



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Διερεύνηση της αποτελεσματικότητας της μεθόδου Covariance Matrix Adaptation- Evolution Strategy (CMA-ES) στην Βελτιστοποίηση χωρίς περιορισμούς

Διπλωματική Εργασία

του φοιτητή

Μπούρα Ν. Αθανάσιου

Επιβλέπων: Αθανάσιος Τόλης, Αναπληρωτής Καθηγητής Σχολής Μηχανολόγων
Μηχανικών ΕΜΠ

Υπόλοιπα μέλη Τριμελούς επιτροπής: Κοκκίνης Βασίλειος, Επίκουρος Καθηγητής
ΣΕΜΦΕ ΕΜΠ, και Κολέτσος Ιωάννης, Αναπληρωτής Καθηγητής ΣΕΜΦΕ ΕΜΠ

Ευχαριστίες

Η διπλωματική εργασία δεν θα μπορούσε να επιτευχθεί δίχως την πολύτιμη συμβολή του επιβλέποντα καθηγητή, Αθανάσιου Τόλη, ο οποίος καθοδηγούσε διαρκώς στην πορεία της εργασίας και με εισήγαγε στην ενασχόληση με ένα πολύ ευρύ και ενδιαφέρον πεδίο. Επιπρόσθετα, αυτή η διπλωματική εργασία είναι αφιερωμένη στον πατέρα μου, Νίκο, και σε όλη την οικογένεια μου για όλα τα εφόδια, την αγάπη και την πίστη που φρόντισαν να μου παρέχουν.

Περιεχόμενα

Έποψη.....	5
Κεφάλαιο 1: Εξελικτικοί αλγόριθμοι	7
1.1 Ιστορική Αναδρομή.....	7
1.1.1. Η θεωρία του Δαρβίνου.....	7
1.1.2 Από τον Δαρβίνο στον Rechenberg.....	8
1.2 Περιγραφή των Εξελικτικών αλγορίθμων.....	9
1.3 Δομή Εξελικτικών Αλγορίθμων	12
1.4 Οικογένεια των Εξελικτικών Αλγορίθμων.....	16
1.4.1 Γενετικοί Αλγόριθμοι (Genetic Algorithms).....	17
1.4.2 Στρατηγικές Εξέλιξης.....	22
Κεφάλαιο 2: Μαθηματικό Υπόβαθρο	26
2.1 Αποσύνθεση Ιδιοτιμών Θετικά Ορισμένου Πίνακα.....	27
2.2 Η Πολυδιάστατη Κανονική Κατανομή	29
2.3 Βελτιστοποίηση μιας Black Box συνάρτησης.....	31
2.4 Εσσιανός Πίνακας και Πίνακας Συνδιακύμανσης.....	32
Κεφάλαιο 3: Ανάλυση του αλγορίθμου	34
3.1 Η Βασική Εξίσωση.....	36
3.2 Μέση Τιμή m	37
3.3 Προσαρμόζοντας τον πίνακα συνδιακύμανσης.....	38
3.3.1 Υπολογίζοντας τον πίνακα συνδιακύμανσης C	38
3.3.2 Αναβάθμιση μ -βαθμού	41
3.3.2 Αναβάθμιση 1-βαθμού	44
3.3.3 Cumulation: Χρησιμοποιώντας την διαδρομή της εξέλιξης.....	45
3.4 Συνδυάζοντας την αναβάθμιση μ -βαθμού και την cumulation	47
3.5 Έλεγχος βήματος.....	48
3.6 Βήματα του Αλγορίθμου	52
Κεφάλαιο 4: Επίλυση Προβλημάτων μέσω του CMAES	54
4.1 Προσέγγιση Λύσεων Προβλημάτων	54
4.2 Σύνοψη Αποτελεσμάτων με την μορφή πίνακα:	75
Συμπεράσματα.....	76
Κατάλογος βιβλιογραφικών αναφορών.....	78

Έποψη

Η εργασία πραγματεύεται την χρήση της CMA (Covariance Matrix Adaptation) πάνω στην επιχειρησιακή έρευνα. Η CMA ανήκει στην οικογένεια των εξελικτικών αλγορίθμων και επιχειρεί ουσιαστικά να δώσει λύση σε πολύπλοκα προβλήματα τα οποία είναι αδύνατον να επιλυθούν με τις παραδοσιακές μεθόδους βελτιστοποίησης.

Για τον όρο επιχειρησιακή έρευνα έχουν δοθεί πολλοί ορισμοί. Χαρακτηριστικά, όπως είπε ο R. Watson-Watt και αναφέρει το σύγγραμμα του Κολέτσου [20], ο οποίος υπήρξε ένας από τους πρωτεργάτες της εισαγωγής και ανάπτυξης του θέματος στην Βρετανική αεροπορία ενώ παράλληλα ήταν ένας από τους εμπνευστές του ονόματος του κλάδου ως «Operational Research» (μαζί με τον A.P. Rowe):

«Η Επιχειρησιακή Έρευνα αποσκοπεί στο να ερευνήσει ποσοτικά εάν ένας οργανισμός παίρνει από την λειτουργία του εξοπλισμού του, τη βέλτιστη δυνατή συνεισφορά σε σχέση με τον ολικό αντικειμενικό σκοπό του, ποιες αλλαγές σε εξοπλισμό και μεθόδους απαιτούνται για τη βελτίωση των αποτελεσμάτων με το μικρότερο δυνατό κόστος σε προσπάθεια και χρόνο και τέλος σε ποιο βαθμό μεταβολές στους επιμέρους αντικειμενικούς σκοπούς θα συνεισέφεραν στην πιο οικονομική και έγκαιρη εκτέλεση του ολικού στρατηγικού αντικειμενικού σκοπού»

Έτσι, ένα πρόβλημα επιχειρησιακής έρευνας περιλαμβάνει τα ακόλουθα στάδια [20]:

- **Ανάλυση του Συστήματος** το οποίο αφορά αφορά την ανάλυση της δομής και του τρόπου λειτουργίας του συστήματος που καλείται να επιλυθεί
- **Διατύπωση Στόχων** κατά το οποίο οφείλουν να τεθούν οι στόχοι του προβλήματος, όπως π.χ. η ελαχιστοποίηση του κόστους
- **Διατύπωση του Μοντέλου** όταν και το πρόβλημα κατασκευάζεται σε μορφή εξισώσεων και ανισώσεων ούτως ώστε να τεθεί προς λύση
- **Επίλυση του Μοντέλου** όπου μέσω μαθηματικών τεχνικών όπως π.χ. της βελτιστοποίησης έχουμε την λύση του προβλήματος
- **Ανάλυση Ευαισθησίας** στην οποία μελετάται η επίδραση που θα έχει μια αλλαγή στις μεταβλητές στην λύση του προβλήματος

Είναι, λοιπόν, προφανής η σχέση μεταξύ Βελτιστοποίησης και Επιχειρησιακής Έρευνας και ως εκ τούτου είναι φανερή και η επίδραση που θα μπορούσε να έχει η CMA πάνω σε αυτήν.

Όπως προαναφέρθηκε, η CMA ανήκει στην οικογένεια των εξελικτικών αλγορίθμων. Από όταν έγινε αντιληπτή η χρησιμότητα τους, οι εξελικτικοί αλγόριθμοι έχουν αποτελέσει ένα σπουδαίο εργαλείο στα χέρια της επιστήμης το οποίο βρίσκει χρήσεις σε διάφορους τομείς όπως π.χ. η βελτιστοποίηση την οποία και θα μελετήσουμε. Στην αρχή βέβαια θα έλεγε κανείς ότι οι αλγόριθμοι αυτοί δεν είχαν την προσοχή που έπρεπε από την επιστημονική κοινότητα, και έτσι για πολλά χρόνια από την αρχή της δημιουργίας τους δεν παρείχαν κάποια σημαντική χρήση στον άνθρωπο. Αυτό βέβαια άλλαξε στην πορεία όταν και αντιλήφθηκαν πολλοί πως η δομή και ο τρόπος λειτουργίας τους τους επιτρέπει να είναι ιδιαίτερα χρήσιμοι σε πολλούς κλάδους.

Έτσι, φτάσαμε στο σημείο δημιουργίας της CMA που κατάφερε να επιλύσει πολύπλοκα και μέχρι τότε, σχεδόν αδύνατα προσεγγίσιμα, προβλήματα. Θα επιχειρήσουμε, λοιπόν, αφού αναλύσουμε τους εξελικτικούς αλγορίθμους και το μαθηματικό υπόβαθρο θα επιχειρήσουμε να εφαρμόσουμε τον αλγόριθμό μας σε διάφορα προβλήματα τα οποία έχουν δοθεί από τους Thomas Runarsson and Xin Yao για να μελετήσουμε το κατά πόσο επιτυγχάνεται η σύγκλιση σε μία λύση.

Θεωρώντας τα δοθέντα προβλήματα ως μοντελοποίηση προβλημάτων στα οποία απαιτείται βελτιστοποίηση, μπορούμε να αντιληφθούμε το πόσο μεγάλη σύνδεση υπάρχει με την επιχειρησιακή έρευνα και την CMA. Βέβαια, λόγω του γεγονότος ότι ο αλγόριθμός μας είναι στοχαστικός, σε κάθε πρόβλημα γίνεται επανάληψη της διαδικασίας του αλγορίθμου 6 φορές και μελέτη της απόκλισης των αποτελεσμάτων ούτως ώστε η προσέγγιση της λύσης να είναι όσο το δυνατόν καλύτερη.

Κεφάλαιο 1: Εξελικτικοί αλγόριθμοι

1.1 Ιστορική Αναδρομή

1.1.1. Η θεωρία του Δαρβίνου

Η θεωρία της Εξέλιξης των Ειδών (Evolution of Species) αναπτύχθηκε από τον Δαρβίνο, όπως ενημερώνει ο Λυκοθανάσης [21], στα μέσα του περασμένου αιώνα. Αρχικά, λόγω της βαθιάς θρησκοληψίας της τότε εποχής προκάλεσε μεγάλη αναστάτωση, αφού οι αρχές της ερχόταν σε σύγκρουση με τις επικρατούσες αντιλήψεις περί προέλευσης της ζωής. Με την πάροδο ενός και πλέον αιώνα, και λόγω της σταδιακής απαγκίστρωσης του ανθρώπου από την θρησκοληψία, ο θόρυβος αυτός δεν έχει κοπάσει πλήρως, όμως η θεωρία έχει γίνει αποδεκτή από το σύνολο των επιστημόνων, διότι κατόρθωσε να πείσει και να δώσει ικανοποιητικές απαντήσεις σε θεμελιώδη ερωτήματα. Συγκεκριμένα, η θεωρία του Δαρβίνου αποσκοπούσε στο να δώσει μια εξήγηση για το φαινόμενο της ζωής, την προέλευσή της και τις βασικές λειτουργίες της. Τα κυριότερα σημεία της, που σχετίζονται και ερμηνεύουν τον τρόπο λειτουργίας των Εξελικτικών Αλγορίθμων, είναι τα εξής:

- Δεν υπάρχει αντικειμενική βάση διαχωρισμού των ζωντανών οργανισμών σε ανώτερους και κατώτερους για το κάθε βιολογικό είδος. Σε κάθε ένα από αυτά, μερικά άτομα αφήνουν περισσότερους απογόνους σε σύγκριση με τα υπόλοιπα και έτσι τα κληροδοτούμενα χαρακτηριστικά τους αυξάνονται στην επόμενη γενιά. Βέβαια το ποιος οργανισμός θα επιβιώσει και θα καταφέρει να πολλαπλασιαστεί εξαρτάται από άλλες συνθήκες όπως είναι οι δυσκολίες και τα εμπόδια που παρουσιάζονται κατά τη διάρκεια της ζωής του κάθε οργανισμού. Έτσι, προκειμένου να μπορέσουν να προσαρμοστούν με την αλλαγή του περιβάλλοντος και να επιβιώσουν αλλάζουν τα χαρακτηριστικά τους και, πιο συγκεκριμένα, τα **χρωμοσώματά** τους (chromosomes), που είναι πολύπλοκα οργανικά μόρια τα οποία κωδικοποιούν τη δομή και τα χαρακτηριστικά τους. Τα χρωμοσώματα αποτελούνται από μικρότερα μέρη, γνωστά ως **γονίδια** (genes). Το σύνολο της γενετικής πληροφορίας που είναι κωδικοποιημένο στα γονίδια ονομάζεται **γονότυπος** (genotype). Η δημιουργία ενός νέου οργανισμού περιλαμβάνει την αποκωδικοποίηση των χρωμοσωμάτων. Το σύνολο των «ορατών» χαρακτηριστικών του και της συμπεριφοράς του, που καθορίζονται από τις πληροφορίες των γονιδίων, συνιστούν το **φαινότυπο** (phenotype).

