθμου

Βήμα 0

- Καθόρισε το πλήθος N των σημείων από τα οποία θα αποτελείται το δείγμα.
- Καθόρισε την παράμετρο γ . Η παράμετρος γ είναι το ποσοστό των μη απορριφθέντων σημείων του αρχικού δείγματος.
- Δημιουργία δύο συνόλων των X^* και $X^{(1)}$. Τα σύνολα αυτά περιέχουν αντίστοιχα, τα σημεία τοπικών ελαχίστων και τα σημεία από τα οποία ξεκίνησε μια τοπική μέθοδος ανεπιτυχώς. Το ότι ξεκίνησε μια μέθοδος ανεπιτυχώς σημαίνει ότι οδηγήθηκε σε τοπικά ελάχιστα που έχουν ήδη εντοπιστεί.

Βήμα 1

- Επέλεξε τυχαία N σημεία και πρόσθεσέ τα στο δείγμα σημείων.

Βήμα 2

- Κατασκεύασε το τροποποιημένο δείγμα σημείων. Από τα αρχικά σημεία θα κρατηθούν γN (αυτά που έχουν την χαμηλότερη τιμή συνάρτησης). Εφάρμοσε ένα βήμα *steepest descent* για αυτά τα σημεία.

Βήμα 3

- Εφαρμογή μιας διαδικασίας ομαδοποίησης στο τροποποιημένο δείγμα σημείων. Παρακάτω θα περιγραφεί εκτενώς πώς τα σημεία κατανέμονται στις διάφορες ομάδες. Αρχικά επιλέγονται ως σημεία πυρήνες εκείνα τα σημεία που ανήκουν στο σύνολο X^* κι εν συνεχεία σημεία από το σύνολο $X^{(1)}$. Αν όλα τα σημεία του δείγματος σημείων έχουν κατανεμηθεί σε κάποια ομάδα τότε πηγαίνε στο βήμα 5.

Βήμα 4

- Επέλεξε από εκείνα τα σημεία που δεν έχουν μπει σε κάποια ομάδα, εκείνο που έχει την χαμηλότερη τιμή της συνάρτησης. Έστω ότι αυτό είναι το σημείο $x^{(1)}$. Εφάρμοσε μια μέθοδο τοπικής ελαχιστοποίησης για την εύρεση ενός **τοπικού ελαχίστου** x^* .
- Αν το x^* είναι ένα **τοπικό ελάχιστο** που δεν έχει ξαναεντοπιστεί (δεν ανήκει δηλαδή στο σύνολο X^*) το προσθέτουμε στο σύνολο X^* . Το νέο τοπικό ελάχιστο θα αποτελέσει τον πυρήνα για μια νέα ομάδα, οπότε γίνεται η εφαρμογή της διαδικασίας της ομαδοποίησης και κάποια νέα σημεία εισέρχονται στη νέα ομάδα. Στη συνέχεια **πηγαίνουμε στο βήμα 1** κι επαναλαμβάνεται η διαδικασία από εκεί και κάτω.
- Αν το x^* είναι ένα **τοπικό ελάχιστο** που έχει ήδη βρεθεί, τότε επιλέγουμε το **αρχικό σημείο** $x^{(1)}$ που οδήγησε στο ίδιο τοπικό ελάχιστο σαν πυρήνα μιας νέας ομάδας και το προσθέτουμε στο σύνολο $X^{(1)}$. Πηγαίνουμε στο βήμα 4 εωσότου όλα τα σημεία να έχουν τοποθετηθεί σε κάποια ομάδα.

Βήμα 5

- Επιλέγουμε από τα τοπικά ελάχιστα που έχουν βρεθεί, εκείνο που έχει την **μικρότερη τιμή** και το επιστρέφουμε σαν το πιθανό ολικό ελάχιστο.

Παρατηρήσεις

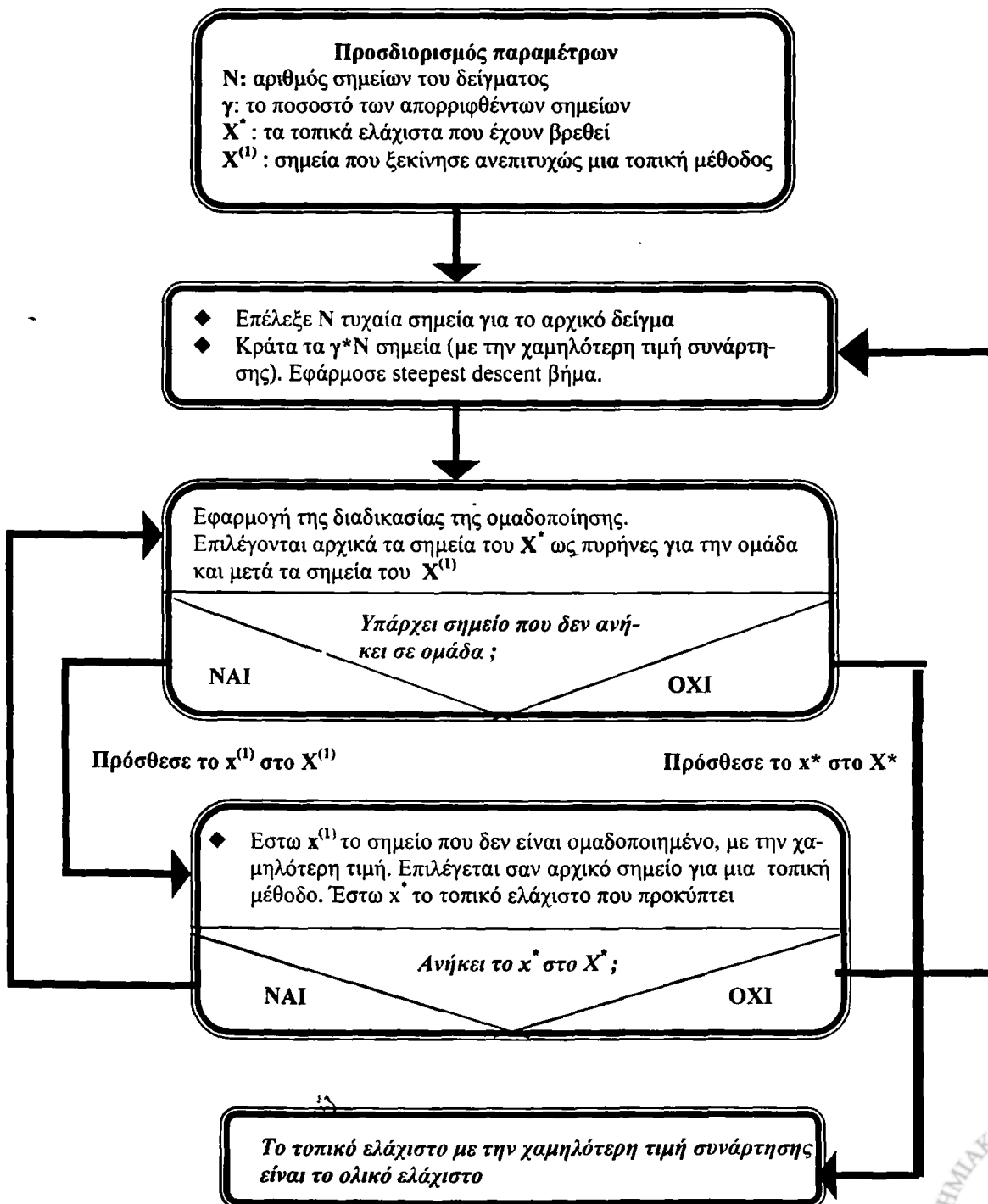
Παράμετροι εισόδου :

- Ο χρήστης καθορίζει το **πλήθος των σημείων που θα αποτελούν το δείγμα σημείων**. Όσο μεγαλύτερο είναι το πλήθος των σημείων του αρχικού δείγματος, τόσο μεγαλύτερη είναι η πιθανότητα να προσεγγίσουμε το σωστό ολικό ελάχιστο. Ενδεικτική τιμή για το πλήθος των σημείων είναι $25 * \text{διάσταση συνάρτησης}$.
- Επίσης ο χρήστης καθορίζει το **ποσοστό των απορριφθέντων σημείων** (απορρίπτονται εκείνα τα σημεία που έχουν την μεγαλύτερη τιμή συνάρτησης). Ενδεικτική τιμή $\gamma=0.3$.
- Ο χρήστης μπορεί να **εισάγει ένα άνω όριο για το πλήθος των τοπικών ελαχίστων**, που θέλει να εντοπιστούν από τον παραπάνω αλγόριθμο. Όταν ο αλγόριθμος ξεπεράσει αυτό το ανώτατο όριο του πλήθους των τοπικών ελαχίστων, τερματίζει. Με αυτό τον τρόπο αυξάνεται η ταχύτητα σύγκλισης του αλγορίθμου, με σημαντικές επιπτώσεις όμως στην αξιοπιστία και στην αποτελεσματικότητά του.
- Επίσης ο χρήστης εισάγει **μια τιμή κατωφλίου για την κρίσιμη απόσταση** που θα πρέπει να πληρεί ένα σημείο για να εισαχθεί σε κάποια ομάδα. Όσο μεγαλύτερη είναι αυτή η τιμή, τόσο πιο ανεκτικοί γινόμαστε στην εισαγωγή σημείων στις διάφορες ομάδες, με αποτέλεσμα ο αλγόριθμος να τερματίσει (αφού κριτήριο τερματισμού του αλγορίθμου, είναι η μη ύπαρξη σημείου, που να μην ανήκει σε κάποια ομάδα). Μικρή τιμή κατωφλίου έχει ως αποτέλεσμα μεγαλύτερη αξιοπιστία στο αποτέλεσμα που επιστρέφει η μέθοδος, επηρεάζοντας όμως αισθητά την ταχύτητά της.

Παράμετροι εξόδου :

Επιστρέφεται η τιμή του ολικού ελαχίστου και η θέση στην οποία παρατηρήθηκε. Επίσης επιστρέφονται και τα διάφορα τοπικά ελάχιστα που βρέθηκαν κατά την εξέλιξη της διαδικασίας. Αυτό είναι χρήσιμο για κάποιες εφαρμογές στις οποίες κρίνεται σκόπιμο να επιστρέφονται και τα διάφορα τοπικά ελάχιστα.

6.3 Σχηματική αναπαράσταση του αλγορίθμου



6.4 Πώς γίνεται η ομαδοποίηση

Εστω ότι το διάνυσμα των μεταβλητών είναι ένα διάνυσμα διάστασης N , δηλαδή $x = (x^{(1)}, x^{(2)}, \dots, x^{(N)})$

Θεωρούμε ως πυρήνα του cluster ένα σημείο τοπικού ελαχίστου x^* , που είναι επίσης ένα διάνυσμα διάστασης N . Ορίζουμε την απόσταση ανάμεσα σε δύο σημεία x_1 και x_2 , σύμφωνα με τον παρακάτω τύπο (Ευκλείδεια απόσταση).

$$D = [(x_1^{(1)} - x_2^{(1)})^2 + (x_1^{(2)} - x_2^{(2)})^2 + \dots + (x_1^{(n)} - x_2^{(n)})^2]^{1/2}$$

Αρχικά τοποθετούμε στην ομάδα εκείνο το σημείο που απέχει την μικρότερη απόσταση από τον πυρήνα της (seed point). Για την εισαγωγή των επόμενων σημείων στην ομάδα ακολουθούμε την παρακάτω διαδικασία :

Βρίσκουμε ένα σημείο που δεν έχει τοποθετηθεί ακόμη σε κάποια ομάδα και ικανοποιεί την συνθήκη :

$$D(x, C) = \min \{d(x, x^*)\} \quad x^* \in C$$

Επιλέγουμε δηλαδή το σημείο εκείνο που απέχει την μικρότερη απόσταση από την ομάδα σε σχέση με τα υπόλοιπα σημεία. Απόσταση ενός σημείου από την ομάδα, θεωρείται η ελάχιστη απόσταση από κάποιο σημείο που ανήκει ήδη στην ομάδα.