- Οι κυριότερες λειτουργίες του φαινομένου της εξέλιξης είναι η **αναπαραγωγή** (reproduction) και η **μετάλλαξη** (mutation). Συγκεκριμένα:

-Η μετάλλαξη αφορά τον τυχαίο τρόπο με τον οποίο αλλάζει η δομή των χρωμοσωμάτων, συνήθως από λανθασμένη αντιγραφή βιολογικών μορίων ή από εξωγενείς παράγοντες (π.χ. ακτινοβολία), αλλάζοντας έτσι κάποιο χαρακτηριστικό. Η αλλαγή αυτή μπορεί να επιφέρει είτε θετικά είτε αρνητικά αποτελέσματα, επηρεάζοντας έτσι την μετέπειτα εξέλιξη.

-Μέσω της αναπαραγωγής, δημιουργείται ένας νέος οργανισμός του οποίου τα χρωμοσώματα αποτελούνται από ίδιο αριθμό γονιδίων του πατέρα και της μητέρας. Έτσι, για κάθε χαρακτηριστικό, το νέο άτομο έχει πάρει το γονίδιο από κάποιον από τους δύο γονείς του. Βέβαια, όταν τα γονίδια συμφωνούν μεταξύ τους το χαρακτηριστικό που θα δοθεί στο παιδί είναι ίδιο ενώ σε άλλη περίπτωση κυριαρχεί το γονίδιο ενός μόνο γονέα, αφήνοντας το γονίδιο του άλλου μόνο ως «κατάλειπο» επόμενων γενεών. Το γονίδιο που τελικά καθορίζει το χαρακτηριστικό λέγεται **κυρίαρχο** ή επικρατές (dominant) και το άλλο **υπολειπόμενο** (recessive). Γονίδια που διεκδικούν την ίδια θέση σε ένα χρωμόσωμα (δηλαδή που είναι υπεύθυνα για το ίδιο χαρακτηριστικό), λέγονται αλληλόμορφα (alleles).

1.1.2 Από τον Δαρβίνο στον Rechenberg

Μελετώντας την θεωρία του Δαρβίνου, ο μηχανισμός αυτός της φυσικής επιλογής φάνηκε ιδιαίτερα ελκυστικός στον Rechenberg, πρωτοπόρο των Γενετικών Αλγορίθμων.

Συγκεκριμένα, έπειτα από πολλές μελέτες τις δεκαετίες του '50 και του '60 διάφοροι επιστήμονες μελέτησαν, ανεξάρτητα ο ένας από τον άλλο, εξελικτικά συστήματα με τη σκέψη ότι η βιολογική διαδικασία της εξέλιξης θα μπορούσε να αποτελέσει τη βάση για την ανάπτυξη νέων αλγορίθμων βελτιστοποίησης σε τεχνολογικά προβλήματα. Η ιδέα πίσω από τα συστήματα αυτά, ήταν η ανάπτυξη ενός πληθυσμού από υποψήφιες λύσεις σε ένα δεδομένο πρόβλημα, χρησιμοποιώντας τελεστές εμπνευσμένους από τη φυσική γενετική μεταβολή και τη φυσική επιλογή.

Έτσι, ο Rechenberg εμπνευσμένος από κάποιες λειτουργίες της φύσης στα συστήματά της θέλησε να τις εφαρμόσει αφού πίστευε ότι θα μπορούσαν, αν ενσωματώνονταν σε

αλγόριθμους για υπολογιστές, να προκύψουν αποδοτικές τεχνικές επίλυσης δύσκολων και δυσεπίλυτων προβλημάτων. Έτσι, στις αρχές της δεκαετίας του '70 προέκυψαν ως αποτέλεσμα της εργασίας του οι “στρατηγικές εξέλιξης” ως μια μέθοδο βελτιστοποίησης για κατασκευές όπως τα λεπτά μεταλλικά φύλλα των αεροπλάνων. Η ιδέα αυτή αναπτύχθηκε παραπέρα από τον Schwefel (1975, 1977) αποτελώντας ένα βήμα που έδωσε ώθηση για την ανάπτυξη του κλάδου αυτού.

Ωστόσο εργασίες που αφορούσαν τους εξελικτικούς αλγορίθμους είχαν γίνει ήδη και από τον Fraser (1957). Παρόλα αυτά, στην εργασία αυτή χρησιμοποιούσε του εξελικτικούς αλγόριθμους μόνο ως εξομοιωτές της φυσικής διαδικασίας της εξέλιξης και όχι ως μέθοδος βελτιστοποίησης. Η εργασία του Box (1957) η οποία παρουσιάστηκε την ίδια εποχή αναφερόταν στη χρήση «εξελικτικών τελεστών» για τη βελτίωση της παραγωγικότητας εργοστασίων. Ωστόσο η μεθοδολογία που προτεινόταν στην εργασία αυτή δεν εφαρμοζόταν σε υπολογιστή αλλά απαιτούσε τη μεσολάβηση ενός ανθρώπου στην παραγωγική διαδικασία. Οπότε ήταν φανερό πως παρόλο που οι σκέψεις είχαν τα σωστά θεμέλια για την ανάπτυξη του κλάδου από εκείνη την εποχή δεν μπορούσε να επιτευχθεί.

Σύμφωνα με τον Goldberg (1989) ένας από τους βασικούς θεμελιωτές των εξελικτικών αλγορίθμων είναι ο J. Holland και ειδικότερα μέσω της εργασίας του «Adaptation in Natural and Artificial Systems» [10] που δημοσιεύτηκε το 1975 ο οποίος θέλησε να εξηγήσει τις διαδικασίες προσαρμογής που ακολουθούν φυσικοί οργανισμοί, συνδυάζοντας παράλληλα συστήματα μίμησης των διαδικασιών αυτών.

1.2 Περιγραφή των Εξελικτικών αλγορίθμων

Μια πρώτη εύστοχη περιγραφή θα μπορούσε να δοθεί από τον Jones [12] ο οποίος έδωσε μια 1η περιγραφή των εξελικτικών αλγορίθμων (EA) ως:

«Ο αλγόριθμος διατηρεί ένα σύνολο από πιθανές λύσεις σε ένα πρόβλημα. Ορισμένες από αυτές τις πιθανές λύσεις χρησιμοποιούνται για τη δημιουργία νέων πιθανών λύσεων μέσω της χρήσης τελεστών. Οι τελεστές επενεργούν επάνω στις πιθανές λύσεις και δημιουργούν νέες ομάδες αυτών. Οι πιθανές λύσεις πάνω στις οποίες επενεργεί ένας τελεστής επιλέγονται με βάση την ποιότητα που έχουν ως λύσεις του υπό εξέταση προβλήματος. Ο αλγόριθμος χρησιμοποιεί επαναληπτικά αυτή την επεξεργασία,

προκειμένου να δημιουργήσει νέες ομάδες πιθανών λύσεων, μέχρι να ικανοποιηθεί κάποιο κριτήριο τερματισμού».

Επιπρόσθετα οι Goldberg και Holland [6] τονίζουν ότι οι εξελικτικοί αλγόριθμοι είναι στοχαστικές μέθοδοι αναζήτησης, οι οποίες έχουν εφαρμοστεί επιτυχώς σε πολλά προβλήματα αναζήτησης, βελτιστοποίησης και μηχανικής μάθησης (machine learning). Σε αντίθεση με τις περισσότερες τεχνικές βελτιστοποίησης, οι ΕΑ χρησιμοποιούν έναν πιθανό πληθυσμό λύσεων, τον οποίο διαχειρίζονται με τέτοιο τρόπο, ώστε εφαρμόζοντας κάποιους τελεστές διαφοροποίησης, για την εύρεση μιας αρκετά ικανοποιητικής, αν όχι ολικά βέλτιστης, λύσης, πάνω σε πολλές φορές πολύπλοκα προβλήματα. Άλλες μέθοδοι εμφανίζουν περιορισμένες δυνατότητες σε προβλήματα με πολύ μεγάλο χώρο εφικτών λύσεων και έχουν αυξημένη πιθανότητα, από τη φύση τους, εγκλωβισμού σε τοπικά βέλτιστα καθιστώντας πολλές φορές την λύση του προβλήματος αδύνατη.

Η λειτουργία των Εξελικτικών Αλγόριθμων είναι η εξής, όπως αναφέρει η Α. Γκουντή [18]:

1. Αρχικά, έχουμε την δημιουργία ενός **πληθυσμού** (population) που αποτελείται από **άτομα** (individuals) των οποίων τα χαρακτηριστικά είναι τυχαία. Κάθε ένα από αυτά τα άτομα λαμβάνει μια **τιμή ικανότητας** (fitness value), ανάλογα με την φύση και τον στόχο του προβλήματος, η οποία και μεταβάλλεται με σκοπό βοηθήσει στην παραγωγή μιας υποψήφια λύσης.
2. Αφού έχουμε παράξει τον αρχικό πληθυσμό μας, αξιολογείται η ικανότητα της κάθε υποψήφιας λύσης, δηλαδή του κάθε ενός ατόμο του πληθυσμού, υπολογίζοντας την αξία της μέσω μιας **συνάρτησης ικανότητας** (fitness function). Αυτή η αξιολόγηση μπορεί να περιλαμβάνει περίπλοκες προσομοιώσεις και υπολογισμούς και προφανώς ο αριθμός των αξιολογήσεων εξαρτάται από τον αριθμό του πληθυσμού.
3. Με βάση αυτήν την αξιολόγηση, επιλέγονται τα άτομα εκείνα του πληθυσμού που έχουν μεγαλύτερη ικανότητα ούτως ώστε να παράξουν την **επόμενη γενιά** (γονείς). Έτσι, οι υποψήφιας λύσεις με χαμηλότερη συνάρτηση ικανότητας απορρίπτονται και αυτές με την μεγαλύτερη εισέρχονται στην «**πισίνα ζευγαρώματος**» (mating pool) με μεγαλύτερη πιθανότητα.

4. Στην φάση της αναπαραγωγής δημιουργούνται **απόγονοι** (offsprings) που εισήλθαν στην «πισίνα ζευγαρώματος», μεταλλάσσοντας ή/και διασταυρώνοντας τους γονότυπους των επιλεγμένων ατόμων με βάση την θεωρία του Δαρβίνου. Ο **ανασυνδυασμός** (recombination), ή αλλιώς **διασταύρωση** (crossover), θα μπορούσε να χαρακτηριστεί ως τελεστής που εφαρμόζεται σε δύο ή περισσότερα επιλεγμένα άτομα (γονείς) και δημιουργεί έναν ή περισσότερους απογόνους που αποτελούν λύσεις του προβλήματός που έχει τεθεί προς λύση ενώ ο τελεστής της **μετάλλαξης** (mutation) εφαρμόζεται σε ένα άτομο και δημιουργεί μία υποψήφια λύση.

5. Οι απόγονοι εν συνεχεία ενσωματώνονται στον πληθυσμό και έτσι προκύπτει ένας νέος πληθυσμός που θα ακολουθήσουν την ίδια διαδικασία για να «διεκδικήσουν» μια θέση στην επόμενη γενιά. Τα άτομα με μεγαλύτερη συνάρτηση ικανότητας έχουν περισσότερες πιθανότητες να επιβιώσουν όπως είδαμε στο βήμα 2.

6. Ακολουθώντας αυτήν την διαδικασία πολλές φορές, ο Εξελικτικός Αλγόριθμος σταματάει όταν φτάσουμε σε κάποια **συνθήκη τερματισμού** που έχει τεθεί από το πρόβλημα, τον χρήστη (υπολογιστικό όριο που έχει τεθεί) ή έχει να κάνει με την φύση του προβλήματος (κατάλληλη ικανότητα επιτευχθεί, κτλ). Το μέγεθος του πληθυσμού διατηρείται σταθερό σε κάθε επανάληψη.

Χαρακτηριστικά, δίνεται το παρακάτω σχήμα που αποτελεί μοντέλο ψευδοκώδικα του τρόπου λειτουργίας ενός εξελικτικού αλγόριθμου, όπως έχει αναλυθεί παραπάνω:

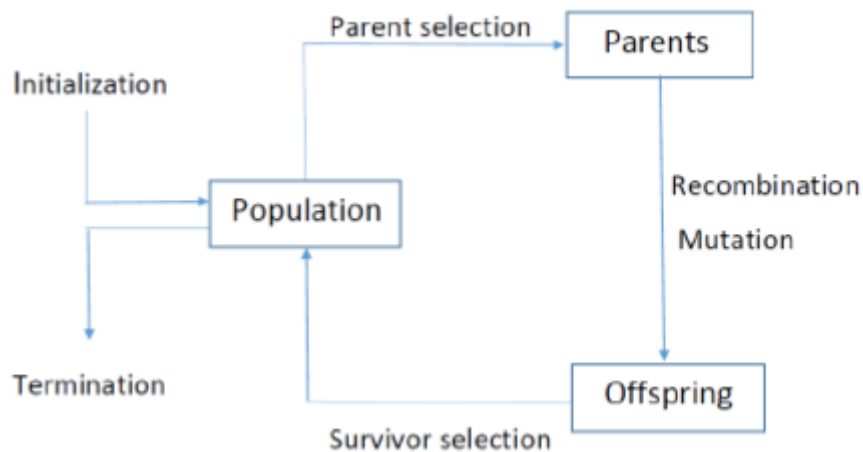
```
BEGIN
  INITIALISE population with random candidate solutions;
  EVALUATE each candidate;
  REPEAT UNTIL (TERMINATION CONDITION is satisfied) DO
    1 SELECT parents;
    2 RECOMBINE pairs of parents;
    3 MUTATE the resulting offspring;
    4 EVALUATE new candidates;
    5 SELECT individuals for the next generation;
  OD
END
```

Σχήμα 1.1: Πηγή: ΕΞΕΛΙΚΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ: Ο ΑΛΓΟΡΙΘΜΟΣ ΤΗΣ ΔΙΑΦΟΡΙΚΗΣ ΕΞΕΛΙΞΗΣ, Α. Γκουντή. Μοντέλο ψευδοκώδικα εξελικτικού αλγόριθμου

1.3 Δομή Εξελικτικών Αλγορίθμων

Θέλοντας να χαρακτηρίσουμε τους εξελικτικούς αλγόριθμους απαιτείται ένα σύνολο διαδικασιών και τελεστών που χρειάζεται να αναλυθούν ώστε να καθορίσουμε την δομή τους. Αυτό περιλαμβάνει, σύμφωνα με την Α. Γκουντή [18]:

- Αρχικοποίηση
- Αντιπροσώπευση(καθορισμός των ατόμων)
- Συνάρτηση αξιολόγησης(ή συνάρτηση ικανότητας)
- Πληθυσμός
- Μηχανισμός επιλογής γονέων
- Τελεστές διασταύρωσης και μετάλλαξης
- Μηχανισμός επιλογής επιβίωσης(αντικατάσταση)
- Κριτήρια τερματισμού



Σχήμα 1.2: Πηγή: ΕΞΕΛΙΚΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ: Ο ΑΛΓΟΡΙΘΜΟΣ ΤΗΣ ΔΙΑΦΟΡΙΚΗΣ ΕΞΕΛΙΞΗΣ, Α. Γκουντή. Γενική διαδικασία ενός εξελικτικού αλγόριθμου ως διάγραμμα ροής.

Αναλύοντας τις επιμέρους διαδικασίες έχουμε [18]:

Αρχικοποίηση

Αφορά την επιλογή των αρχικών πληθυσμών, ή **αρχικοποίηση** (initialization), μέσα στους αλγόριθμους. Γενικά, όπως προαναφέρθηκε η αρχικοποίηση είναι απλή στους περισσότερους Εξελικτικούς Αλγόριθμους αφού η επιλογή του αρχικού πληθυσμού γίνεται τυχαία. Βέβαια, σε μερικές περιπτώσεις για την αρχικοποίηση μπορούν να χρησιμοποιηθούν διάφορες μέθοδοι με στόχο έναν αρχικό πληθυσμό με υψηλή συνάρτηση ικανότητας ούτως ώστε να μπορέσει να πετύχει πιο εύκολα τον στόχο του,

δηλαδή σε λιγότερες επαναλήψεις της διαδικασίας μας. Βέβαια, οι τεχνικές αυτές διαφέρουν ανάλογα και με την φύση του κάθε προβλήματος και προφανώς τον στόχο τον οποίο έχουμε.

Αντιπροσώπευση

Το βασικό βήμα για να καθορίσουμε έναν εξελικτικό αλγόριθμο είναι μπορέσουμε να εμφωλεύσουμε ιδιότητες που απορρέουν από τον κόσμο που ζούμε, μέσα στο περιβάλλον που δημιουργείται ο Εξελικτικός Αλγόριθμος. Έτσι, για κάθε πρόβλημα, κάθε γενότυπος αναπαριστά μια πιθανή λύση σε ένα πρόβλημα και η αποκωδικοποίησή του μας δίνει τους φαινότυπους. Η **αναπαράσταση** (representation), λοιπόν, μας δίνει τη δυνατότητα να κάνουμε την χαρτογράφηση γενότυπων-φαινότυπων και να τις μελετήσουμε. Αξίζει να σημειωθεί πως ο χώρος των φαινότυπων μπορεί να διαφέρει από αυτού των γενότυπων και πως όλη η εξελικτική αναζήτηση γίνεται στον χώρο των γενότυπων. Μια (βέλτιστη) λύση, ή ένας καλός φαινότυπος, επιτυγχάνεται αποκωδικοποιώντας τον καλύτερο γενότυπο στις συνθήκες του προβλήματος μετά τον τερματισμό και αναπαρίσταται στον χώρο των γενότυπων.

Συνάρτηση αξιολόγησης

Ο ρόλος της **συνάρτησης αξιολόγησης** (evaluation function) είναι να παρουσιάζει το κατά πόσο ανταποκρίνεται ο πληθυσμός στις απαιτήσεις στις οποίες πρέπει να προσαρμοστεί ο Εξελικτικός Αλγόριθμος, τις οποίες έχει θέσει το πρόβλημα. Σχηματίζει μια σωστή βάση για τον τελεστή της επιλογής με απώτερο σκοπό να διευκολύνει τυχόν βελτιώσεις. Τεχνικά, είναι μια συνάρτηση ή διαδικασία που προσδίδει ένα μέτρο ποιότητας στους γενότυπους και τους δίνει μια τιμή ικανότητας (fitness value), δηλαδή καθορίζει την χρησιμότητα των διαφορετικών χαρακτηριστικών των υποψήφιων λύσεων στο πρόβλημα. Τυπικά, αυτή η συνάρτηση αποτελείται από ένα μέτρο ποιότητας στον χώρο των φαινότυπων και στην αντίστροφη αναπαράσταση. Στην Βελτιστοποίηση ισχύει ότι αν στόχος του προβλήματος είναι η ελαχιστοποίηση, είναι προτιμότερη μια λύση με μικρότερη τιμή ικανότητας ενώ το αντίθετο ισχύει όταν ο στόχος είναι η μεγιστοποίηση.

Στους Εξελικτικούς Αλγόριθμους η συνάρτηση αξιολόγησης καλείται κοινώς συνάρτηση ικανότητας (fitness function). Τέλος, θα πρέπει να σημειωθεί στην περίπτωση της βελτιστοποίησης, για το πρόβλημα που χρειάζεται να λυθεί από έναν Εξελικτικό Αλγόριθμο ο όρος «αντικειμενική συνάρτηση» χρησιμοποιείται στο αρχικό

πρόβλημα και ο όρος «συνάρτηση αξιολόγησης ή ικανότητας» είναι ένας μετασχηματισμός της.

Πληθυσμός

Ο πληθυσμός αποτελεί ένα σύνολο από αντίγραφα ενός γενότυπου-ατόμου που, όπως προαναφέραμε, αντιπροσωπεύει τις πιθανές λύσεις. Τα άτομα του πληθυσμού είναι δυναμικά αντικείμενα, αφού μεταβολή στον αριθμό της επανάληψης επιφέρει μεταβολή και στα άτομα καθαυτά. Το μέγεθος του είναι καθορισμένο από την αρχή και στις περισσότερες εφαρμογές εξελικτικών αλγορίθμων είναι σταθερό και δεν αλλάζει κατά την διάρκεια της εξελικτικής αναζήτησης.

Η ποικιλομορφία του πληθυσμού είναι ένα μέτρο του αριθμού των διαφορετικών λύσεων που παρουσιάζονται. Δηλαδή για κάθε ένα μέλος του πληθυσμού, θα μπορούσαμε να πούμε ότι έχουμε μια λύση του, προφανώς όχι αναγκαστικά ενδεδειγμένη. Να σημειωθεί ότι μόνο μια τιμή ικανότητας δεν δείχνει την παρουσία μόνο ενός φαινότυπου και ομοίως ένας φαινότυπος δεν δείχνει απαραίτητα έναν γενότυπο. Το αντίστροφο δεν ισχύει: ένας γενότυπος δείχνει μόνο έναν φαινότυπο και μια τιμή ικανότητας.