Δεν είναι όμως ακόμη βέβαιο ότι το σημείο αυτό θα εισαχθεί στην συγκεκριμένη ομάδα. Θα πρέπει να ελέγξουμε, αν η απόσταση αυτή είναι μικρότερη από μια τιμή κατωφλίου που δίνεται ως παράμετρος εισόδου.

Επίσης για την εισαγωγή ενός συγκεκριμένου σημείου σε κάποια ομάδα, ελέγχεται και το κριτήριο της παραγώγου (gradient criterion).

Κριτήριο παραγώγου

- Αν το σημείο πυρήνας (seed point) της ομάδας, είναι κάποιο τοπικό ελάχιστο, έστω x^* , τότε για να ανήκει το σημείο x στην ίδια ομάδα, θα πρέπει να ισχύει η συνθήκη

$$(x^* - x) \nabla f(x) < 0$$

- Αν το σημείο πυρήνας της ομάδας, είναι το αρχικό σημείο που οδήγησε σε τοπικό ελάχιστο που έχει ήδη βρεθεί, έστω $x^{(1)}$, τότε για να ανήκει το νέο σημείο x στην ίδια ομάδα θα πρέπει να ισχύει

$$(x^{(1)} - x) (\nabla f(x^{(1)}) - \nabla f(x)) > 0$$

Ανάλυση του κριτηρίου της παραγώγου

Γύρω από την περιοχή ενδιαφέροντος του τοπικού ελαχίστου x^* , ισχύει η παρακάτω σχέση για κάποιο x .

$f(x) = f(x^*) + \frac{1}{2} (x - x^*)^T H^* (x - x^*)$, όπου H^* ο Εσσιανός πίνακας στο x^* .

Θέτουμε $g(x) = H^* (x - x^*)$

Αν λοιπόν τα σημεία x και x' ανήκουν στην περιοχή του τοπικού ελαχίστου x^* ισχύουν οι παρακάτω σχέσεις για καθένα από αυτά τα σημεία :

$$H^* (x^* - x) = -g(x) \quad (1)$$

$$H^* (x^* - x') = -g(x') \quad (2)$$

Από τις (1) και (2) προκύπτει η σχέση

$$H^* (x' - x) = g(x') - g(x)$$

$$\text{Και : } (x' - x)^T H^* (x' - x) = (x' - x)^T (g(x') - g(x))$$

Επειδή ο πίνακας H^* πρέπει να είναι θετικά ορισμένος, θα ισχύει :

$$(x' - x)^T (g(x') - g(x)) > 0$$

Αν η σχέση αυτή δεν ικανοποιείται, τότε τα σημεία x και x' δεν ανήκουν στην περιοχή προσέλκυσης (region of attraction) του ίδιου ελαχίστου.

6.5 Πειραματικά αποτελέσματα

	Διάσταση	Ελ. τιμή	Ολ. Ελάχιστο	Αποτιμήσεις
Sixhump	2	-1.0316284	-1.0316285	1054
Goldstein	2	3	3	1245
Rastrigin	2	-2	-2	2277
Griewank2	2	0	0	5267
Neural	15	0.000000008	0	15673
Griewank10	10	0	0	52396

Σχόλια - συμπεράσματα

Η μέθοδος της *ομαδοποίησης*, φαίνεται ότι υπερέχει έναντι των υπολοίπων μεθόδων. Πετυχαίνει ακριβή αποτελέσματα με μικρό αριθμό κλήσεων της συνάρτησης ακόμη και σε δύσκολα προβλήματα ολικής ελαχιστοποίησης. Ειδικά σε περιπτώσεις με λίγα τοπικά ελάχιστα, η ταχύτητα σύγκλισης του αλγορίθμου αυτού είναι πολύ μεγάλη. Χρησιμοποιήθηκε η τοπική μέθοδος *bfgs* για όλα τα παραδείγματα, εκτός από την συνάρτηση σφάλματος του νευρωνικού δικτύου, όπου χρησιμοποιήθηκε η μέθοδος *Levenberg-Marquadt*. Βέβαια η ύπαρξη κάποιων παραμέτρων, που είδαμε νωρίτερα όπως το πλήθος των σημείων του αρχικού δείγματος και η επιλογή της τιμής κατωφλίου για την απόσπαση των σημείων που ανήκουν στην ίδια ομάδα, έχουν καθοριστική σημασία για την επιτυχία της μεθόδου.

Συγκεντρωτικά συγκριτικά αποτελέσματα στοχαστικών μεθόδων

Οι πίνακες που ακολουθούν περιγράφουν τα συγκεντρωτικά αποτελέσματα των στοχαστικών μεθόδων που υλοποιήθηκαν. Έτσι μπορούμε να έχουμε μια συνολική εικόνα, από την οποία προκύπτουν χρήσιμα συμπεράσματα για την αξιοπιστία και την απόδοση της κάθε μεθόδου. Η πρώτη στήλη περιέχει το όνομα της συνάρτησης δοκιμής (δες παράρτημα) και την πραγματική τιμή του ολικού ελαχίστου. Οι υπόλοιπες στήλες περιγράφουν το πλήθος των αποτιμήσεων της συνάρτησης και το επιστρεφόμενο αποτέλεσμα για κάθε μέθοδο. Ο πίνακας 1 περιέχει αποτελέσματα των μεθόδων της ελεγχόμενης τυχαίας αναζήτησης (απλή και βελτιωμένη έκδοση του αλγορίθμου του Price) καθώς επίσης και των δύο εκδόσεων του αλγορίθμου simulated annealing. Ο πίνακας 2 περιγράφει τα αποτελέσματα της μεθόδου SALT (Simulated Annealing and Local Tuning), του απλού γενετικού αλγορίθμου, της υβριδικής μεθόδου (γενετικοί με τοπική ελαχιστοποίηση) και της μεθόδου ομαδοποίησης (clustering).

Πίνακας 1

Συνάρτηση Τιμή ολ. Ελαχίστου	Απλός Price Αποτιμήσεις Ελαχ. τιμή	Βελτίωση Price Αποτιμήσεις Ελαχ. Τιμή	SA1 Αποτιμήσεις Ελαχ. τιμή	SA2 Αποτιμήσεις Ελαχ. τιμή
Sixhump -1.0316285	2203 -1.0316284	1878 -1.0316284	5329 -1.0316284	6083 -1.0316284
Goldstein 3	2856 3.0000026	2345 3	5123 3.000423	6123 3.000672
Rastrigin -2	3521 -1.9999	2209 -1.9999	4804 -1.9888888	5503 -1.9999998
Griewank2 0	3224 0.00000008	1932 0.00000001	12404 0.0000001	14671 0.00000014
Neural 0	150000 0.00043	150000 0.000061345	150000 0.0000198	150000 0.0000532
Griewank10 0	86123 0.000123	81778 0.0000083	150000 0.0002238	150000 0.0002145

Πίνακας 2

Συνάρτηση Τιμή ολ. Ελαχίστου	SALT Αποτιμήσεις Ελαχ. τιμή	Γενετικός Αποτιμήσεις Ελαχ. Τιμή	Υβριδική Αποτιμήσεις Ελαχ. τιμή	Ομαδοποίηση Αποτιμήσεις Ελαχ. Τιμή
Sixhump -1.0316285	2429 -1.0316284	29830 -1.0316283	1690 -1.0316284	1054 -1.0316284
Goldstein 3	2678 3	25456 3.0003456	2365 3	1245 3
Rastrigin -2	5901 -2	41813 -1.9999999	5206 -2	2277 -2
Griewank2 0	6103 0	33550 0	7828 0	5267 0
Neural 0	127855 0.00000076	150000 0.000015	27892 0.00000004	15673 0.000000008
Griewank10 0	103751 0.00000013	81933 0.000067	10197 0.00000017	52396 0

Σχόλια – συμπεράσματα

Από τα παραπάνω πειραματικά αποτελέσματα φαίνεται σαφώς η ανωτερότητα των υβριδικών μεθόδων, των μεθόδων δηλαδή που χρησιμοποιούν και τοπική ελαχιστοποίηση. Έτσι λοιπόν οι μέθοδοι SALT, η υβριδική μέθοδος των γενετικών αλγορίθμων και η ομαδοποίηση, επιστρέφουν αποτελέσματα μεγάλης ακρίβειας, πολύ κοντά στην τιμή του πραγματικού ολικού ελαχίστου. Επίσης το πλήθος των αποτιμήσεων της συνάρτησης για τις υβριδικές μεθόδους είναι πολύ μικρότερο, συγκριτικά με τις υπόλοιπες, κυρίως στην περίπτωση συναρτήσεων με πολλές παραμέτρους (Griewank10, Neural). Από τις μη υβριδικές μεθόδους καλύτερα αποτελέσματα δίνει η πρώτη έκδοση του simulated annealing SA1. Οι μη υβριδικές μέθοδοι δίνουν επίσης ικανοποιητικά αποτελέσματα κι έχουν το πλεονέκτημα ότι δεν απαιτούν την ύπαρξη της παραγώγου. Τέλος, η μέθοδος που υπερέχει έναντι των υπολοίπων, όπως προκύπτει από τα πειραματικά αποτελέσματα, είναι η μέθοδος της ομαδοποίησης.

Κεφάλαιο 7 - Γενικά στοιχεία για τις αιτιοκρατικές (deterministic) μεθόδους

7.1 Εισαγωγή

Οι μέθοδοι που ανήκουν σε αυτή την κατηγορία προσπαθούν να λύσουν το πρόβλημα της ολικής ελαχιστοποίησης *αιτιοκρατικά*. Το μεγάλο τους πλεονέκτημα είναι ότι εγγυώνται την εύρεση σωστής λύσης για το ολικό ελάχιστο. Η βασική φιλοσοφία αυτών των μεθόδων είναι η εξαντλητική αναζήτηση σε όλο το πεδίο ορισμού των παραμέτρων, και ο υπολογισμός του διαστήματος τιμών της συνάρτησης μέσω κατάλληλων συναρτήσεων. Αυτές οι μέθοδοι επιτυγχάνουν μόνο όταν υπολογίζουν τα διαστήματα τιμών της συνάρτησης (για συγκεκριμένο διάστημα ορισμού των παραμέτρων) με μεγάλη ακρίβεια, λαμβάνοντας υπόψη και τα διάφορα *roundoff errors* (λάθη στρογγύλευσης). Στον τομέα αυτό η *ανάλυση διαστημάτων* (*interval analysis*) κρίνεται πιο αποδοτική. Οι μέθοδοι αυτές στηρίζονται κυρίως στις τεχνικές της διαμέρισης και της αναγνώρισης ορίων των τιμών της συνάρτησης. Είναι γνωστές με τον όρο μέθοδοι *διακλάδωσης και φράγματος* (*branch and bound*). Με την διαμέριση χωρίζεται το αρχικό διάστημα ορισμού των παραμέτρων σε μικρότερα υποδιαστήματα (*branch*) και στη συνέχεια μέσω κατάλληλων συναρτήσεων (πχ με χρήση συναρτήσεων της ανάλυσης διαστημάτων) υπολογίζονται τα όρια του διαστήματος τιμών της συνάρτησης, για καθένα από αυτά τα υποδιαστήματα. Οι πιο δημοφιλείς εκπρόσωποι αυτής της κατηγορίας είναι οι μέθοδοι διαστήματος (*interval methods*). Το μεγάλο μειονέκτημα αυτών των μεθόδων είναι οι αυξημένες απαιτήσεις τους σε υπολογιστικούς πόρους, κυρίως στις περιπτώσεις συναρτήσεων με πολλές παραμέτρους και πεδίο ορισμού μεγάλου εύρους. Ο υπολογιστικός χρόνος αυτών των μεθόδων, όταν τα διαστήματα ορισμού των παραμέτρων είναι μεγάλου εύρους, είναι πολλές φορές απαγορευτικός. Σε αυτά τα προβλήματα κρίνονται πιο κατάλληλες οι στοχαστικές μέθοδοι.

Στο κεφάλαιο αυτό θα δούμε τις γενικές αρχές των μεθόδων *φράγματος και διακλάδωσης* και στη συνέχεια θα δώσουμε τα απαραίτητα εισαγωγικά στοιχεία και τις αρχές που διέπουν την ανάλυση διαστημάτων. Στο σημείο αυτό θα πρέπει να τονίσουμε ότι παραθέτουμε τα τελείως απαραίτητα στοιχεία της ανάλυσης διαστημάτων, που θα βοηθήσουν τον αναγνώστη στην κατανόηση των μεθόδων διαστήματος για το πρόβλημα της ολικής ελαχιστοποίησης.

7.2 Μέθοδοι εύρεσης των ορίων των τιμών μιας συνάρτησης

Οι μέθοδοι που ανήκουν σε αυτή την κατηγορία [23] επιστρέφουν διαστήματα στα οποία ανήκουν οι τιμές που μπορεί να πάρει η συνάρτηση, για δεδομένο πεδίο ορισμού των παραμέτρων της. Σε προβλήματα ολικής ελαχιστοποίησης γνωρίζουμε επομένως, ότι η τιμή του ολικού ελαχίστου δεν μπορεί να είναι μικρότερη από το αριστερό άκρο του διαστήματος τιμών της συνάρτησης, που επέστρεψε η μέθοδος. Βέβαια, όσο μικρότερο είναι το πλάτος του επιστρεφόμενου διαστήματος, τόσο καλύτερη είναι η προσέγγιση της πραγματικής τιμής του ολικού ελαχίστου.

7.2.1 Τεχνικές για τον υπολογισμό ορίων των διαστημάτων τιμών της συνάρτησης

- Η προσέγγιση του Lipschitz

Η τεχνική αυτή μας επιστρέφει ένα κάτω φράγμα των τιμών της συνάρτησης για δεδομένο πεδίο ορισμού των παραμέτρων της. Βασίζεται στην υπόθεση ότι υπάρχει μια σταθερά του Lipschitz L , τέτοια ώστε

$$|f(x_1) - f(x_2)| \leq L \|x_1 - x_2\|$$

Αν η τιμή της συνάρτησης είναι γνωστή για κάποιο σημείο, έστω x_1 , τότε το κάτω άκρο του διαστήματος τιμών της συνάρτησης για όλα τα x , που είναι ανάμεσα στα x_1 και x_2 , θα δίνεται από τον τύπο $f(x_1) - L \|x - x_1\|$.

- Η γραμμική προσέγγιση κάτω ορίου (The linear lower bound approach)

Η προσέγγιση αυτή καθορίζει γραμμικά κάτω όρια για μια συνάρτηση f , διασπώντας την αρχική συνάρτηση f σε πιο απλές συναρτήσεις, μονότονες και κυρτές. Στη συνέχεια, προσπαθούμε να συνδυάσουμε τα κάτω όρια των απλών συναρτήσεων, ώστε να πάρουμε τελικά το κάτω όριο της αρχικής μας συνάρτησης. Αυτή η μέθοδος προϋποθέτει βέβαια, ότι η συνάρτηση μπορεί να υποδιαιρεθεί σε άλλες πιο απλές συναρτήσεις, με τις ιδιότητες της μονοτονικότητας και της κυρτότητας. Όμως έχει αποδειχτεί ότι κάτι τέτοιο μπορεί να γίνει μόνο για απλά προβλήματα ολικής ελαχιστοποίησης, μιας μόνο παραμέτρου.

- Η προσέγγιση διαστήματος

Η προσέγγιση αυτή ακολουθεί τις αρχές της αριθμητικής διαστήματος που θα αναλύσουμε εκτενώς στη συνέχεια. Χρησιμοποιώντας λοιπόν, τις βασικές πράξεις της αριθμητικής διαστήματος, μπορούμε να προσεγγίσουμε με μεγάλη ακρίβεια τα διαστήματα τιμών της συνάρτησης. Για κάθε πραγματική συνάρτηση μπορεί να γραφτεί η επέκτασή διαστημάτων της (interval extension). Μια τέτοια συνάρτηση παίρνει σαν είσοδο ένα διάστημα κι επιστρέφει το διάστημα τιμών της συνάρτησης.

7.2.2 Ένας απλός αλγόριθμος εύρεσης των ορίων μιας συνάρτησης

Έστω μια συνάρτηση F_L η οποία μας επιστρέφει το κάτω όριο του διαστήματος τιμών της συνάρτησης, για το πεδίο ορισμού των παραμέτρων της. Αν θεωρήσουμε λοιπόν δεδομένη την F_L , μπορούμε με εξαντλητική αναζήτηση σε υποδιαστήματα του πεδίου ορισμού της συνάρτησης, να φτάσουμε στην πραγματική τιμή του ολικού ελαχίστου. Ένας απλός αλγόριθμος για την εύρεση της ελάχιστης τιμής, θα έχει τα παρακάτω βήματα. Αρχικά χωρίζεται το αρχικό σύνολο X του πεδίου ορισμού των παραμέτρων, σε μικρότερα υποσύνολα S_i^x . Αν είναι διαθέσιμη η συνάρτηση F_L , μπορεί εύκολα να υπολογιστεί το κάτω όριο του διαστήματος τιμών της αρχικής συνάρτησης (με εφαρμογή της $F_L(S_i^x)$), για καθένα από τα υποσύνολα S_i^x . Επίσης κρατείται και η μέγιστη έως τώρα τιμή της συνάρτησης, έστω U_{F^*} . Κάθε υποσύνολο S_i^x για το οποίο ισχύει $F_L(S_i^x) > U_{F^*}$ μπορεί να απορριφθεί, επειδή είναι προφανές ότι αποκλείεται να περιέχει το ολικό ελάχιστο, αφού το κάτω όριο του διαστήματος τιμών της συνάρτησης στο υποσύνολο S_i^x , είναι μεγαλύτερο από την τρέχουσα μέγιστη τιμή της. Οι παραπάνω διαδικασίες, δηλαδή ο διαμερισμός του αρχικού συνόλου (partitioning), η εύρεση των ορίων του διαστήματος τιμών της συνάρτησης (bounding) και η πιθανή απόρριψη κάποιων υποσυνόλων (rejection) επαναλαμβάνονται για συνεχώς μικρότερα υποσύνολα, μέχρι την ικανοποίηση κάποιου κριτηρίου τερματισμού. Η ένωση των συνόλων που δεν έχουν απορριφθεί κατά την διεξαγωγή της

παραπάνω διαδικασίας, αποτελούν πιθανές λύσεις ολικών ελαχίστων για το συγκεκριμένο πρόβλημα.

Η επιτυχία του παραπάνω αλγορίθμου εύρεσης ορίων εξαρτάται σε μεγάλο βαθμό από την συνάρτηση εύρεσης κάτω ορίου των τιμών της συνάρτησης (the lower bounding function). Όσο πιο κοντά βρίσκεται το κάτω όριο που επιστρέφει η παραπάνω συνάρτηση στην πραγματική τιμή του ολικού ελαχίστου, τόσο πιο ακριβές θα είναι το αποτέλεσμα. Αν για παράδειγμα το ολικό ελάχιστο μιας συνάρτησης είναι μηδέν, θα επιθυμούσαμε η επιστρεφόμενη τιμή της συνάρτησης του κάτω ορίου να ήταν περίπου μηδέν. Ωστόσο, οι συναρτήσεις αυτές χαρακτηρίζονται κάποιες φορές από μεγάλη ανακρίβεια. Το επιθυμητό είναι να διέπονται από την ιδιότητα της *ασυμπτωτικής ακρίβειας*, δηλαδή το κάτω όριο των διαστήματος τιμών της συνάρτησης να πλησιάζει το ολικό ελάχιστο, όσο το πλάτος του αρχικού διαστήματος ορισμού των παραμέτρων μικραίνει.

7.2.3 Ένας γενικός αλγόριθμος εύρεσης ορίων (bounding algorithm)

Όλες οι μέθοδοι εύρεσης ορίων (*bounding methods*), όπως είναι οι αλγόριθμοι φράγματος και διακλάδωσης, οι μέθοδοι της επικάλυψης (*covering*), μέθοδοι του *Lipschitz* και οι μέθοδοι διαστήματος, υλοποιούν τον ακόλουθο γενικό αλγόριθμο.

- Διαμέριση του αρχικού διαστήματος ορισμού των παραμέτρων της συνάρτησης, σε μικρότερα υποδιαστήματα (*partition*).
- Εύρεση των ορίων του διαστήματος τιμών της συνάρτησης (και του διαστήματος τιμών της παραγώγου, για κάποιες περιπτώσεις) στα διάφορα υποδιαστήματα (*bounding*).
- Απόρριψη εκείνων των υποδιαστημάτων, τα οποία είναι προφανές ότι δεν μπορούν να περιέχουν την τιμή του ολικού ελαχίστου (*rejection phase*).

Η ένωση των εναπομείναντων διαστημάτων θα περιέχει το σημείο του πραγματικού ολικού ελαχίστου. Ο παραπάνω αλγόριθμος ακολουθεί την αρχή της εξαντλητικής αναζήτησης (*exhaustive search*). Εξαντλητικά λοιπόν διερευνά όλο το σύνολο ορισμού της συνάρτησης κι έτσι το αποτέλεσμα που επιστρέφεται είναι αιτιοκρατικό κι όχι στοχαστικό, αφού δεν εξετάζεται ένα υποσύνολο του συνόλου ορισμού των παραμέτρων.

7.2.3.1 Η φάση της διαμέρισης του αρχικού συνόλου

Ο τρόπος με τον οποίο θα γίνει η διαμέριση του αρχικού συνόλου σε μικρότερες υποπεριοχές, είναι άμεση συνάρτηση του τύπου της συνάρτησης και της γεωμετρίας του συνόλου ορισμού των παραμέτρων της. Θα πρέπει επίσης η διαμέριση να γίνεται με τέτοιο τρόπο, ώστε να είναι εύκολος ο υπολογισμός του διαστήματος τιμών της συνάρτησης για το υποσύνολο που προκύπτει. Για παράδειγμα η *αριθμητική διαστημάτων* χρησιμοποιεί τα *υπερορθογώνια* (*hyperrectangles*), για να υπολογίσει τα όρια του διαστήματος των τιμών της συνάρτησης. Η διαδικασία λοιπόν της διαμέρισης είναι η διαίρεση ενός μεγάλου υπερρθογωνίου σε μικρότερα υπερρθογώνια. Το επιλεγόμενο γεωμετρικό σχήμα διαμέρισης, θα πρέπει να καλύπτει όλο το σύνολο ορισμού των παραμέτρων, ώστε να διασφαλιστεί ότι έγινε έλεγχος όλης της περιοχής (εξαντλητική αναζήτηση), όπου μπορούν να εμφανιστούν ολικά ελάχιστα. Αν για παράδειγμα το επιλεγόμενο σχήμα διαμέρισης απορρίπτει κάποιες περιοχές, τότε μπορεί να απομακρυνθούν και σημαντικές περιοχές, που περιέχουν το πραγματικό ολικό ελάχιστο. Συνεπώς η επιλογή του υπερκύκλου

(*hypercircle*) ως σχήμα διαμέρισης δεν είναι αποδοτική, αφού μπορεί να απορρίψει σημαντικές υποπεριοχές.

Σύμφωνα λοιπόν με τα παραπάνω, η επιλογή του *υπερορθογωνίου* κρίνεται ως η πιο κατάλληλη για τους εξής λόγους :

- Η διαμέριση γίνεται πολύ εύκολα.
- Καλύπτει όλη την περιοχή ορισμού των παραμέτρων της συνάρτησης, χωρίς επικαλύψεις.
- Οι υπολογισμοί του διαστήματος τιμών της συνάρτησης είναι πιο εύκολοι.

Αν η αρχική περιοχή των παραμέτρων δεν είναι από μόνη της ένα υπερορθογώνιο, τότε μπορούμε να την κλείσουμε σε ένα υπερορθογώνιο και στη συνέχεια να απορρίψουμε εκείνες τις περιοχές, οι οποίες είτε δεν υπακούουν στους περιορισμούς του προβλήματος, είτε δεν είναι δυνατό να περιέχουν ολικά ελάχιστα.