Μηχανισμός επιλογής γονέων

Ο ρόλος του **μηχανισμού επιλογής γονέων** (parent selection mechanism) είναι να ξεχωρίσει και να επιλέξει μέσα στον πληθυσμό άτομα εκείνα τα οποία κρίνει ως «ικανότερα», βάση των τιμών ικανότητάς τους, και να τους δώσει την δυνατότητα να γίνουν γονείς της επόμενης γενιάς. Βέβαια, στους Εξελικτικούς αλγορίθμους, η επιλογή εξαρτάται και από κάποια πιθανότητα. Έτσι, τα ικανότερα άτομα του πληθυσμού, δηλαδή αυτά την υψηλότερη συνάρτηση ικανότητας, έχουν μεγαλύτερη πιθανότητα αναπαραγωγής στην επόμενη γενιά ενώ τα λιγότερα ικανά έχουν εξίσου μεγάλη πιθανότητα να αντικατασταθούν από νέα. Μαζί με τον μηχανισμό επιλογής επιβίωσης είναι υπεύθυνοι για τις βελτιώσεις της ποιότητας των λύσεων. Σε αντίθεση με τους τελεστές διασταύρωσης και μετάλλαξης που δρουν σε ένα ή δύο άτομα-γονείς, οι μηχανισμοί επιλογής γονέων δρουν σε όλα άτομα του πληθυσμού.

Τελεστές ανασυνδυασμού (ή διασταύρωσης) και μετάλλαξης

Πρόκειται για τους τελεστές οι οποίοι είναι υπεύθυνοι για τον ανασυνδυασμό και την μετάλλαξη που εφαρμόζονται στον πληθυσμό των ατόμων που έχουν επιλεγεί για

αναπαραγωγή, δηλαδή των γονέων, και που δημιουργείται από τον μηχανισμό επιλογής γονέων και επιτυγχάνουν την δημιουργία των απογόνων ή, σε όρους προβλημάτων, των νέων λύσεων.

Ο **τελεστής της μετάλλαξης** έχει μία παράμετρο εισόδου και για αυτό είναι μοναδιαίος. Πιο συγκεκριμένα, εφαρμόζεται σε έναν γενότυπο και δίνει μια μετάλλαξη του, το παιδί ή αλλιώς τον απόγονο. Επίσης, ισχύει ότι είναι στοχαστικός αφού οι εκάστοτε απόγονοι εξαρτώνται από τις συνέπειες μιας σειράς τυχαίων επιλογών.

Εν αντιθέσει με τον τελεστή της μετάλλαξης, ο **τελεστής του ανασυνδυασμού** ή αλλιώς της διασταύρωσης είναι δυαδικός, αφού κατά την διάρκεια της λειτουργίας του διασταυρώνει τις πληροφορίες (χαρακτηριστικά) από δύο γονείς-γενότυπους ούτως ώστε να τις διαβιβάσει σε ένα ή δύο παιδιά-φαινότυπους. Βέβαια, ο ανασυνδυασμός είναι και αυτός στοχαστικός τελεστής αφού η επιλογή των πληροφοριών του κάθε γονέα που θα διασταυρωθούν καθώς και ο τρόπος με τον οποίο θα γίνει αυτό είναι τυχαία γεγονότα. Εδώ αξίζει να αναφέρουμε ότι διασταυρώσεις που χρησιμοποιούν πάνω από δύο άτομα του πληθυσμού ως γονείς είναι μαθηματικά εφικτοί, αλλά δεν προσβλέπονται από την βιολογία και πιθανόν γι' αυτό δεν είναι τόσο συχνοί σε χρήση, αν και πολλές μελέτες έχουν δείξει ότι έχουν θετικές συνέπειες στην εξέλιξη.

Ο τελεστής του ανασυνδυασμού λειτουργεί ως εξής: «ζευγαρώνοντας» δυο άτομα με διαφορετικά χαρακτηριστικά δημιουργούμε απογόνους με χαρακτηριστικά και των δύο. Έτσι, τα χαρακτηριστικά αυτά κάποιες φορές μπορεί να μην είναι τα επιθυμητά αλλά οι εξελικτικοί αλγόριθμοι μέσω των διαδικασιών που ακολουθούν, προσπαθούν με την πάροδο του χρόνου να μετριάσουν τέτοιου είδους αποτελέσματα. Βέβαια, σε αντίθεση με ότι συμβαίνει στον κόσμο της βιολογίας όπου η διασταύρωση είναι απαραίτητη για την αναπαραγωγή, εντούτοις σε κάποιους εξελικτικούς αλγορίθμους υπάρχει κάποια πιθανότητα να μην εφαρμοστεί.

Μηχανισμός επιλογής επιβίωσης

Ο ρόλος του **μηχανισμού επιλογής επιβίωσης** (survivor selection mechanism), παρομοίως με τον μηχανισμό επιλογής γονέων, είναι να ξεχωρίσει μέσα στον πληθυσμό τα ικανότερα άτομα με βάση τις τιμές ικανότητάς τους. Ο μηχανισμός επιλογής επιβίωσης, βέβαια, δεν καλείται παρά μόνο όταν έχουν δημιουργηθεί οι απόγονοι των επιλεγμένων γονέων.

Εφόσον το μέγεθος του πληθυσμού στους Εξελικτικούς Αλγορίθμους είναι σταθερό, η επιλογή πρέπει να γίνει με βάση το κατά πόσο αξίζουν ή όχι τα άτομά του να περάσουν στην επόμενη γενιά, δηλαδή να επιβιώσουν. Αυτή η απόφαση συνήθως βασίζεται στην τιμή ικανότητάς τους ή στην ηλικία τους. Συγκεκριμένα, ο μηχανισμός αυτός βαθμολογεί και κατατάσσει τους υπονήφιους γονείς και απόγονους και, με βάση την ηλικία και την ικανότητά τους, επιλέγει τους πιο κατάλληλους. Συχνά ο όρος αυτός καλείται και **αντικατάσταση** (replacement) αφού οι νέοι και καλύτεροι απόγονοι πολλές φορές δίνουν την θέση τους σε κάποιους καλύτερους.

Κριτήρια τερματισμού

Τα κριτήρια τερματισμού διαφέρουν ανά πρόβλημα και συχνά έχουν να κάνουν με την φύση τους αλλά και με το «μέτρο» προσέγγισης που αυτός που επιδιώκει να λύσει το πρόβλημα θεωρεί επαρκές. Έτσι, δημιουργούνται πολλά διαφορετικά κριτήρια που ο καθένας μπορεί να χρησιμοποιήσει. Για παράδειγμα στο πρόβλημα που αφορά επίπεδο βέλτιστης ικανότητας, πιθανόν από ένα γνωστό βέλτιστο μιας αντικειμενικής συνάρτησης, τότε ένα μέτρο διαφοράς της λύσης να μπορεί να είναι αρκετά καλή επιλογή. Παρόλα αυτά, λόγω της στοχαστικότητας της φύσης τους οι Εξελικτικοί Αλγόριθμοι δεν εγγυώνται ότι μπορούν πάντα να φτάσουν στο βέλτιστο αφού η κατανομή του αρχικού πληθυσμού είναι τυχαίο γεγονός και έτσι το κριτήριο αυτό μπορεί να μην ικανοποιηθεί και ο αλγόριθμος να μην σταματήσει ποτέ. Για αυτόν το λόγο αυτό μπορεί να χρησιμοποιηθούν και άλλα κριτήρια όπως είναι ο μέγιστος αριθμός γενεών ή η σύγκλιση σε μία λύση και το χρονικό διάστημα τρεξίματος του αλγορίθμου.

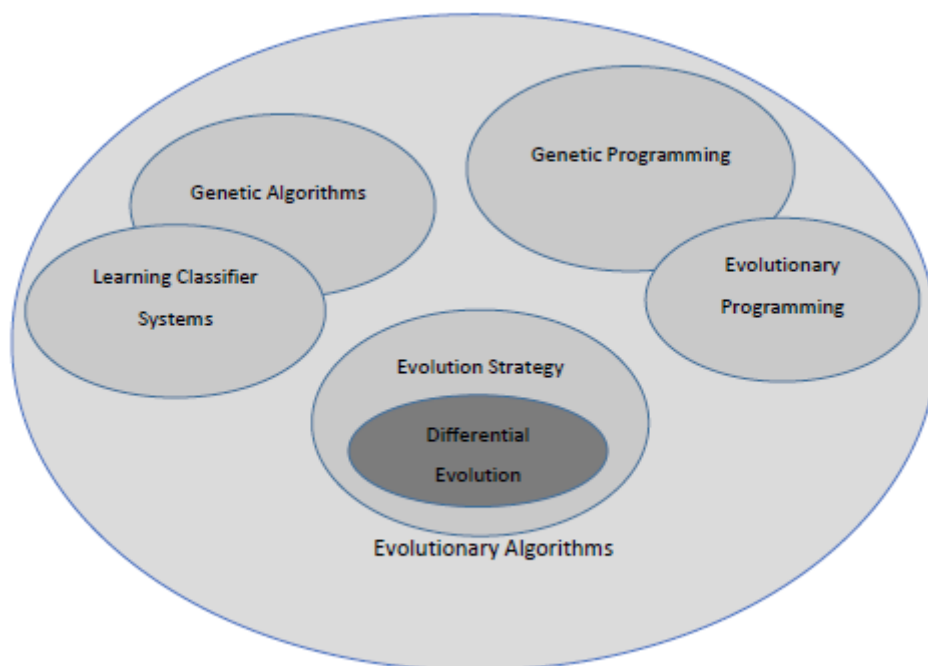
1.4 Οικογένεια των Εξελικτικών Αλγορίθμων

Οι Εξελικτικοί Αλγόριθμοι (ΕΑ) ή αλλιώς του Evolutionary Computation (EC) αποτελούν υπερσύνολο άλλων πέντε ειδών αλγορίθμων [18]. Παρόλο που η αρχική έρευνα που αφορά τον EC χρονολογείται στις δεκαετίες του '50 και του '60 (Bremermann), η πραγματική ανάπτυξη του όρου συντελέστηκε την δεκαετία του '90 με στόχο την συνένωση των 28 προσπαθειών σε πέντε κυρίως τομείς, οι οποίοι είναι:

1. Γενετικοί Αλγόριθμοι-ΓΑ (Genetic Algorithms-J.H.Holland) [10]
2. Γενετικός Προγραμματισμός-ΓΠ (Genetic Programming-J.R.Koza) [13]
3. Εξελικτικός Προγραμματισμός-ΕΠ (Evolutionary Programming- L.J.Fogel D.B.Fogel) [5]

4. Εξελικτική Στρατηγική-ΕΣ (Evolution Strategy- I.Rechenberg- H.P.Schwefel) [14]

5. Μανθάνοντα Συστήματα Ταξινομητών-ΜΣΤ (Learning Classifier Systems)



Σχήμα 3: Πηγή: ΕΞΕΛΙΚΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ: Ο ΑΛΓΟΡΙΘΜΟΣ ΤΗΣ ΔΙΑΦΟΡΙΚΗΣ ΕΞΕΛΙΞΗΣ, Α. Γκουντή, Η οικογένεια των εξελικτικών αλγορίθμων

1.4.1 Γενετικοί Αλγόριθμοι (Genetic Algorithms)

Η βασική ιδέα που κρύβεται πίσω από τους Γενετικούς Αλγόριθμους είναι η μίμηση των μηχανισμών της φύσης. Χαρακτηριστικό παράδειγμα, όπως αναφέρει ο Λυκοθανάσης [21] αποτελούν οι λαγοί και πώς αναπαράγονται και εξελίσσονται από γενιά σε γενιά. Έστω ότι έχουμε έναν συγκεκριμένο πληθυσμό από λαγούς. Όπως είναι φυσικό, πολλοί ή και όλοι θα εμφανίζουν διαφορετικά χαρακτηριστικά. Για παράδειγμα, κάποιος από αυτούς θα είναι πιο γρήγορος και πιο εύστροφος από τους άλλους. Αυτοί οι γρηγορότεροι και εξυπνότεροι λαγοί έχουν λιγότερες πιθανότητες να αποτελέσουν γεύμα κάποιας αλεπούς και, άρα από τη στιγμή που καταφέρνουν να επιβιώσουν θα ασχοληθούν με την αναπαραγωγή του είδους τους. Φυσικά, θα υπάρχει και ένας μικρός αριθμός αργών και λιγότερο εύστροφων λαγών, που θα καταφέρουν να επιβιώσουν επειδή στάθηκαν σχετικά τυχεροί. Όλοι αυτοί οι λαγοί, που έχουν καταφέρει να επιβιώσουν, θα αρχίσουν την παραγωγή της επόμενης γενιάς τους, μιας γενιάς που όπως όλες θα συνδυάζει όλα τα χαρακτηριστικά των μελών της προηγούμενης, συνδυασμένα με διάφορους τρόπους μεταξύ τους. Έτσι, μερικοί αργοί

λαγοί θα αναμειχθούν με κάποιους γρήγορους, κάποιιοι γρήγοροι με άλλους γρήγορους, κάποιιοι εύστροφοι λαγοί με κάποιους μη εύστροφους και ούτω καθεξής. Βέβαια όπως είναι λογικό, αναλογικά με την προηγούμενη χρονιά, ο αριθμός των λαγών που δεν ήταν έξυπνοι και γρήγοροι, θα καλύπτει μικρότερο ποσοστό του πληθυσμού. Έτσι, οι μικροί λαγοί της επόμενης γενιάς θα είναι, κατά μέσο όρο, γρηγορότεροι και έξυπνότεροι από τους προγόνους τους, αφού από την προηγούμενη γενιά επιβίωσαν περισσότεροι γρήγοροι και έξυπνοι λαγοί. Ευτυχώς, για την διατήρηση της φυσικής ισορροπίας, και οι αλεπούδες υφίστανται την ίδια διαδικασία αναπαραγωγής, διαφορετικά οι λαγοί θα γινόντουσαν υπερβολικά γρήγοροι και έξυπνοι για να μπορούν να τους πιάσουν.

Όπως τονίσαμε παραπάνω, οι Γενετικοί Αλγόριθμοι χρησιμοποιούν ορολογία δανεισμένη από το χώρο της φυσικής Γενετικής όπως ακριβώς και οι Εξελικτικοί που είναι υπερσύνολό τους. Συγκεκριμένα, συνοψίζοντας ουσιαστικά την θεωρία του Δαρβίνου που αναλύσαμε στο Κεφάλαιο 1.1 έχουμε ότι τα χρωμοσώματα αποτελούνται από διάφορα στοιχεία που ονομάζονται γονίδια και είναι διατεταγμένα σε γραμμική ακολουθία. Τα γονίδια είναι υπεύθυνα των χαρακτηριστικών του καθενός.

Κάθε **γονότυπος** (που στις περισσότερες περιπτώσεις είναι ένα μόνο χρωμόσωμα) αναπαριστά μια πιθανή λύση σε ένα πρόβλημα. Το μεταφρασμένο περιεχόμενο ενός συγκεκριμένου χρωμοσώματος καλείται φαινότυπος και καθορίζεται από τον χρήστη, ανάλογα με τις ανάγκες και τις απαιτήσεις του.

Μια διαδικασία εξέλιξης που εφαρμόζεται πάνω σε έναν πληθυσμό χρωμοσωμάτων αντιστοιχεί σε ένα εκτενές ψάξιμο μέσα σε ένα χώρο από πιθανές λύσεις. Απαραίτητη προϋπόθεση για την επιτυχημένη έκβαση μια τέτοιας αναζήτησης αποτελεί η εξισορρόπηση δύο διαδικασιών που είναι προφανώς αντικρουόμενες, της εκμετάλλευσης και διατήρησης των καλύτερων λύσεων και της όσο το δυνατόν καλύτερης εξερεύνησης όλου του διαστήματος.

Λόγω της ευρείας χρήσης των Γενετικών Αλγορίθμων στον τομέα της βελτιστοποίησης πολλοί άνθρωποι έχουν την εντύπωση ότι οι Γενετικοί Αλγόριθμοι είναι αποκλειστικά αλγόριθμοι βελτιστοποίησης. Πράγματι, όπως συμβαίνει και την περίπτωση που θα μελετήσουμε, η βελτιστοποίηση είναι μεγάλο μέρος των εφαρμογών τους. Γενικά όμως, οι Γενετικοί Αλγόριθμοι πρέπει να αντιμετωπίζονται και ως μια ιδεατή

προσομοίωση μιας φυσικής διαδικασίας, τέτοια ώστε να ενσωματώνει τους στόχους και τους σκοπούς της διαδικασίας αυτής, προσαρμοσμένα ώστε να επιλύουν τα προβλήματα που έχουν τεθεί.

Όπως προαναφέραμε, κατά τη διάρκεια της τελευταίας δεκαετίας το ενδιαφέρον για τις διαδικασίες βελτιστοποίησης έχει αυξηθεί τόσο πολύ, ώστε συνεχώς δημιουργόντουσαν καινούργια, πιο πολύπλοκα και με πολύ αυστηρούς περιορισμούς προβλήματα, που να μπορούν να επιλυθούν μόνο προσεγγιστικά από τους σημερινούς υπολογιστές. Οι Γενετικοί Αλγόριθμοι αποσκοπούν στην εξυπηρέτηση τέτοιου είδους προβλημάτων και αποτελούν ένα από τα σημαντικότερα εργαλεία στην μελέτη τους. Παρόλο που ανήκουν στην κατηγορία των στοχαστικών αλγορίθμων, στους οποίους η αρχική κατανομή είναι ένα τυχαίο γεγονός, διαφέρουν σε πολύ μεγάλο βαθμό από τους αλγόριθμους που εφαρμόζουν τυχαίες μεθόδους αναζήτησης και βελτιστοποίησης, αφού είναι σε θέση να συνδυάζουν τα κατάλληλα στοιχεία και από άμεσες και από στοχαστικές τεχνικές αναζήτησης προκειμένου να πετύχουν το βέλτιστο αποτέλεσμα. Προφανώς, η φύση της στοχαστικότητας, και ιδιαίτερα η τυχαιότητα, δεν μπορεί να εξαλειφθεί τελείως αλλά οι Γενετικοί Αλγόριθμοι καταφέρνουν να το μετριάσουν κατά ένα μεγάλο ποσοστό. Αυτός είναι και ο κύριος λόγος για τον οποίο οι ΓΑ θεωρούνται πιο εύρωστοι από τις υπάρχουσες μεθόδους άμεσης αναζήτησης. Ένα άλλο εξίσου σημαντικό χαρακτηριστικό τους είναι ότι διατηρούν έναν πληθυσμό πιθανών λύσεων πάνω στον οποίο πειραματίζονται, σε αντίθεση με όλες τις άλλες μεθόδους αναζήτησης που επεξεργάζονται ένα μόνο σημείο του διαστήματος αναζήτησης.

Ένας από τους βασικούς κανόνες λειτουργίας ενός Γενετικού Αλγορίθμου είναι ότι πραγματοποιεί αναζήτηση σε πολλές κατευθύνσεις με το να διατηρεί έναν πληθυσμό από πιθανές λύσεις και να τον μεταλλάσει μέσω των διαφόρων διαδικασιών που πραγματοποιεί. Ουσιαστικά, πληθυσμός υφίσταται μια προσομοιωμένη γενετική εξέλιξη. Σε κάθε γενιά, οι ικανοποιητικές λύσεις αναπαράγονται, ενώ οι μη ικανοποιητικές αφαιρούνται. Ο διαχωρισμός και η αξιολόγηση των διαφόρων λύσεων γίνεται με την βοήθεια μιας αντικειμενικής συνάρτησης ή συνάρτησης ικανότητας (objective ή fitness function), η οποία παίζει το ρόλο του περιβάλλοντος μέσα στο οποίο εξελίσσεται ο πληθυσμός. Επίσης στη βιβλιογραφία αναφέρεται και ως συνάρτηση αξιολόγησης και συνάρτηση καταλληλότητας.

Η δομή ενός απλού γενετικού αλγορίθμου, σύμφωνα με τον Λυκοθανάση [21], περιγράφεται ως εξής:

1. Κατά την διάρκεια της επαναληπτικής εκτέλεσης t ενός αλγορίθμου, διατηρεί ένα πληθυσμό από πιθανές λύσεις:

$$P(t) = \{x_1^t, \dots, x_n^t\}$$

2. Γίνεται αξιολόγηση κάθε λύσης μέσω μίας συνάρτησης αξιολόγησης η οποία δίνει ένα μέτρο της καταλληλότητάς της και αποφασίζουμε κατά πόσο δεχόμαστε κάποια λύση του πληθυσμού μέσω ενός κριτηρίου τερματισμού. Σε περίπτωση που δεν είμαστε ικανοποιημένοι, αφού ολοκληρωθεί η αξιολόγηση όλων των στοιχείων του πληθυσμού, έχουμε την δημιουργία ενός νέου πληθυσμού στην επόμενη εκτέλεση, που προκύπτει από την επιλογή των πιο κατάλληλων στοιχείων του πληθυσμού της προηγούμενης γενιάς τα οποία και υφίστανται μετατροπές μέσω των διαδικασιών της **μετάλλαξης** (mutation) και της **διασταύρωσης** (crossover ή mating) σχηματίζοντας νέες πιθανές λύσεις. Μέσω της διασταύρωσης έχουμε τον συνδυασμό στοιχείων δύο χρωμοσωμάτων των γονέων για την δημιουργία απογόνων ανταλλάσσοντας αντίστοιχα κομμάτια από τους γονείς. Αντίστοιχα, η μετάλλαξη αλλάζει αυθαίρετα ένα ή περισσότερα γονίδια ενός συγκεκριμένου χρωμοσώματος μέσω μιας τυχαίας αλλαγής γονιδίων και με πιθανότητα ίση με το **ρυθμό μετάλλαξης** (mutation rate). Έτσι, θα λέγαμε ότι μπορούν να δημιουργηθούν λύσεις που δεν είναι πολύ σχετικές με τον υπάρχον πληθυσμό.