7.2.3.2 Επιλογή των συναρτήσεων, που καθορίζουν τα όρια του διαστήματος τιμών της συνάρτησης (*bounding functions*)

Η επιτυχία των αλγορίθμων φράγματος κρίνεται κατά μεγάλο ποσοστό, από την ορθή επιλογή των συναρτήσεων, που καθορίζουν τα όρια του διαστήματος τιμών της συνάρτησης. Θα πρέπει λοιπόν οι συναρτήσεις αυτές, να διέπονται από τα ακόλουθα χαρακτηριστικά γνωρίσματα :

- Να μπορούν να δίνουν όρια αξιόπιστα, κι ασυμπτωτικά ακριβή.
- Να υλοποιούνται εύκολα.
- Να μπορούν να εφαρμόζονται για οποιαδήποτε συνάρτηση.
- Να έχουν τη δυνατότητα γενίκευσης.
- Εύκολο να προγραμματιστούν και να χρησιμοποιηθούν, ακόμη κι από έναν χρήστη που δεν είναι πολύ εξειδικευμένος (φιλικές προς τον χρήστη).

7.2.3.3 Λάθος στρογγύλευσης (Roundoff error)

Οι παραπάνω συναρτήσεις θα πρέπει να μας επιστρέφουν όρια για τις τιμές της συνάρτησης, που να ικανοποιούν τους περιορισμούς ακρίβειας της μηχανής. Κάθε τέτοια συνάρτηση θα πρέπει να προνοεί για την αντιμετώπιση *roundoff errors* (λάθη στρογγύλευσης).

Παράδειγμα λάθους στρογγύλευσης (*roundoff error*)

Η ύπαρξη αριθμών διπλής ακρίβειας στις σύγχρονες γλώσσες προγραμματισμού, δεν έχει λύσει το πρόβλημα εμφάνισης λαθών στρογγύλευσης στους διάφορους επιστημονικούς υπολογισμούς. Ας δούμε το παρακάτω παράδειγμα του **M.Rump** [23], που φανερώνει την σημασία των *roundoff* λαθών. Έστω λοιπόν η συνάρτηση

$$f(x) = 333.75 y^2 + x^2 (x^2 y^2 - y^6 - 121 y^4 - 2) + 5.5 y^8 + (x/2) * y$$

Η τιμή της συνάρτησης αυτής υπολογίστηκε για τις εξής τιμές των x και y :
 $x = 77617$ και $y = 33096$.



Η τιμή της συνάρτησης υπολογίστηκε για αριθμούς απλής ακρίβειας, διπλής κι εκτεταμένης, με ακρίβεια 6, 14, και 35 δεκαδικών ψηφίων αντίστοιχα, και πήραμε τα ακόλουθα αποτελέσματα

- Απλής ακρίβειας $f = 6.33825 \times 10^{29}$
- Διπλής ακρίβειας $f = 1.1726039400532$
- Εκτεταμένης ακρίβειας $f = 1.1726039400531786318588349045201838$

Παρατηρώντας τα παραπάνω αποτελέσματα, μπορεί κανείς να υποθέσει ότι οι υπολογισμοί με την χρήση αριθμών απλής ακρίβειας, είναι λάθος, ενώ τα δύο επόμενα αποτελέσματα είναι σωστά, αφού συμφωνούν στα 13 πρώτα δεκαδικά ψηφία. Ωστόσο, όλοι οι υπολογισμοί είναι λάθος, ακόμη και το επιστρεφόμενο πρόσημο!!!

Το πραγματικό αποτέλεσμα, όπως αυτό επιστρέφεται με χρήση της *αριθμητικής διαστημάτων* (θα αναλυθεί εκτενώς στη συνέχεια της αναφοράς μας) είναι το παρακάτω διάστημα

**[-0.827396059946821368141165095479816292005,
-0.827396059946821368141165095479816291986]**

Βλέπουμε λοιπόν από το προηγούμενο παράδειγμα την κρισιμότητα των *λαθών στρογγύλευσης*, αφού μπορούν να επηρεάσουν σε πολύ μεγάλο βαθμό την ακρίβεια των αποτελεσμάτων στους επιστημονικούς υπολογισμούς, με αριθμούς κινητής υποδιαστολής (floating point numbers).

Η αριθμητική διαστημάτων αντιμετωπίζει το παραπάνω πρόβλημα και κρίνεται κατάλληλη σε μαθηματικά προβλήματα μεγάλης ακρίβειας, όπως είναι και το πρόβλημα της ολικής ελαχιστοποίησης.

7.3 Γενικά στοιχεία της αριθμητικής διαστημάτων [24]

Ας θεωρήσουμε ένα κλειστό διάστημα X πραγματικών αριθμών με άνω και κάτω άκρα a και b αντίστοιχα, δηλαδή $X = [a,b]$. Ως *αριθμό-διάστημα* (*interval number*) θεωρούμε ένα τέτοιο διάστημα X . Το διάστημα αυτό αποτελείται από το σύνολο $\{x: a \leq x \leq b\}$ των πραγματικών ανάμεσα στο a και στο b συμπεριλαμβανομένων και των άκρων a και b . Οι όροι *διάστημα* (*interval*) κι *αριθμός-διάστημα* (*interval number*), θα χρησιμοποιούνται από εδώ και στο εξής ισοδύναμα.

Ενας πραγματικός αριθμός x είναι ισοδύναμος με ένα διάστημα $[x,x]$. Αυτή είναι και η πιο απλή περίπτωση *αριθμών-διαστημάτων*. Ένα τέτοιο διάστημα θα ορίζεται από εδώ και στο εξής ως «εκφυλισμένο» (*degenerate*). Έτσι λοιπόν ο πραγματικός αριθμός 2 είναι ισοδύναμος με τον *αριθμό-διάστημα* $[2,2]$. Επίσης, μπορούμε να γράφουμε απλά x και να εννοούμε το διάστημα $[x,x]$.

Τα άκρα του διαστήματος ενός διαστήματος $[a,b]$ είναι πολύ πιθανό, να μην μπορούν να αναπαρασταθούν από αριθμούς της μηχανής. Τότε, *στρογγυλοποιούμε* τον a (αριστερό άκρο του διαστήματος) στον μεγαλύτερο αριθμό της μηχανής που είναι μικρότερος ή ίσος από τον a , και *στρογγυλοποιούμε* τον b στον μικρότερο αριθμό μηχανής που είναι μεγαλύτερος ή ίσος από τον b . Η διαδικασία αυτή ονομάζεται *εξωτερική στρογγύλευση* (*outward rounding*) και χρησιμοποιείται πολύ στις διάφορες πράξεις της *αριθμητικής διαστημάτων*, όπως θα δούμε και στη συνέχεια.

7.3.1 Διάφοροι ορισμοί

- Ένα διάστημα $X = [a,b]$ είναι θετικό, αν $a \geq 0$, αυστηρά θετικό αν $a > 0$, αρνητικό αν $b \leq 0$ κι αυστηρά αρνητικό αν $b < 0$
- Δύο διαστήματα $[a,b]$ και $[c,d]$ είναι ίσα αν και μόνο αν $a = c$ και $b = d$
- Οι αριθμοί-διαστήματα είναι μερικώς διατεταγμένοι. Έχουμε $[a,b] < [c,d]$ αν και μόνο αν $b < c$

7.3.2 Βασικές πράξεις της αριθμητικής διαστημάτων

Ας θεωρήσουμε τα σύμβολα $+$, $-$, $*$, $/$ που συμβολίζουν τις πράξεις της πρόσθεσης, της αφαίρεσης, του πολλαπλασιασμού και της διαίρεσης αντίστοιχα. Αν με το σύμβολο op συμβολίσουμε καθεμία από τις παραπάνω αριθμητικές πράξεις δύο πραγματικών αριθμών x και y , τότε η αντίστοιχη λειτουργία για δύο αριθμούς-διαστήματα X και Y θα είναι

$$X \text{ op } Y = \{ x \text{ op } y : x \in X, y \in Y \}$$

Έτσι το αποτέλεσμα της πράξης $X \text{ op } Y$, είναι ένα διάστημα, στο οποίο ανήκουν όλοι εκείνοι οι πραγματικοί αριθμοί, που προκύπτουν ως το αποτέλεσμα της πράξης $x \text{ op } y$ για κάθε $x \in X$ και κάθε $y \in Y$.

Ο παραπάνω ορισμός παράγει τους παρακάτω κανόνες για τον προσδιορισμό των άκρων του διαστήματος που προκύπτει μετά την πράξη $X \text{ op } Y$, δύο διαστημάτων $X = [a, b]$ και $Y = [c, d]$

Πρόσθεση

$$X + Y = [a + c, b + d]$$

Αφαίρεση

$$X - Y = [a - d, b - c]$$

Πολλαπλασιασμός

- | | |
|----------------------------------|--|
| • Αν $a \geq 0$ και $c \geq 0$ | $X * Y = [ac, bd]$ |
| • Αν $a \geq 0$ και $c < 0 < d$ | $X * Y = [bc, bd]$ |
| • Αν $a \geq 0$ και $d \leq 0$ | $X * Y = [bc, ad]$ |
| • Αν $a < 0 < b$ και $c \geq 0$ | $X * Y = [ad, bc]$ |
| • Αν $a < 0 < b$ και $d \leq 0$ | $X * Y = [bd, ad]$ |
| • Αν $b \leq 0$ και $c \geq 0$ | $X * Y = [ad, bc]$ |
| • Αν $b \leq 0$ και $c < 0 < d$ | $X * Y = [ad, ac]$ |
| • Αν $b \leq 0$ και $d \leq 0$ | $X * Y = [bd, ac]$ |
| • Αν $a < 0 < b$ και $c < 0 < d$ | $X * Y = [\min(bc, ad), \max(ac, bd)]$ |

Διαίρεση

$$\frac{1}{Y} = \left[\frac{1}{d}, \frac{1}{c} \right] \quad (0 \notin Y)$$

$$\frac{X}{Y} = X * \left(\frac{1}{Y} \right) \quad (0 \notin Y)$$

Επίσης ορίζουμε την πράξη ύψωσης σε δύναμη

$$\begin{aligned} X^n &= [1, 1] && \text{αν } n = 0 \\ &[a^n, b^n] && \text{αν } a \geq 0 \text{ ή αν } a \leq 0 \leq b \text{ και } n \text{ είναι περιττός} \\ &[b^n, a^n] && \text{αν } b \leq 0 \\ &[0, \max(a^n, b^n)] && \text{αν } a \leq 0 \leq b \text{ κι ο } n \text{ είναι άρτιος για } n=1, 2, \dots \end{aligned}$$

7.3.3 Το πρόβλημα της εξάρτησης (The Dependency problem)

Ας υποθέσουμε ότι αφαιρούμε το διάστημα $X = [a, b]$ από τον εαυτό του. Χρησιμοποιώντας τον προηγούμενο κανόνα για την αφαίρεση διαστημάτων στην αριθμητική διαστημάτων, παίρνουμε σαν αποτέλεσμα το διάστημα $[a-b, b-a]$. Αυτό ίσως προκαλεί εντύπωση, γιατί αναμενόμενο αποτέλεσμα θα ήταν το διάστημα $[0, 0]$. Παρ'όλα αυτά όμως το αποτέλεσμα είναι $\{x - y : x \in X, y \in Y\}$, αντί για $\{x - x : x \in X\}$.

Γενικά, όταν μια μεταβλητή εμφανίζεται περισσότερες από μια φορά σε κάποιον υπολογισμό της αριθμητικής διαστημάτων, αντιμετωπίζεται σαν διαφορετική μεταβλητή σε κάθε της εμφάνιση. Έτσι η πράξη $X - X$ είναι η ίδια με την $X - Y$ με το Y να είναι ίσο, αλλά ανεξάρτητο από το X . Αυτό έχει ως αποτέλεσμα να έχουμε πιο πλατιά διαστήματα (*wide intervals*) και συνεπώς λιγότερη ακρίβεια στους υπολογισμούς.

Ο ορισμός που δώσαμε για την πράξη ύψωσης σε δύναμη ενός διαστήματος, αντιμετωπίζει με επιτυχία το παραπάνω πρόβλημα της εξάρτησης. Για παράδειγμα για $n = 2$, ο ορισμός είναι ισοδύναμος με :

$$X^2 = \{x^2 : x \in X\}, \text{ που είναι διαφορετικός από τον ορισμό}$$

$$X * X = \{x * y : x \in X, y \in X\}$$

Επομένως είναι διαφορετικό το αποτέλεσμα της πράξης $[-1, 2]^2 = [0, 4]$ από το αποτέλεσμα $[-1, 2] * [-1, 2] = [-2, 4]$.