Ένας Γενετικός Αλγόριθμος για ένα συγκεκριμένο πρόβλημα πρέπει να αποτελείται από τα παρακάτω πέντε τμήματα [20]:

1. Μια γενετική αναπαράσταση των πιθανών λύσεων του προβλήματος.
2. Την δημιουργία ενός αρχικού πληθυσμού των πιθανών λύσεων. Αυτός, όπως ακριβώς στους Εξελικτικούς, μπορεί να μην είναι πάντα τυχαίος, ανάλογα με την φύση και την πολυπλοκότητα του προβλήματος
3. Μια αντικειμενική συνάρτηση αξιολόγησης που αποσκοπεί στο να κατατάσσει τις λύσεις με βάση την καταλληλότητά τους.
4. Γενετικούς τελεστές που εμπλέκονται στην διαδικασία παραγωγής των παιδιών.

5. Τιμές για διάφορες παραμέτρους που χρησιμοποιεί ο γενετικός αλγόριθμος (μέγεθος πληθυσμού κ.τ.λ.).

Πλεονεκτήματα των Γενετικών Αλγορίθμων

Όπως μας ενημερώνει ο Λυκοθανάσης [20], με το πέρασμα του χρόνου, η χρήση των Γενετικών Αλγορίθμων έχει αυξηθεί ραγδαία και παρατηρείται σε διάφορες εφαρμογές. Αυτό συμβαίνει, κυρίως, επειδή:

1. Μπορούν να λύσουν δύσκολα προβλήματα με μεγάλο δείκτη αξιοπιστίας και με μεγάλη ταχύτητα. Ένας από τους σημαντικούς λόγους χρήσης των Γενετικών Αλγορίθμων είναι η μεγάλη τους αποδοτικότητα. Αυτή η αποδοτικότητα είναι ευδιάκριτη από το πλήθος των δύσκολων προβλημάτων που έχουν επιλυθεί μέσω αυτών και από το γεγονός ότι οι Γενετικοί Αλγόριθμοι δεν φαίνεται να δυσκολεύονται σε σημεία προβλημάτων που για άλλες μέθοδοι παρουσιάζουν τεράστιες δυσκολίες.

2. Μπορούν να συνεργαστούν με πολλά διαφορετικά μοντέλα και συστήματα. Οι Γενετικοί Αλγόριθμοι προσφέρουν το σημαντικό πλεονέκτημα της χρήσης τους με προσθετικό τρόπο στα μοντέλα που χρησιμοποιούνται σήμερα, μη απαιτώντας την επανασχεδιάσή τους. Μπορούν εύκολα να συνεργαστούν με τον υπάρχοντα κώδικα, χωρίς μεγάλο κόπο. Αυτό συμβαίνει, διότι χρησιμοποιούν μόνο πληροφορίες της διαδικασίας ή συνάρτησης που πρόκειται να βελτιστοποιήσουν, δίχως να ενδιαφέρει άμεσα ο ρόλος της μέσα στο σύστημα ή η όλη δομή του συστήματος.

3. Μπορούν εύκολα να επεκταθούν και να εξελιχθούν. Οι Γενετικοί Αλγόριθμοι δεν αντιστέκονται σε αλλαγές, επεκτάσεις και μετεξελίξεις, ανάλογα με την κρίση του σχεδιαστή αφού η φύση τους τους εξασφαλίζει τέτοια προνόμια. Σε πολλές εφαρμογές, μια μικρή παραλλαγή στον Γενετικό Αλγόριθμο, που παρακούει τις ιδιότητες της φύσης που ακολουθεί δεν είναι απλά ανεκτή, αλλά ίσως κιόλας επιβάλλεται.

4. Μπορούν να συμμετέχουν σε υβριδικές μορφές με άλλες μεθόδους ως υποστηρικτικές. Δηλαδή, σε κάποιες περιπτώσεις που ένας αλγόριθμος έχει πιο πολύ ισχύ πάνω σε ένα πρόβλημα τότε μπορεί ένας Γενετικός Αλγόριθμος να χρησιμοποιηθεί σε ένα υβριδικό σχήμα με τον υπάρχον.

5. Βρίσκουν εφαρμογή σε πολύ περισσότερα πεδία από κάθε άλλη μέθοδο. Το χαρακτηριστικό, που τους εξασφαλίζει αυτό το πλεονέκτημα, είναι η ελευθερία

επιλογής των κριτηρίων που καθορίζουν την επιλογή μέσα στο τεχνητό περιβάλλον με εφαρμογές όπως π.χ. στην οικονομία και στην ψυχολογία.

6. Το πεδίο αναζήτησης είναι πολύ ευρύ αφού η αναζήτηση κατά την διάρκεια ενός Γενετικού Αλγορίθμου γίνεται σε πολλά σημεία ταυτόχρονα και όχι μόνο σε ένα. Σε πολλές μεθόδους βελτιστοποίησης, ισχύει ότι η επεξεργασία γίνεται βήμα προς βήμα, πηγαίνοντας προσεχτικά από σημείο σε σημείο του πεδίου ορισμού του προβλήματος γεγονός που κρίνεται πολύ πιο «χρονοβόρο» από ότι στους Γενετικούς Αλγόριθμους. Ουσιαστικά οι γενετικοί αλγόριθμοι ενεργούν ταυτόχρονα πάνω σε ένα ευρύ σύνολο σημείων ξεκινώντας την αναζήτησή τους από ένα πληθυσμό και κατόπιν παράγουν διαδοχικά καινούργιους πληθυσμούς. Έπειτα, τρέχοντας ο αλγόριθμος δημιουργεί νέους πληθυσμούς που συγκλίνουν προς την επιθυμητή λύση.

1.4.2 Στρατηγικές Εξέλιξης

Σύμφωνα με την scholarpedia [11] οι στρατηγικές εξέλιξης (ES) αποτελούν μια υποκατηγορία μεθόδων άμεσης αναζήτησης (και βελτιστοποίησης) που εμπλέκονται στην φύση και ανήκουν στην τάξη των εξελικτικών αλγορίθμων. Μάλιστα όπως τονίζουν και οι Salimans, Ho, Chen, Sidor και Sutskever [15] οι στρατηγικές εξέλιξης είναι της κλάσης των αλγορίθμων βελτιστοποίησης Black Box (βλέπε Κεφ.2.3) ενώ ο πιο γνωστός αλγόριθμος της οικογένειας αυτής είναι η CMA, την οποία και θα μελετήσουμε. Η 1^η εμφάνισή των στρατηγικών εξέλιξης χρονολογείται στα μέσα της δεκαετίας του 1960 όταν οι P. Bienert, I. Rechenberg και H.-P. O Schwefel στο Πολυτεχνείο του Βερολίνου ανέπτυξαν τα πρώτα βιονικά-εμπνευσμένα συστήματα για την εξέλιξη των βέλτιστων σχημάτων των ελαχίστων σωμάτων σύρματος σε αεροδυναμική σήραγγα χρησιμοποιώντας την αρχή εξέλιξης του Δαρβίνου.

Οι στρατηγικές εξέλιξης μπορούν να εφαρμοστούν σε όλους τους τομείς βελτιστοποίησης, συμπεριλαμβανομένων συνεχών, διακριτών, συνδυαστικών χώρων αναζήτησης Y χωρίς και με περιορισμούς καθώς και με μικτούς χώρους αναζήτησης παρέχοντας μας έτσι πολλά πλεονεκτήματα με την χρήση τους. Με βάση τον Beyer [3] δεδομένου του προβλήματος βελτιστοποίησης

$$y^* = \operatorname{argopt}_{y \in Y} f(y),$$

η συνάρτηση $f(y)$ που πρέπει να βελτιστοποιηθεί, που αναφέρεται και ως αντικειμενική συνάρτηση (ή στόχος), μπορεί να παρουσιαστεί σε μαθηματική μορφή,

μέσω προσομοιώσεων ή ακόμα και σε όρους μετρήσεων που λαμβάνονται από πραγματικά αντικείμενα. Οι στρατηγικές εξέλιξης μπορούν επίσης να εφαρμοστούν σε ένα σύνολο αντικειμενικών λειτουργιών στο πλαίσιο βελτιστοποίησης πολλαπλών αντικειμένων (βλ. Επίσης Πολυπολιτισμικούς Εξελικτικούς Αλγορίθμους και Πολυπαραγοντική Αναζήτηση). Μια σημαντική εκδοχή τους αποτελούν η κανονική.

Canonical ES Versions

Οι κανονικές εκδοχές των στρατηγικών εξέλιξης (ES) υποδηλώνονται με

$$(\mu / \rho, \lambda) - \text{ES} \text{ και } (\mu / \rho + \lambda) - \text{ES}, \text{ αντίστοιχα}$$

όπου

μ ο αριθμός των γονέων,

$\rho \leq \mu$ ο αριθμός ανάμιξης (δηλ. Ο αριθμός των γονέων που εμπλέκονται στην αναπαραγωγή ενός απογόνου)

λ ο αριθμός των απογόνων

Οι γονείς είναι έχουν επιλεγεί από το (πολλαπλάσιο) σύνολο είτε του απογόνου, που αναφέρεται ως επιλογή-κόμμα (πρέπει να ισχύει $\mu < \lambda$), είτε οι γονείς και οι απόγονοι, αμφότεροι επιλέγονται από την επιπλέον-επιλογή. Η επιλογή βασίζεται στην κατάταξη της fitness function, $F(y)$, των ατόμων λαμβάνοντας τα μ καλύτερα άτομα (αναφέρεται επίσης και ως επιλογή περικοπής). Σε γενικές γραμμές, ένα άτομο

$$a := (y, s, F(y))$$

περιλαμβάνει το διάνυσμα της παραμέτρου ενός αντικειμένου $y \in Y$ που πρέπει να βελτιστοποιηθεί, ένα σύνολο παραμέτρων στρατηγικής s , που απαιτούνται ειδικά σε αυτοπροσδιοριζόμενες στρατηγικές εξέλιξης, και την αντικειμενική συνάρτηση $F(y)$ που είναι ισοδύναμη με την αντικειμενική συνάρτηση $f(y)$, δηλαδή $F(y) \equiv f(y)$ το οποίο ισχύει στην απλούστερη περίπτωση. Η διάκριση μεταξύ $F(y)$ και $f(y)$ είναι απαραίτητη, αφού το $F(y)$ μπορεί να είναι το αποτέλεσμα μιας τοπικής αναζήτησης που εφαρμόζεται στην $f(y)$ που πρέπει να βελτιστοποιηθεί ή ακόμα και μπορεί να είναι αποτέλεσμα μιας άλλης στρατηγικής εξέλιξης (όπως η Meta-ES). Επιπλέον, το παρατηρούμενο $F(y)$ μπορεί να είναι αποτέλεσμα μιας θορυβώδους διαδικασίας αξιολόγησης $f(y)$.