Εδώ θα πρέπει να σημειώσουμε ότι αν μια μεταβλητή τύπου διαστήματος εμφανίζεται μόνο μια φορά σε κάποιο τύπο της συνάρτησης, τότε δεν ερχόμαστε αντιμέτωποι με το πρόβλημα της εξάρτησης. Έτσι εξάρτηση μπορεί να έχουμε σε μια συνάρτηση $f(X, Y)$ της μορφής :

$$f(X, Y) = (X - Y) / (X + Y), \text{ αλλά όχι αν αυτή είναι γραμμένη με την μορφή } f(X, Y) = 1 - 2 / (1 + X / Y)$$



Αν υπολογίσουμε την συνάρτηση f σύμφωνα με την τελευταία μορφή, το διάστημα που προκύπτει σαν αποτέλεσμα είναι ακριβώς το σύνολο τιμών της συνάρτησης $f(x, y)$ όταν $x \in X$ και $y \in Y$

7.3.4 Συναρτήσεις διαστημάτων (Interval Functions) - Ιδιότητες

7.3.4.1 Συναρτήσεις πραγματικών τιμών για διαστήματα (*real-valued interval functions*)

Υπάρχουν πολλές συναρτήσεις που παίρνουν ως ορίσματα διαστήματα κι επιστρέφουν πραγματική τιμή. Οι πιο σημαντικές από αυτές είναι

- το μέσο ή το κέντρο ενός διαστήματος $X = [a, b]$ (**midpoint**)

$$m(x) = \frac{a+b}{2}$$

- το πλάτος ενός διαστήματος (**width of an interval**)

$$w(X) = b - a$$

- η απόλυτη τιμή του διαστήματος X

$$|X| = \max \{ |a|, |b| \}$$

η απόλυτη τιμή ονομάζεται και **magnitude**

7.3.4.2 Συναρτήσεις διαστήματος (interval functions)

Μια συνάρτηση διαστήματος (**interval function**) είναι μια συνάρτηση που παίρνει σαν όρισμα διάστημα και η επιστρεφόμενη τιμή της είναι επίσης διάστημα. Ας θεωρήσουμε μια συνάρτηση f πραγματικών τιμών, με πραγματικές μεταβλητές x_1, x_2, \dots, x_n και μια συνάρτηση διαστήματος F με μεταβλητές τύπου διαστήματος X_1, X_2, \dots, X_n . Η συνάρτηση διαστήματος F ονομάζεται κι επέκταση διαστήματος της f (*interval extension of f*) αν

$$F(x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n) \quad \forall x_i (i=1, \dots, n)$$

Αν τα ορίσματα της F είναι διαστήματα «εκφυλισμένα» (*degenerate*) (υπενθυμίζουμε ότι ένα διάστημα είναι «εκφυλισμένο», αν $x_i = x_i$, δηλαδή το κάτω και το άνω άκρο είναι ίσα), τότε και το $F(X_1, \dots, X_n)$ είναι ένα διάστημα επίσης «εκφυλισμένο» ίσο με $f(x_1, x_2, \dots, x_n)$.

Ο ορισμός προϋποθέτει την ύπαρξη ακριβούς αριθμητικής διαστημάτων. Στην πράξη όταν έχουμε και στρογγύλευση ισχύει :

$$f(x_1, x_2, \dots, x_n) \in F(x_1, x_2, \dots, x_n), \text{ όπου } F \text{ είναι μια επέκταση διαστήματος της } f.$$

Η ιδιότητα της inclusion monotonicity

Μια συνάρτηση διαστήματος ονομάζεται **inclusion monotonic** [28] αν $X_i \subset Y_i$ ($i=1,2$) τότε ισχύει $F(X_1, \dots, X_n) \subset F(Y_1, \dots, Y_n)$

Επίσης η αριθμητική διαστημάτων είναι inclusion monotonic. Αν το op παριστάνει κάποιον από τους τελεστές +, -, *, / και $X_i \subset Y_i$ ($i = 1,2$) συνεπάγεται ότι

$$(X_1 \text{ op } X_2) \subset (Y_1 \text{ op } Y_2)$$

Αν χρησιμοποιείται εξωτερική στρογγύλευση (*outward rounding*), τότε η αριθμητική διαστήματος είναι inclusion monotonic.

Για απλοποίηση κάποιων συμβολισμών θα θεωρούμε από εδώ και στο εξής με τον τύπο $f(X_1, \dots, X_n)$ μια επέκταση διαστήματος της συνάρτησης πραγματικών τιμών $f(x_1, \dots, x_n)$. Θα θεωρούμε λοιπόν ότι κάθε συνάρτηση της οποίας τα ορίσματα είναι διαστήματα (intervals), είναι επίσης μια συνάρτηση διαστήματος.

Το γεγονός ότι οι τελεστές της αριθμητικής διαστημάτων είναι inclusion monotonic, βεβαιώνει ότι και ρητές (rational) συναρτήσεις διαστήματος είναι επίσης inclusion monotonic. Το ακόλουθο παράδειγμα δείχνει ότι αν μια ρητή συνάρτηση αποτιμάται (υπολογίζεται) με διαφορετικούς τρόπους σε διαφορετικά διαστήματα, μπορεί να πάψει να είναι inclusion monotonic.

Ας υποθέσουμε ότι γράφουμε τη συνάρτηση $f(x) = x(1-x)$ στην μορφή $f(x) = c(1-c) + (1-2c)(x-c) - (x-c)^2$

Αυτοί οι δύο τύποι είναι ισοδύναμοι για κάποια τιμή του c . Ας θεωρήσουμε το διάστημα $X = [0,1]$ και $C = m(X) = 0.5$ (υπενθυμίζουμε ότι η συνάρτηση m επιστρέφει το μέσο (midpoint) του διαστήματος). Αποτιμάται η συνάρτηση σύμφωνα με τον προηγούμενο τύπο και παίρνουμε σαν αποτέλεσμα $f([0,1]) = [0, 0.25]$.

Αντικαθιστούμε το $X = [0,1]$ με το $X' = [0,0.9]$. Επίσης αντικαθιστούμε το c με το $c' = m(X') = 0.45$. Τότε παίρνουμε σαν αποτέλεσμα $f(X') = [0,0.2925]$. Παρατηρούμε ότι η $f(X')$ δεν περιέχεται στην $f(X)$ ακόμη κι αν $X' \subset X$. Η ιδιότητα της inclusion monotonicity έπαψε να ισχύει, επειδή αλλάξαμε το τύπο της συνάρτησης αντικαθιστώντας το c με το c' .

Μπορεί κάποιος να ισχυριστεί ότι ο τύπος της συνάρτησης δεν άλαξε αφού και στην μία και στην άλλη περίπτωση το c είναι ίσο με τα μέσα των αντίστοιχων διαστημάτων (δηλαδή $c = m(X)$ και $c' = m(X')$). Το μέσο αυτών των διαστημάτων δεν υπολογίζεται όμως χρησιμοποιώντας μόνο τους τελεστές της πρόσθεσης, αφαίρεσης, πολλαπλασιασμού και διαίρεσης. Πρέπει να κληθεί και μια ξεχωριστή διαδικασία για τον υπολογισμό των άκρων του μέσου του διαστήματος.

Το θεώρημα που ακολουθεί φανερώνει ότι για ρητές (rational) συναρτήσεις ισχύει η ιδιότητα της inclusion monotonicity.

Θεώρημα :

Ας θεωρήσουμε μια συνάρτηση $F(X_1, X_2, \dots, X_n)$, η οποία είναι μια ρητή συνάρτηση διαστήματος. Υποθέτουμε ακόμη ότι η συνάρτηση F αποτιμάται με σταθερό τύπο, που αποτελείται από σταθερή ακολουθία αριθμητικών πράξεων, που περιλαμβάνουν μόνο τις πράξεις της πρόσθεσης, της αφαίρεσης, του πολ/πλασιασμού και της διαίρεσης. Τότε αποδεικνύεται ότι η συνάρτηση F είναι *inclusion monotonic*.

Η απόδειξη αυτού του θεωρήματος είναι προφανής και προκύπτει από την ιδιότητα τα μονοτονικότητας των παραπάνω αριθμητικών τελεστών.

Κατά τον προσδιορισμό εκείνου του διαστήματος, που περιέχει το σύνολο τιμών της πραγματικής συνάρτησης, γίνεται χρήση της μονοτονικότητας. Πολλές φορές δεν χρησιμοποιούμε την μονοτονικότητα μόνο των τεσσάρων βασικών τελεστών, αλλά την χρησιμοποιούμε και για άλλες συναρτήσεις, όπως θα δούμε στη συνέχεια. Αρκεί να αποδεικνύουμε κάθε φορά, ότι το αποτέλεσμα καλύπτει την ιδιότητα της *inclusion monotonicity*.

Ας εξετάσουμε την περίπτωση των μη ρητών συναρτήσεων. Έστω μια πραγματική μη ρητή συνάρτηση f με πραγματικές παραμέτρους (συμβολίζονται από ένα πραγματικό διάνυσμα $x = (x_1, x_2, \dots, x_n)$). Υποθέτουμε επίσης ότι προσεγγίζουμε την μη ρητή συνάρτηση f , με μια συνάρτηση $r(x)$ που είναι ρητή, οπότε έχουμε το ακόλουθο σφάλμα προσέγγισης :

$$|f(x) - r(x)| < \varepsilon \text{ για όλα τα } x \text{ τέτοια ώστε } a_i \leq x \leq b_i$$

$$\text{Τότε, } f(X_1, \dots, X_n) \subset r(X_1, \dots, X_n) + [-\varepsilon, \varepsilon] \text{ για κάθε διάστημα } X_i \subset [a_i, b_i]$$

Συμπεραίνουμε λοιπόν ότι το διάστημα τιμών της συνάρτησης f , μπορεί να υπολογιστεί, αποτιμώντας την συνάρτηση $r(X_1, \dots, X_n)$, χρησιμοποιώντας την αριθμητική διαστημάτων και προσθέτοντας τα όρια για την συνάρτηση λάθους $[-\varepsilon, \varepsilon]$.

Η αποτίμηση της μη ρητής συνάρτησης διαστήματος F είναι *inclusion monotonic*, αν η συνάρτηση r που χρησιμοποιείται για την προσέγγιση της, είναι επίσης *inclusion monotonic*. Το παρακάτω θεώρημα είναι ίσως το πιο βασικό θεώρημα στην ανάλυση διαστημάτων.

Θεώρημα : Ας θεωρήσουμε ότι η $F(X_1, \dots, X_n)$ είναι μια επέκταση διαστήματος *inclusion monotonic*, μιας πραγματικής συνάρτησης $f(x_1, x_2, \dots, x_n)$. Τότε η $F(X_1, \dots, X_n)$ περιέχει το διάστημα του συνόλου τιμών της συνάρτησης $f(x_1, \dots, x_n)$ για όλα τα $x_i \in X_i$

Πρακτική αποτίμηση των συναρτήσεων διαστήματος [24]

Σε αυτό το σημείο θα περιγράψουμε παραθέτωνας κάποια παραδείγματα, πώς γίνεται η αποτίμηση των συναρτήσεων διαστήματος, με απλά λόγια πώς μπορούμε να υπολογίσουμε για μια συγκεκριμένη συνάρτηση με δεδομένο πεδίο ορισμού, το διάστημα εκείνο που περιλαμβάνει όλες τις τιμές της συνάρτησης.

Όταν κάνουμε επιστημονικούς υπολογισμούς με πραγματικούς αριθμούς, προσεγγίζουμε μη ρητές συναρτήσεις με ρητές. Έχουμε επίσης μια λίστα από υπορουτίνες που υπολογί-

ζουν προσεγγιστικά τιμές συναρτήσεων, όπως για παράδειγμα η τριγωνομετρική, η εκθετική κτλ.

Χρειαζόμαστε και στην περίπτωση των συναρτήσεων διαστήματος μια παρόμοια λίστα υπορουτίνων. Ας θεωρήσουμε ένα διάστημα $X = [a, b]$ κι ας υποθέσουμε ότι θέλουμε μια υπορουτίνα για τον υπολογισμό της εκθετικής $\exp(X)$. Επειδή γνωρίζουμε ότι η εκθετική συνάρτηση είναι μονοτονική, τότε μπορούμε να πούμε ότι $\exp(X) = [\exp(a), \exp(b)]$.

Ο τρόπος για να παράγουμε αυτό το διάστημα υπολογίζεται ως εξής :

Υπολόγισε το $\exp(a)$ χρησιμοποιώντας μια αριθμητική υπορουτίνα διπλής ακρίβειας, (όχι υπορουτίνα της ανάλυσης διαστημάτων). Στρογγυλοποίησε το αποτέλεσμα στον μεγαλύτερο αριθμό μηχανής ο οποίος δεν υπερβαίνει την διπλής ακρίβειας τιμή του $\exp(a)$. Αυτό είναι και το επιθυμητό άκρο του διαστήματος για την συνάρτηση διαστήματος $\exp(X)$. Στη συνέχεια υπολόγισε το $\exp(b)$ χρησιμοποιώντας μια υπορουτίνα διπλής ακρίβειας. Στρογγυλοποίησε το αποτέλεσμα στον μικρότερο απλής ακρίβειας αριθμό μηχανής, που είναι τουλάχιστον ίσος με την διπλής ακρίβειας τιμή του $\exp(b)$. Αυτό είναι και το επιθυμητό δεξί άκρο για την συνάρτηση διαστήματος $\exp(X)$. Ας σημειώσουμε εδώ ότι η ακρίβεια των αποτελεσμάτων, εξαρτάται από την ακρίβεια των υπορουτίνων διπλής ακρίβειας (οι υπορουτίνες αυτές είναι για απλούς αριθμούς κι όχι για αριθμούς διαστήματα).

Ας δούμε ένα ακόμη παράδειγμα για να διασαφηνιστεί περισσότερο πώς γίνεται ο υπολογισμός μιας συνάρτησης διαστήματος με την χρήση της μονοτονικότητας. Θέλουμε λοιπόν να υπολογίσουμε το αποτέλεσμα για την συνάρτηση διαστήματος $\arctan(X)$ με ακρίβεια στο τρίτο δεκαδικό ψηφίο. Θα εκεμεταλλευτούμε και το γεγονός της μονοτονικότητας για την συνάρτηση $\arctan(x)$. Θα χρησιμοποιήσουμε αρχικά μια συνάρτηση για να την προσεγγίσουμε.

Το πολυώνυμο $P(x) = x(0.079331x^2 - 0.288679x + 0.995354)$ (1)
προσεγγίζει την συνάρτηση $\arctan(x)$ για $x \in [-1, 1]$

Το λάθος της παραπάνω προσέγγισης έχει ως άνω φράγμα :
 $\max |p(x) - \arctan(x)| < 0.0061$

Με βάση λοιπόν την παραπάνω προσέγγιση και το άνω φράγμα για το διάστημα τιμών της $\arctan(X)$ έχουμε :

$$\arctan(X) \in p(x) + [-0.00061, 0.00061]$$

Για $x \in [-1, 1]$, η συνάρτηση διαστήματος $p(X)$ προκύπτει αν αντικαταστήσουμε το x με το X .

Η υπορουτίνα για τον υπολογισμό της $\arctan(x)$ θα υπολογίσει το δεξιό μέλος της παραπάνω σχέσης, και θα επιστρέψει το διάστημα σαν αποτέλεσμα για την συνάρτηση διαστήματος. Στην πράξη κατά τον υπολογισμό της συνάρτησης $p(X)$, θα γράφαμε το πολυώνυμο με κάποια άλλη μορφή, όχι όπως δίνεται από τον τύπο (1). Χρησιμοποιούμε λοιπόν τον παρακάτω τύπο

$$P(X) = X[0.079331(X^2 - 1.18946)^2 + 0.733734] \quad (2)$$



Με αυτόν τον τρόπο μειώνονται οι εμφανίσεις του X και αποκτούμε έτσι καλύτερα αποτελέσματα για το διάστημα X . Ωστόσο, κατά την μετάβασή μας από τον τύπο (1) στον τύπο (2) έγιναν κάποια λάθη στρογγύλευσης. Θα πρέπει λοιπόν να ενημερώσουμε τα όρια του διαστήματος λάθους. Από το παραπάνω παράδειγμα φαίνεται πόσο προσεκτικοί θα πρέπει να είμαστε όταν γράφουμε υπορουτίνες της ανάλυσης διαστημάτων για την αποτίμηση μη ρητών συναρτήσεων.

Κεφάλαιο 8 – Περιγραφή αλγορίθμων διαστήματος

8.1 Εισαγωγή

Οι μέθοδοι διαστήματος, χρησιμοποιούν τις βασικές αρχές της ανάλυσης διαστημάτων για την επίλυση προβλημάτων ολικής ελαχιστοποίησης. Ο κύριος αλγόριθμος είναι ακολουθιακός κι αιτιοκρατικός και χρησιμοποιεί τις γνωστές τεχνικές διακλάδωσης και φράγματος (*branch and bound*), που περιγράψαμε στο προηγούμενο κεφάλαιο. Επίσης μπορεί να συνδυαστεί και με κάποιες άλλες τεχνικές που επιταχύνουν την λειτουργία του, όπως για παράδειγμα εξέταση κάποιων κριτηρίων (πχ έλεγχος μέσου) κι εφαρμογή τοπικών μεθόδων ελαχιστοποίησης. Οι τεχνικές διακλάδωσης και φράγματος υπενθυμίζουμε, ότι διαχωρίζουν την αρχική περιοχή ορισμού των παραμέτρων σε μικρότερες υποπεριοχές (*branching*), όπου και υπολογίζονται τα όρια του διαστήματος τιμών της συνάρτησης (*bounding*). Η εξέταση κάποιων κριτηρίων, έχει ως αποτέλεσμα την απόρριψη κάποιων περιοχών, οι οποίες αποκλείεται να περιέχουν την τιμή του πραγματικού ολικού ελαχίστου. Συνήθως, ο αλγόριθμος αυτός τερματίζει όταν το εύρος των διαστημάτων που απομένουν (μετά από διαδοχικούς τεμαχισμούς της αρχικής περιοχής), γίνει πολύ μικρό (μικρότερο από κάποια τιμή κατωφλίου).

8.2 Περιγραφή του αλγορίθμου [26]

8.2.1 Γενικό σχήμα

Τα δεδομένα εισόδου του αλγορίθμου είναι τα εξής

- Ένα αρχικό διάστημα X ορισμού των παραμέτρων. Το διάστημα X είναι ένα υπερρθογώνιο κι ορίζεται ως εξής :
$$X = \{ x: a_i \leq x_i \leq b_i, \text{ με } i = 1, 2, \dots, n \}$$

$$= ([a_1, b_1], \dots, [a_n, b_n]).$$
- Η συνάρτηση διαστήματος F παίρνει σαν όρισμα κάποιο διάστημα κι επιστρέφει το πεδίο τιμών της συνάρτησης για το συγκεκριμένο διάστημα.
- Κάποια τιμή κατωφλίου για τον έλεγχο κριτηρίου τερματισμού (το κριτήριο τερματισμού θα περιγραφεί στη συνέχεια).

Βήμα 1 : Υπολόγισε τα όρια των τιμών της συνάρτησης για το αρχικό υπερρθογώνιο X (με αποτίμηση της συνάρτησης διαστήματος $F(X)$).

Βήμα 2 : Υπολόγισε την τιμή του κάτω ορίου του διαστήματος τιμών της συνάρτησης $y = \min F(X)$.

Βήμα 3 : Αρχικοποίησε μια λίστα $L = (X, y)$. Κάθε κόμβος της λίστας θα περιέχει δύο στοιχεία, το διάστημα X και την ελάχιστη τιμή y της συνάρτησης στο συγκεκριμένο διάστημα.

Βήμα 4 : Επιλογή κατεύθυνσης τεμαχισμού του υπερρθογωνίου. Επιλέγεται ως κατεύθυνση τεμαχισμού εκείνη η διάσταση με το μέγιστο πλάτος. Το πλάτος w ενός διαστήματος $X=[a, b]$ δίνεται από τον τύπο $w = b - a$.

Αν λοιπόν το υπερορθογώνιο είναι ένα διάνυσμα $X = X_1 * X_2 * \dots * X_m$ (όπου m η διάσταση της συνάρτησης), για την κατεύθυνση τεμαχισμού k θα ισχύει $k = \max\{w(X_i)\}$, για $i = 1, 2, \dots, m$.

Βήμα 5 : Διχοτόμησε το υπερορθογώνιο X στην κατεύθυνση k . Έτσι προκύπτουν τα μικρότερα κουτιά V_1, V_2 , τέτοια ώστε $X = V_1 \cup V_2$.

Βήμα 6 : Υπολόγισε τα όρια των τιμών της συνάρτησης για τα νέα υπερορθογώνια V_1, V_2 . (Υπολογισμός των τιμών $F(V_1), F(V_2)$).

Βήμα 7 : Υπολόγισε τα κάτω όρια του διαστήματος τιμών της συνάρτησης για τα υπερορθογώνια V_1 και V_2 , $u_i = \min F(V_i)$, για $i = 1, 2$.

Βήμα 8 : Αφαίρεσε το υπερορθογώνιο $X = (X, y)$ από την λίστα L .

Βήμα 9 : Βάλε στην λίστα τα νέα υπερορθογώνια V_1, V_2 , με αύξουσα σειρά ως προς την τιμή του κάτω ορίου τιμών της συνάρτησης. Το πρώτο κουτί στην λίστα θα είναι αυτό που έχει την μικρότερη τιμή y δηλαδή την μικρότερη τιμή κάτω ορίου του διαστήματος τιμών της συνάρτησης.

Βήμα 10 : Έλεγχος κριτηρίου μέσου (*midpoint test*) για όλα τα υπερορθογώνια της λίστας κι αφαίρεση αυτών που δεν το ικανοποιούν.

Βήμα 11 : Έλεγχος κριτηρίου παραγώγου κι αφαίρεση των υπερορθογωνίων που δεν το ικανοποιούν.

Βήμα 12 : Βγάλε το πρώτο υπερορθογώνιο από την λίστα.

Βήμα 13 : Έλεγχος κριτηρίου τερματισμού. Αν ικανοποιείται τέλος, αλλιώς πήγαινε στο βήμα 4.

8.2.2 Περιγραφή κριτηρίου του μέσου (*midpoint criterion*)

Έστω mX το μέσο ενός υπερορθογωνίου X . Το μέσο για ένα διάστημα $[a, b]$ δίνεται από τον τύπο : $m(x) = \frac{a + b}{2}$

Γίνεται αποτίμηση της συνάρτησης διαστήματος F για το μέσο (το μέσο θεωρείται «εκφυλισμένο» διάστημα), οπότε έχουμε $f(mX) = [L_{fmX}, U_{fmX}]$, όπου L_{fmX}, U_{fmX} τα κάτω κι άνω όρια του διαστήματος τιμών της συνάρτησης στο μέσο.

Το U_{fmX} θα αποτελεί ένα άνω φράγμα του ολικού ελαχίστου της συνάρτησης. Έτσι θα ισχύει $U_{fmX} \geq f_*$, όπου f_* η τιμή του πραγματικού ολικού ελαχίστου. Αν τώρα υπολογίσουμε τα διαστήματα τιμών της συνάρτησης για κάποιο άλλο υπερορθογώνιο Y , θα έχουμε $f(Y) = [L_{fY}, U_{fY}]$, και μπορούμε να κάνουμε τον παρακάτω έλεγχο. Έστω U_{f*} η χαμηλότερη έως τώρα τιμή των U_{fmX} (δηλαδή των ανωτάτων ορίων των διαστημάτων τιμών της συνάρτησης)

- Αν $L_{fY} > U_{f*}$ το υπερορθογώνιο Y δεν μπορεί να περιέχει το ολικό ελάχιστο κι έτσι απορρίπτεται.