Ο εννοιολογικός αλγόριθμος της στρατηγικής εξέλιξης- $\mu / \rho, \lambda$) σύμφωνα με τον Beyer [3], δίνεται παρακάτω:

($\mu / \rho, \lambda$) -Αναπροσαρμογή-Εξέλιξη-Στρατηγική του οποίου τα βήματα είναι:

- 1) Αρχικοποίηση του πληθυσμού των γονέων $P_\mu = \{a_1, \dots, a_\mu\}$
- 2) Δημιουργία του λ απογόνους \tilde{a} για τον σχηματισμό του πληθυσμού των απογόνων $\tilde{P}_\mu = \{\tilde{a}_1, \dots, \tilde{a}_\lambda\}$ όπου κάθε απόγονος \tilde{a} δημιουργείται από κάποιον όταν:
 - i) Επιλέξει (τυχαία) ρ γονείς από P_μ (αν $\rho = \mu$ πρέπει να πάρει όλα τα γονικά άτομα αντ' αυτού).
 - ii) Ανασυνδιάσει όλους τους ρ επιλεγμένους γονείς a ούτως ώστε να σχηματίσουν ένα ανασυνδυασμένο άτομο r .
 - iii) Μεταλλάξει τα σύνολα των παραμέτρων στρατηγικής s του ανασυνδυασμένου r .
 - iv) Μεταλλάξει το σύνολο αντικειμενικών παραμέτρων y του ανασυνδυασμένου r χρησιμοποιώντας τη μεταβλητή παράμετρο στρατηγικής που είναι υπεύθυνος για να γίνει ο έλεγχος των στατιστικές ιδιοτήτων της μετάλλαξης της παραμέτρου ενός αντικειμένου.
- 3) Επιλογή νέου γονικού πληθυσμού (χρησιμοποιώντας τη ντετερμινιστική επιλογή περικοπής) από είτε
 - i) τον πληθυσμό των απογόνων \tilde{P}_λ , που αναφέρεται ως επιλογή-κόμμα, που συνήθως χαρακτηρίζεται ως " (μ, λ) -επιλογή", ή
 - ii) τον απόγονο \tilde{P}_λ και τον γονικό πληθυσμό P_μ (αυτό αναφέρεται ως συν-επιλογή, που συνήθως χαρακτηρίζεται ως " $(\mu + \lambda)$ -επιλογή"
- 4) Επιστροφή στο βήμα 2 μέχρι να ολοκληρωθεί το κριτήριο τερματισμού

Ανάλογα με τον χώρο αναζήτησης και την αντικειμενική συνάρτηση $f(y)$, ο ανασυνδυασμός και / ή η μετάλλαξη των παραμέτρων της στρατηγικής μπορεί είτε να συμβαίνουν είτε όχι σε συγκεκριμένες παραλλαγές του αλγορίθμου. Για παράδειγμα, μια στρατηγική εξέλιξης ($\mu / \rho, \lambda$), ή ισοδύναμα ($\mu / \rho, \lambda$)-ES, δεν χρησιμοποιεί ανασυνδυασμό. Παράγει τους νέους γονείς μ για την επόμενη γενιά τόσο από τους παλιούς γονείς μ όσο και από τους απογόνους λ (που παράγονται από αυτούς τους γονείς) λαμβάνοντας τα καλύτερα μ άτομα (σε σχέση με τον παρατηρούμενο $F(y)$).

Οι στρατηγικές εξέλιξης τύπου $(\mu / \rho, \lambda)$ αναφέρονται επίσης ως ESs σταθερής κατάστασης, δηλ. στρατηγικές χωρίς κενό παραγωγής: παράγουν μόνο έναν απόγονο ανά γενιά. Μετά την αξιολόγηση της ικανότητάς του $F(y)$, το χειρότερο άτομο αφαιρείται από τον πληθυσμό. Οι στρατηγικές αυτού του τύπου είναι ιδιαίτερα χρήσιμες σε παράλληλους υπολογιστές, όταν οι χρόνοι υπολογισμού των ικανοτήτων των ατόμων είναι μη σταθεροί, επιτρέποντας έτσι την ασύγχρονη παράλληλη επεξεργασία.

Όσον αφορά τη συνθήκη τερματισμού, μπορούν να χρησιμοποιηθούν μέτρα απόστασης στην fitness function ή στον χώρο αναζήτησης, καθώς και ο μέγιστος αριθμός γενεών.

Κεφάλαιο 2: Μαθηματικό Υπόβαθρο

Αφού, λοιπόν, αναλύσαμε τους γενετικούς αλγόριθμους, θα επιχειρήσουμε τώρα να αναλύσουμε το μαθηματικό υπόβαθρο που χρειάζεται κάποιος να κατέχει ούτως ώστε να μπορεί να κατανοήσει σε βάθος την λειτουργία του αλγόριθμου που θα μελετήσουμε. Έτσι, θα μπορέσουμε να μελετήσουμε τον αλγόριθμο ενδελεχώς στο επόμενο κεφάλαιο. Για την συγγραφή των κεφαλαίων αυτών χρησιμοποιήθηκαν πολλές αναφορές από συγγράμματα του Hansen [7][8][9], ο οποίος και μελέτησε εις βάθος την CMA.

Συμβολισμοί-Ορολογία

EMNA Estimation Matrix Adaptation

λ αριθμός πληθυσμού. Ισχύει ότι $\lambda \geq 2$

μ αριθμός των γονέων. Ισχύει ότι $\mu \leq \lambda$

$\mu_{eff} = (\sum_{t=1}^{\mu} w_t^2)^{-1}$, η αποτελεσματική επιλογή μαζών για την μέση τιμή

$\sum w_j = \sum_{t=1}^{\mu} w_t$, το άθροισμα όλων των βαρών. Ισχύει ότι $w_i \leq 0$ για κάθε $i > \mu$

$\sum |w_t|^+ = \sum_{t=1}^{\mu} w_t = 1$, το άθροισμα των θετικών βαρών

$\sum |w_t|^- = -(\sum w_j - \sum |w_t|^+) = -\sum_{t=\mu+1}^{\lambda} w_t \geq 0$, το άθροισμα των αρνητικών βαρών

$B \in \mathbb{R}^n$ ένας ορθοκανονικός πίνακας του οποίου οι στήλες του είναι ιδιοδιανύσματα του C και ο οποίος έχει αριθμό γραμμών ίσο με το νούμερο των διαγώνιων στοιχείων του D

$C^{(g)} \in \mathbb{R}^{n \times n}$, ο πίνακας συνδιακύμανσης για την γενιά g

c_{tt} , τα διαγώνια στοιχεία του C

$c_c \leq 1$, ο ρυθμός εκμάθησης για το cumulation για την αναβάθμιση 1-βαθμού του πίνακα συνδιακύμανσης

$c_1 \leq 1 - c_{\mu}$, ο ρυθμός εκμάθησης για την αναβάθμιση 1-βαθμού του πίνακα συνδιακύμανσης

$c_{\mu} \leq 1 - c_1$, ο ρυθμός εκμάθησης για την αναβάθμιση μ -βαθμού του πίνακα συνδιακύμανσης

$c_{\sigma} < 1$, ο ρυθμός εκμάθησης για το cumulation για τον έλεγχο βήματος

$D \in \mathbb{R}^n$ είναι ένας διαγώνιος πίνακας του οποίου τα στοιχεία είναι οι ρίζες των ιδιοτιμών του πίνακα συνδιακύμανσης C και ο οποίος έχει στοιχεία τα d_t

$d_\sigma \approx 1$, παράμετρος απόσβεσης για αναβάθμιση μεγέθους βημάτων,

$$f \in \mathbb{R}^n \mapsto \mathbb{R}$$

$$f_{sphere}: \mathbb{R}^n \mapsto \mathbb{R}, x \mapsto \|x\|^2 = x^T x = \sum_{t=1}^n x_t^2$$

$g \in N_0$, αριθμός γενεών

$I \in \mathbb{R}^{n \times n}$, ο ταυτοτικός πίνακας

$m^{(g)} \in \mathbb{R}^n$, μέση τιμή της κατανομής αναζήτησης στην γενιά g

$p \in \mathbb{R}^n$, το μονοπάτι της εξέλιξης

w_i , όπου $i = 1, \dots, \lambda$ τα βάρη στα οποία γίνεται ο ανασυνδιασμός

$x_k^{(g+1)}$, ο k -ος απόγονος από κάθε $(g+1)$ -γενιά

$x_{i:\lambda}^{(g+1)}$, το σημείο με το καλύτερο fitness function από κάθε γενιά.

$$y_k^{(g+1)} = (x_k^{(g+1)} - m^{(g)}) / \sigma^{(g)},$$

2.1 Αποσύνθεση Ιδιοτιμών Θετικά Ορισμένου Πίνακα

Σύμφωνα με τον Herve Abdi [1], οι ιδιοτιμές και τα ιδιοδιανύσματα είναι μεγέθη (ονομάζονται και χαρακτηριστικά μεγέθη ενός πίνακα) τα οποία μας βοηθούν στην αποσύνθεση των ιδιοτιμών ενός τετραγωνικού πίνακα, που αποτελεί την ανάλυση της δομής ενός πίνακα και βρίσκει ιδιαίτερη χρήση στον τομέα της ανάλυσης πολλών μεταβλητών (π.χ. μπορούμε να βρούμε εύκολα μέσω αυτής της μορφής το μέγιστο ή ελάχιστο). Βέβαια, εδώ πρέπει να τονίσουμε ότι αποσύνθεση δεν γίνεται για κάθε τετραγωνικό πίνακα.

Ορισμός: Ένα συμμετρικός πίνακας $C \in \mathbb{R}^{n \times n}$, λέγεται θετικά ορισμένος αν για κάθε διάνυσμα $x \in \mathbb{R}^n \setminus \{0\}$ ισχύει ότι $x^T A x > 0$.

Για κάθε θετικά ορισμένο και συμμετρικό πίνακα C ισχύει ότι έχει μία ορθοκανονική βάση ιδιοδιανυσμάτων, $B = [b_1, b_2 \dots b_n]$ του οποίου οι ιδιοτιμές είναι $d_1^2, d_2^2, \dots, d_n^2 \geq 0$ και για τα οποία ισχύει ότι:

$$C b_i = d_i^2 b_i$$

Έτσι για την αποσύνθεση του A ισχύει ότι:

$$C = B D^2 B^T \tag{2}$$

όπου

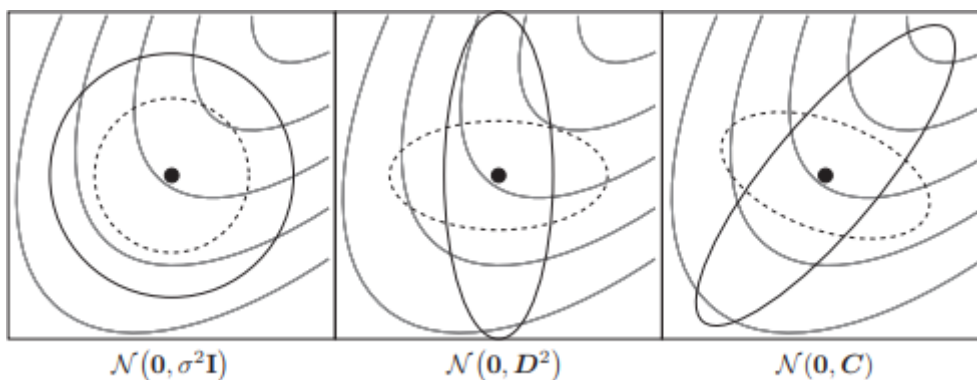
B ορθοκανονική βάση ιδιοδιανυσμάτων του πίνακα