Σημειώνουμε στο σημείο αυτό, ότι αντί για το μέσο μπορούμε να πάρουμε και κάποιο άλλο τυχαίο σημείο του κουτιού X και να κάνουμε ακριβώς τους ίδιους ελέγχους.

8.2.3 Έλεγχος κριτηρίου παραγώγου

Εστω ένα υπερρθογώνιο X . Υπολογίζουμε τις μερικές παραγώγους g_i ($i=1, \dots, m$, όπου m η διάσταση της συνάρτησης). Στη συνέχεια παίρνουμε τα όρια τιμών της κάθε παραγώγου για το αρχικό υπερρθογώνιο X , υπολογίζοντας αντίστοιχα τα $g_i(X)$.

Αν $0 \notin g_i(X)$ για κάποιο $i = 1, \dots, m$, τότε η παράγωγος δεν μπορεί να είναι μηδέν στο συγκεκριμένο διάστημα X , αφού οι τιμές των μερικών παραγώγων στο ολικό ελάχιστο θα πρέπει να είναι μηδέν. Μπορούμε λοιπόν να απορρίψουμε το συγκεκριμένο υπερρθογώνιο.

Ο έλεγχος του κριτηρίου της παραγώγου μπορεί σε κάποιες περιπτώσεις να απορρίψει και διαστήματα μεγάλου πλάτους. Ας θεωρήσουμε την παρακάτω συνάρτηση με μια παράμετρο. Ο τύπος της είναι ο εξής :

$$F(x) = 24x^4 - 142x^3 + 303x^2 - 276x + 93$$

Εστω $X = [-10^{30}, 0]$. Υπολογίζοντας την παράγωγο g προκύπτει το εξής διάστημα τιμών $G(X) = [-0.9603 \times 10^{92}, -278]$. Ετσι μπορούμε να απορρίψουμε όλο το αρχικό διάστημα X , αφού η τιμή της παραγώγου δεν μπορεί να γίνει μηδέν. Το αποτέλεσμα που πήραμε με χρήση του κριτηρίου της παραγώγου είναι εντυπωσιακό, αφού καταφέραμε να απορρίψουμε ένα διάστημα πλάτους 10^{30} !!

8.2.4 Κριτήρια τερματισμού

Το κριτήριο τερματισμού που χρησιμοποιείται συνήθως στις μεθόδους διαστήματος για ολική ελαχιστοποίηση είναι : τερματισμός, όταν το μεγαλύτερο πλάτος των κουτιών της λίστας, είναι μικρότερο από μια τιμή κατωφλίου ϵ_x . Συνήθως η τιμή αυτή επιλέγεται να είναι πολύ μικρή, ώστε το διάστημα τιμών της συνάρτησης να έχει επίσης πολύ μικρό πλάτος. Βέβαια αυτό δεν είναι πάντα σωστό, για αυτό χρησιμοποιείται πολλές φορές και το κριτήριο τερματισμού $w(F(X)) < \epsilon_x$ [27] (δηλαδή το εύρος των τιμών της συνάρτησης να είναι πολύ μικρό). Όσο μικρότερο είναι το πλάτος του διαστήματος τιμών της συνάρτησης, τόσο καλύτερη προσέγγιση έχουμε για την τιμή του πραγματικού ολικού ελαχίστου.

8.2.5 Βελτιώσεις του παραπάνω αλγορίθμου [26]

Ο παραπάνω αλγόριθμος δεν είναι αποδοτικός στην εύρεση ολικών ελαχίστων, συναρτήσεων με πολλά τοπικά ελάχιστα, πολλές παραμέτρους και μεγάλο πλάτος του διαστήματος ορισμού των παραμέτρων. Ο έλεγχος του κριτηρίου του μέσου κι ο έλεγχος των παραγώγων δεν επαρκούν για την ομαλή σύγκλιση του αλγορίθμου, με αποτέλεσμα να δημιουργούνται λίστες με μεγάλο πλήθος υπερρθογωνίων. Σκεφτείτε την περίπτωση μιας συνάρτησης με πολλά τοπικά ελάχιστα διασκορπισμένα σε όλο το πλάτος του πεδίου ορισμού. Το κριτήριο ελέγχου των παραγώγων δεν απορρίπτει πολλά υπερρθογώνια με αποτέλεσμα η λίστα αυτή συνεχώς να διογκώνεται. Υπάρχουν και κάποια άλλα κριτήρια που σε πολλές περιπτώσεις, απορίπτουν μεγάλο ποσοστό υπερρθογωνίων. Για την χρησιμοποίηση αυτών των κριτηρίων θα πρέπει να υπολογίσουμε τις συναρτήσεις διαστήματος για τα στοιχεία των πινάκων *Jacobian* και *Hessian*. Αυτό βέβαια δεν είναι πάντα εύκολο για όλες τις συναρτήσεις, και δυσκολεύει σε μεγάλο βαθμό την υλοποίηση. Γενικά, πάντως μπορούμε να πούμε ότι σε δύσκολα προβλήματα, είναι απαραίτητος ο έλεγχος αρκετών κριτηρίων. Επίσης αυξημένες είναι και οι ανάγκες σε υπολογιστικούς πόρους.

Τα υπόλοιπα κριτήρια για την απόρριψη υπερорθογωνίων, τα οποία χρησιμοποιούνται μόνο στην περίπτωση παραγωγίσιμων συναρτήσεων είναι τα ακόλουθα :

- Έλεγχος μη κυρτότητας

Ας θεωρήσουμε ένα διάστημα B κι ένα υποδιάστημα X του B . Αν το ολικό ελάχιστο βρίσκεται στο σημείο x^* του διαστήματος B , τότε η συνάρτηση f θα πρέπει να είναι *κυρτή* (convex) στην γειτονιά του x^* . Με άλλα λόγια θα πρέπει ο *Εσσιανός* πίνακας (**Hessian**) να είναι *μη αρνητικά ορισμένος* στο x^* . Αν αποδειχτεί ότι ο *Εσσιανός* πίνακας δεν είναι *μη αρνητικά ορισμένος* στο διάστημα X , τότε το διάστημα X θα πρέπει να απορριφθεί.

Η συνθήκη που θα πρέπει να ισχύει για να είναι ο *Εσσιανός* πίνακας *μη αρνητικά ορισμένος* είναι ότι τα στοιχεία της διαγωνίου του θα πρέπει να είναι *μη αρνητικά*. Με την βοήθεια λοιπόν της ανάλυσης διαστήματος υπολογίζουμε τις συναρτήσεις διαστήματος $h_{ii}(X)$ για κάθε i . Αν ισχύει $h_{ii}(X) < 0$ για κάποιο i , τότε το κουτί X μπορεί να απορριφθεί.

- Χρήση μιας Interval Newton μεθόδου

Εστω η διανυσματική συνάρτηση g των *μερικών παραγώγων* της f . Θέλουμε να βρούμε τα σημεία, όπου η παράγωγος g μηδενίζεται για κάποιο υπερорθογώνιο X . Εστω J ο **Jacobian** πίνακας της g και x κάποιο σημείο στο υπερорθογώνιο X . Για κάθε σημείο y μέσα στο X , ισχύει

$$g(y) \in g(x) + J(X)(y-x)$$

Αν το y μηδενίζει την παραγάγο g , τότε το y είναι στο σύνολο λύσεων της $g(x) + J(X)(y-x) = 0$

Μπορούμε λοιπόν λύνοντας την παραπάνω εξίσωση να βρούμε ένα υπερорθογώνιο Y , που περιέχει όλα τα σημεία y που είναι λύσεις της παραπάνω εξίσωσης. Τότε το υπερорθογώνιο X αντικαθίσταται από το $X \cap Y$. Η τομή αυτών των συνόλων μπορεί να είναι το κενό σύνολο, οπότε απορρίπτεται όλο το υπερорθογώνιο X .

Βελτιώσεις στην απόδοση των μεθόδων διαστήματος [27] για ολική ελαχιστοποίηση μπορούμε να έχουμε και με την προσθήκη των παρακάτω στοιχείων :

- Συνδυασμός με κάποια *τοπική μέθοδο ελαχιστοποίησης*. Οι μέθοδοι αυτές μπορούν να χρησιμοποιηθούν είτε στην αρχή του αλγορίθμου για την επιλογή καταλληλότερου αρχικού υπερорθογωνίου, είτε σε ενδιάμεσα στάδια για την λήψη χαμηλοτέρων τιμών της συνάρτησης.
- *Παράλληλοποίηση* του αλγορίθμου. Κάθε παράλληλος επεξεργαστής μπορεί να εφαρμόσει τον αλγόριθμο για ένα συγκεκριμένο τμήμα του αρχικού διαστήματος ορισμού των παραμέτρων της συνάρτησης.

8.2 Αποτελέσματα υλοποίησης

Ο αλγόριθμος αυτός υλοποιήθηκε σε FORTRAN77 κι έγινε χρήση της βιβλιοθήκης INTLIB που περιέχει χρήσιμες συναρτήσεις της ανάλυσης διαστήματος [25]. Τα αποτελέσματα αναφέρονται στο πλήθος των υπερρθογωνίων της τελικής λίστας, στο πλήθος των απορριφθέντων υπερρθογωνίων μετά τον έλεγχο του κριτηρίου του μέσου, στο πλήθος των απορριφθέντων υπερρθογωνίων μετά τον έλεγχο της παραγώγου, οι τελικές αποτιμήσεις της παραγώγου, οι τελικές αποτιμήσεις της συνάρτησης και τέλος η ελάχιστη τιμή που βρήκε ο αλγόριθμος.

1. Six hump camel function

$$f(x_1, x_2) = (4 - 2.1x_1^2 + x_1^4/3) x_1^2 + x_1 x_2 + (-4 + 4x_2^2) x_2^2$$

Η τιμή του ολικού ελαχίστου είναι $f^* = -1.0316285$

Το πεδίο ορισμού των παραμέτρων είναι :
 $-2.5 < x_1 < 2.5$ και $-1.5 < x_2 < 1.5$

Τελικό πλήθος υπερρθογωνίων	15
Απορριφθέντα υπερρθογώνια μετά τον έλεγχο του κριτηρίου του μέσου	716
Απορριφθέντα κουτιά μετά τον έλεγχο του κριτηρίου της παραγώγου	1137
Συνολικό πλήθος αποτιμήσεων της συνάρτησης	2987
Συνολικό πλήθος αποτιμήσεων της παραγώγου	2156
Επιστρεφόμενη τιμή ολικού ελαχίστου	-1.03162849

2. Three hump camel function

$$f(x_1, x_2) = 12x_1^2 - 6.3x_1^4 + x_1^6 + 6(x_2 - x_1)$$

Η τιμή του ολικού ελαχίστου είναι $f^* = 0$

Το πεδίο ορισμού των παραμέτρων είναι :
 $-5 < x_1 < 5$ και $5 < x_2 < 5$

Τελικό πλήθος υπερορθογωνίων	19
Απορριφθέντα υπερορθογώνια μετά τον έλεγχο του κριτηρίου του μέσου	805
Απορριφθέντα κουτιά μετά τον έλεγχο του κριτηρίου της παραγώγου	446
Συνολικό πλήθος αποτιμήσεων της συνάρτησης	3796
Συνολικό πλήθος αποτιμήσεων της παραγώγου	3162
Επιστρεφόμενη τιμή ολικού ελαχίστου	-2.96224585E-33

3. Rosenbrock function

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (x_1 - 1)^2$$

Η τιμή του ολικού ελαχίστου είναι $f^* = 0$

Το πεδίο ορισμού των παραμέτρων είναι :

$-5 < x_1 < 5$ και $5 < x_2 < 5$

Τελικό πλήθος υπερορθογωνίων	11
Απορριφθέντα υπερορθογώνια μετά τον έλεγχο του κριτηρίου του μέσου	467
Απορριφθέντα κουτιά μετά τον έλεγχο του κριτηρίου της παραγώγου	875
Συνολικό πλήθος αποτιμήσεων της συνάρτησης	2007
Συνολικό πλήθος αποτιμήσεων της παραγώγου	1789
Επιστρεφόμενη τιμή ολικού ελαχίστου	-3.21452723E-65

Επίλογος

Η παρούσα διατριβή ασχολείται με το θέμα εύρεσης του ολικού ελαχίστου μιας συνάρτησης. Περιγράφεται ένα σύνολο στοχαστικών μεθόδων, που είναι υλοποιημένες σε MCL, μια γλώσσα προγραμματισμού για το περιβάλλον βελτιστοποίησης Merlin. Όπως μπορεί κανείς να διαπιστώσει κι από τα πειράματα, στις περισσότερες περιπτώσεις έχουμε ακριβή αποτελέσματα, ακόμη και σε δύσκολα προβλήματα ολικής ελαχιστοποίησης. Οι υβριδικές μέθοδοι, είναι αυτές που υπερέχουν έναντι των άλλων μεθόδων. Θα πρέπει να τονίσουμε ότι η γλώσσα MCL είναι πολύ ευέλικτη στην υλοποίηση τέτοιων μεθόδων, αφού μπορεί να εκμεταλλευτεί άμεσα τις τεχνικές τοπικής ελαχιστοποίησης του Merlin.

Οι μελλοντικές επεκτάσεις στο συγκεκριμένο θέμα είναι πολλές, αφού η ολική ελαχιστοποίηση, είναι ένα αρκετά δύσκολο πρόβλημα, που αποτελεί σημαντικό αντικείμενο έρευνας. Σημαντική βελτίωση των παραπάνω μεθόδων, τόσο στην ταχύτητα όσο και στην αξιοπιστία των αποτελεσμάτων, μπορεί να επιτευχθεί μέσω της παραλληλοποίησης των μεθόδων. Οι περισσότερες από αυτές τις μεθόδους μπορούν να υλοποιηθούν και παράλληλα, με μικρές αλλαγές στους αλγορίθμους που παρουσιάστηκαν. Επίσης ο συνδυασμός των διαφόρων μεθόδων ολικής ελαχιστοποίησης, μπορεί να επιφέρει ικανοποιητικά αποτελέσματα. Ο συνδυασμός στοχαστικών κι αιτιοκρατικών μεθόδων κρίνεται επίσης επιτυχής σε δύσκολα προβλήματα ολικής ελαχιστοποίησης. Οι μέθοδοι που περιγράφηκαν στην παρούσα διατριβή είναι σίγουρα οι πιο αποδεκτές μέθοδοι, που επιλύουν τα περισσότερα από τα προβλήματα της ολικής βελτιστοποίησης. Βέβαια στο σημείο αυτό πρέπει να τονίσουμε την αναγκαιότητα ύπαρξης εξειδικευμένων μεθόδων για συγκεκριμένα προβλήματα.

ΠΑΡΑΡΤΗΜΑ

Για τα διάφορα πειράματα χρησιμοποιήθηκε μια σειρά από συναρτήσεις [31]. Οι συναρτήσεις αυτές είναι αντιπροσωπευτικά παραδείγματα για δοκιμές, των διαφόρων μεθόδων ολικής ελαχιστοποίησης. Γενικά θεωρούνται δύσκολα προβλήματα, με πολλά τοπικά ελάχιστα και πολλές παραμέτρους. Το αποτέλεσμα μιας απλής τοπικής μεθόδου σε κάθε μία από αυτές τις συναρτήσεις, απέχει πολύ από την πραγματική τιμή του ολικού ελαχίστου. Χρησιμοποιήθηκαν λοιπόν οι παρακάτω συναρτήσεις

2. Sixhump camel back function

$$F(x_1, x_2) = (4 - 2.1x_1^2 + x_1^4/3) x_1^2 + x_1 x_2 + (-4 + 4x_2^2) x_2^2$$

Η τιμή του ολικού ελαχίστου είναι $f^* = -1.0316285$.

Το πεδίο ορισμού των παραμέτρων είναι :
 $-2.5 < x_1 < 2.5$ και $-1.5 < x_2 < 1.5$.

2. Goldstein and Price

$$F(x) = [1 + (x_1 + x_2 + 1)^2 (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1 x_2 + 3x_2^2)] \\ [30 + (2x_1 - 3x_2)^2 (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1 x_2 + 27x_2^2)]$$

Το ολικό ελάχιστο βρίσκεται στο σημείο $x^* = (0, -1)^T$
και η τιμή του είναι $f(x^*) = 3$.

3. Rastrigin

$$f(x) = x_1^2 + x_2^2 - \cos 18x_1 - \cos 18x_2$$

Πεδίο ορισμού $-1 < x_i < 1$, $i=1, 2$.

Το ολικό ελάχιστο είναι $f^* = -2$ στο σημείο $x^* = (0, 0)$.

Υπάρχουν περίπου 50 τοπικά ελάχιστα στο παραπάνω πεδίο ορισμού.

4. Griewank2

$$F(x) = \frac{x_1^2 + x_2^2}{4000} + 1 - \left(x_1 \cdot \frac{x_2}{\sqrt{2}}\right)$$

Πεδίο ορισμού $-100 < x_i < 100$, $i=1, 2$.

Το ολικό ελάχιστο είναι $f^* = 0$ στο σημείο $x^* = (0, 0)$.

Υπάρχουν περίπου 500 τοπικά ελάχιστα στο παραπάνω πεδίο ορισμού.

5. Griewank10

$$F(x) = \sum_{i=1}^{10} \frac{x_i^2}{4000} + 1 - \prod_{i=1}^{10} \frac{x_i}{\sqrt{i}}$$



Πεδίο ορισμού $-100 < x_i < 100$, $i = 1, \dots, 10$.

Το ολικό ελάχιστο είναι $f^* = 0$ στο σημείο $x^* = (0, \dots, 0)$.

Υπάρχουν κάποιες χιλιάδες τοπικά ελάχιστα στο παραπάνω πεδίο ορισμού.

6. Συνάρτηση σφάλματος ενός νευρωνικού δικτύου

Χρησιμοποιήθηκε ένα MLP με ένα κρυμμένο επίπεδο, μια είσοδο και μια έξοδο για την προσέγγιση ενός πολυωνύμου. Η έξοδος y του νευρωνικού δικτύου για μια είσοδο x δίνεται από την παρακάτω σχέση

$$y = \sum_{i=1}^k w_k g(u_k x + u_0)$$

y : η έξοδος του νευρωνικού δικτύου.

w_k : το βάρος σύνδεσης από τον κρυμμένο νευρώνα k στην έξοδο y .

u_k : το βάρος σύνδεσης από την είσοδο x στον κρυμμένο νευρώνα k .

u_0 : η πόλωση.

g : η σιγμοειδής συνάρτηση

$$g_j = \frac{1}{1 + \exp(-u_j)}, \text{ όπου } u_j \text{ είναι η τιμή ενεργοποίησης του νευρώνα } j \text{ και } g_j$$

είναι η τιμή εξόδου του.

k : ο αριθμός των κρυμμένων νευρώνων.

Χρησιμοποιήθηκαν 30 πρότυπα κι ελαχιστοποιήθηκε η παρακάτω συνάρτηση τετραγωνικού σφάλματος :

$$E = \sum_{i=1}^N \{ y(x^n; w) - t^n \}^2$$

N : το πλήθος των προτύπων εκπαίδευσης.

$y(x^n; w)$: η έξοδος του νευρωνικού δικτύου για το πρότυπο x^n .

w : το διάνυσμα των βαρών (συνιστώσες του είναι τα βάρη w_k (από τον κρυμμένο νευρώνα στην έξοδο), τα βάρη u_k (από την είσοδο στον κρυμμένο νευρώνα) και u_0 (η πόλωση)).

t^n : η επιθυμητή τιμή για την έξοδο, όταν το πρότυπο εισόδου είναι το x^n .

Η επιθυμητή ελάχιστη τιμή της παραπάνω συνάρτησης σφάλματος είναι το 0.

Χρησιμοποιήθηκε ένα MLP με 5 κόμβους (15 παράμετροι) για την προσέγγιση του πολυωνύμου $p(x) = 2x^5 + 3x^3 + 2x + 1$. Το πλήθος των δεδομένων είναι 30.

(Οι τιμές για την είσοδο x είναι στο διάστημα $-5 < x < 5$).

Αναφορές

- [1] D.G.Papageorgiou, I.N.Demitropoulos, I.E.Lagaris. "Merlin-3.0 A Multidimensional optimization environment". *Comp.Phys.Commun.*109 : 227-249, 1998.
- [2] D.G.Papageorgiou, I.N.Demitropoulos, I.E.Lagaris. "The Merlin Control Language for strategic optimization". *Comp.Phys.Commun.*109 : 250-270, 1998.
- [3] A.H.G Rinnooy Kan, G.T Timmer. "Global Optimization". G.L. Nemhauser et al., Eds, *Handbooks in OR & MS, Vol.1*, 1982.
- [4] A.H.G Rinnooy Kan, G.T Timmer. "Stochastic Methods for Global Optimization". *American Journal of Mathematical and Management Sciences*, 1983.
- [5] F. Glover. "Tabu Search –Part I". *ORSA Journal on Computing* 2(1) : 4-32, 1989.
- [6] F. Glover. "Tabu Search –Part II". *ORSA Journal on Computing* 2(1) : 4-32, 1990.
- [7] S.H. Brooks. "A Discussion of Random Methods for seeking Maxima". *Operation Research*, 6 : 244-251, 1958.
- [8] R.W. Becker and G.V. Lago. "A Global Optimization Algorithm". *Proceedings of the Eighth Allerton Conferences on Circuits and Systems Theory*, 1970.
- [9] W.L.Price. "A controlled random search procedure for global optimization". In *Towards Global Optimization 2*, L.C.W.Dixon and G.P.Szego (eds.), North-Holland, Amsterdam, 1974.
- [10] P.Brachetti, M.De Felice Ciccoli, G. Di Pillo and S. Lucidi. "A new version of the Price's algorithm for global optimization". *Journal of Global Optimization* 10 : 165-184, 1997.
- [11] S.Kiprpatrick, C. D. Gelatt and M. P.Vecchi. "Optimization by simulated annealing". *Science* 220 : 671-680, 1983.
- [12] S. Geman and D. Geman. "Stochastic relaxation, Gibbs distribution and Bayesian restoration of images". *IEEE Trans. Patt. Anal. Mac. Int.*6 : 721-741, 1984.
- [13] P. J. M van Laarhoven and E. H. L. Aarts. "Simulated Annealing: Theory and Applications". D.Riedel, Boston, 1997.
- [14] E. Aarts and J. Korst. "Simulated Annealing and Boltzman Machines". John Wiley and Sons, 1980.
- [15] L.Ingber. "Simulated Annealing : Practice versus theory". *J.Mathl.Comput.Modelling* 18 : 29-57, 1993.
- [16] W.L. Goffe, G. D. Ferrier and J. Rogers. "Global Optimization of Statistical Functions with Simulated Annealing". *J.Econometrics* 60 : 65-100, 1994.
- [17] X. Yao. "A new simulated annealing algorithm". *Intern. J. Computer Math.*Vol.56 : 161-168, 1994.
- [18] A.W. Jones and G. W. Forbes. "Adaptive Simulated Annealing". *Journal of Global Optimization* 6:1-37, 1995
- [19] R.L.Yang. "Convergence of the Simulated Annealing Algorithm for Continuous Global Optimization". *Journal of Optimization Theory and Applications*, Vol. 104, No 3 : 691-716, 2000.
- [20] L. Ozdamar, M. Demirhan. "Experiments with new stochastic global optimization search techniques". *Computers & Operations Research* 27 : 841-865, 2000.
- [21] R. Desai, R. Patil. "SALO : Combining Simulated Annealing and Local optimization for efficient global optimization". In *Proceedings of the 9th Florida AI Research Symposium (FLAIRS- '96)*, Key West, FL : 233-237, 1996.
- [22] D.E. Goldberg. "Genetic Algorithms in search optimization and Machine Learning". New-York, Addison-Wesley, 1989.
- [23] A.P Leclerc. "Efficient and global optimization". Dissertation, The Ohio State University, 1994.
- [24] E.Hansen. "Global Optimization Using Interval Analysis". *Pure and Applied Mathematics, A series of monographs and textbooks*, 1992.