$D^2 = DD = \text{diag}(d_1, \dots, d_2)^2 = \text{diag}(d_1^2, \dots, d_2^2)$ είναι ένας διαγώνιος πίνακας με τιμές τις ιδιοτιμές του C

$D = \text{diag}(d_1, d_2, \dots, d_n)$ είναι ένας διαγώνιος πίνακας με τιμές το τετράγωνο των ιδιοτιμών του C

Κατα αυτόν τον τρόπο ο υπολογισμός του αντιστρόφου πίνακα του A ανάγεται στον εξής υπολογισμό:

$$\begin{aligned} C^{-1} &= (BD^2B^T)^{-1} \\ &= B^T^{-1}D^{-2}B^{-1} \\ &= BD^{-2}B^T \\ &= B\text{diag}\left(\frac{1}{d_1^2}, \frac{1}{d_2^2}, \dots, \frac{1}{d_n^2}\right)B^T \end{aligned}$$



Σχήμα 6: Πηγή: «The CMA Evolution Strategy: A Tutorial», N. Hansen Βλέπε Κεφάλαιο 2.3

Από την (2) προκύπτει ότι ορίζουμε ως τετραγωνική ρίζα του πίνακα C ως:

$$C^{\frac{1}{2}} = BDB^T \quad (3)$$

και έτσι

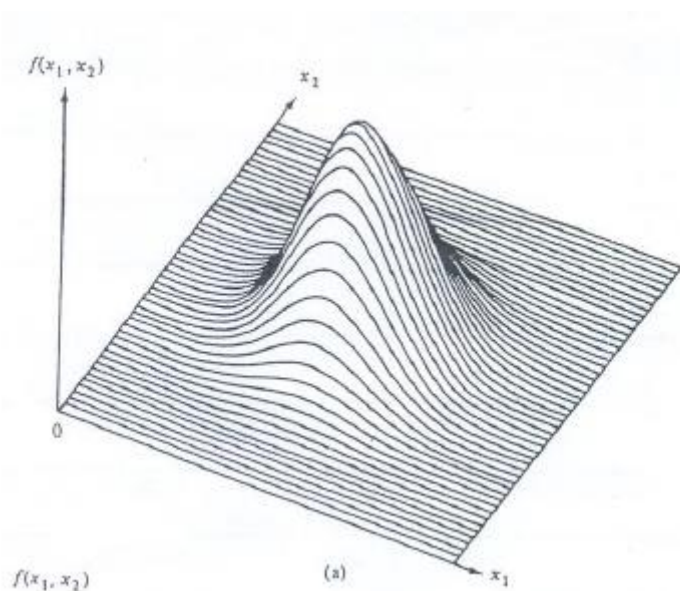
$$\begin{aligned} C^{-\frac{1}{2}} &= BD^{-1}B^T \\ &= B\text{diag}\left(\frac{1}{d_1}, \dots, \frac{1}{d_2}\right)B^T \end{aligned}$$

2.2 Η Πολυδιάστατη Κανονική Κατανομή

Η πολυδιάστατη κανονική κατανομή αποτελεί την γενίκευση της μονομεταβλητής κανονικής κατανομής για διαστάσεις $n \geq 2$, με συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{-1/2}} e^{-(x-\mu)\Sigma^{-1}(x-\mu)/2}$$

Έτσι, όπως και στην κανονική κατανομή μίας μεταβλητής ισχύει ότι ακολουθεί σχήμα καμπάνας με «κορυφή» την μέση τιμή μ μόνο που στην πολυδιάστατη υπάρχει και εξάρτηση από τον πίνακα συνδιακύμανσης C . Χαρακτηριστικά, για 2 διαστάσεις, όπου $C = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$, $\sigma_{11} = \sigma_{22}$ και $Cor(X_1, X_2) = 0$ με έχουμε την εξής κατανομή από την Καλκούνου [19] :



Σχήμα 7: Πηγή: «Πολυμεταβλητή κανονική ανάλυση», Καλκούνου

Γενικά, η πολυδιάστατη κανονική κατανομή που ορίζεται για το τυχαίο διάνυσμα X διαστάσεων $p \times 1$ έχει από κοινού συνάρτηση πυκνότητας πιθανότητας.

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{-1/2}} e^{-(x-\mu)\Sigma^{-1}(x-\mu)/2}$$

Όπου

μ , το διάνυσμα που περιέχει τις αναμενόμενες τιμές κάθε μεταβλητής

Σ , ο πίνακας συνδιακύμανσης των μεταβλητών ενός τυχαίου διανύσματος X

Για τον πίνακα συνδιακύμανσης υπάρχει μια ελκυστική γεωμετρική ερμηνεία: Μπορεί να ερμηνευτεί αποκλειστικά μέσω της (υπερ-)έλλειψης $\{x \in \mathbb{R}^n | x^T C^{-1} x = 1\}$ όπως φαίνεται και στο γράφημα (1). Η έλλειψη αποτελεί επιφάνεια ίδιας πυκνότητας της κατανομής. Οι κύριοι άξονες της έλλειψης ανταποκρίνονται στην στις ιδιοσυναρτήσεις του πίνακα συνδιακύμανσης C ενώ τα μήκη των τετραγωνικών αξόνων στις ιδιοτιμές. Γράφοντας τον πίνακα στην μορφή $C = B(D)^2 B^T$. Αν $D = \sigma I$, όπου $\sigma \in \mathbb{R}_{>0}$ και I είναι ο ταυτοτικός πίνακας, $C = \sigma^2 I$ και η έλλειψη ιστροπική (γράφημα 1 αριστερά). Αν $B = I$ τότε ο $C = D^2$ είναι ένας διαγώνιος πίνακας και το ελλειψοειδές είναι άξονας παράλληλα προσανατολισμένος. Στο σύστημα συντεταγμένων που δίνεται από τις στήλες του B , τότε η κατανομή $N(0, C)$ είναι πάντα ασυσχέτιστη.

Η κανονική κατανομή $N(m, C)$ μπορεί να γραφεί με πολλούς διαφορετικούς τρόπους

$$\begin{aligned}
 N(m, C) &\sim m + N(0, C) \\
 &\sim m + C^{\frac{1}{2}} N(0, I) \\
 &\sim m + B D \underbrace{B^T N(0, I)}_{\sim N(0, C)} \\
 &\sim m + B \underbrace{D N(0, I)}_{\sim N(0, D^2)} \tag{4}
 \end{aligned}$$

όπου " \sim " υποδηλώνει ισότητα στην κατανομή, και $C^{\frac{1}{2}} = B D B^T$. Η τελευταία σειρά μπορεί να ερμηνευτεί καλύτερα ξεκινώντας από τα δεξιά προς τα αριστερά. Έχουμε:

- Η $N(0, I)$ παράγει μια σφαιρική (ισότροπη) κατανομή όπως στο Σχήμα 1, αριστερά.
- Ο D ζυγίζει τη σφαιρική κατανομή μέσα στους άξονες των συντεταγμένων όπως στο σχήμα 1, μέση. $D N(0, I) \sim N(0, D^2)$, το έχει n ανεξάρτητες συνιστώσες. Ο πίνακας D μπορεί να ερμηνευθεί ως πίνακας βηματικού μεγέθους και οι διαγώνιες καταχωρήσεις είναι οι τυπικές αποκλίσεις των συστατικών.
- Ο B ορίζει ένα νέο προσανατολισμό για το ελλειψοειδές, όπου οι νέοι κύριοι άξονες του ελλειψοειδούς αντιστοιχούν στις στήλες του B . Σημειώστε ότι το B έχει $\frac{n^2-n}{2}$ βαθμούς ελευθερίας.

Η εξίσωση (3) είναι χρήσιμη για τον υπολογισμό κατανεμημένων διανυσμάτων $N(m, C)$, διότι $N(0, I)$ είναι ένας φορέας ανεξάρτητων $(0, 1)$ κανονικά κατανεμημένων αριθμών που

2.3 Βελτιστοποίηση μιας Black Box συνάρτησης

Ορισμός: Σύμφωνα με τον Andrea Cassioli [4] ως black-box συνάρτηση θεωρούμε μία συνάρτηση $f: \mathbb{R}^n \mapsto \mathbb{R}$ της οποίας η αναλυτική μορφή είναι άγνωστη. Έτσι το πρόβλημα βελτιστοποίησης που καλούμαστε να λύσουμε είναι το:

$$\begin{aligned} \min_x f(x) \\ g_j(x) &\leq 0 & j = 1, \dots, n \\ x_i &\in [l_i, u_i] \subseteq [-\infty, +\infty] & i = 1, \dots, n \end{aligned}$$

Έτσι, σε τέτοιου είδους συναρτήσεις μπορούμε να εξάγουμε συμπεράσματα σχετικά με την κλίση, την τιμή και τον αν ορίζεται ή όχι σε αυτά τα σημεία η συνάρτηση βασισμένοι σε γνωστές τιμές της συνάρτησης. Αναλυτικότερα, στον τομέα της βελτιστοποίησης μιας black-box συνάρτησης προσπαθούμε να βρούμε το minimum της συνάρτησης βασισόμαστε σε όσες τιμές της διαθέτουμε. Συχνά βέβαια ο υπολογισμός του ολικού ελαχίστου είναι είτε αδύνατος είτε όχι τόσο σχετικός με το ζητούμενο.

Ο αριθμός των τιμών της συνάρτησης που απαιτούνται για την βελτιστοποίησή της ονομάζεται κόστος αναζήτησης search cost και για αυτό συναρτήσεις που για τις οποίες χρειαζόμαστε μεγάλο αριθμό πληροφοριών ($\ll 200$) τις ονομάζουμε costly functions ενώ τις άλλες cheap functions και ανάλογα την κατηγορία στην οποία ανήκουν, συνήθως έχουν διαφορετική αντιμετώπιση επίλυσης.

Σύμφωνα με τον Nikolaus Hansen [7] ένας τυχαίος αλγόριθμος βελτιστοποίησης black-box είναι ο εξής:

```

Initialize distribution parameters  $\theta^{(0)}$ 
For generation  $g = 0, 1, 2, \dots$ 
  Sample  $\lambda$  independent points from distribution  $P(x|\theta^{(g)}) \rightarrow x_1, \dots, x_\lambda$ 
  Evaluate the sample  $x_1, \dots, x_\lambda$  on  $f$ 
  Update parameters  $\theta^{(g+1)} = F_\theta(\theta^{(g)}, (x_1, f(x_1)), \dots, (x_\lambda, f(x_\lambda)))$ 
  break, if termination criterion met

```

Σχήμα 8: Πηγή: «The CMA Evolution Strategy: A Tutorial», N. Hansen

Στην CMA, η κατανομή αναζήτησης, P , είναι μια πολυμεταβλητή κανονική κατανομή. Με δεδομένες όλες τις διακυμάνσεις και covariances, η κανονική κατανομή έχει τη μεγαλύτερη εντροπία όλων των κατανομών στον \mathbb{R}^n . Υπενθυμίζουμε ότι ως εντροπία ορίζεται η ποσότητα της πληροφορίας της πηγής και ισούται με την αναμενόμενη τιμή της πληροφορίας των ενδεχομένων. Επιπλέον, οι κατευθυντήριες γραμμές δεν διακρίνονται με κανέναν τρόπο. Και οι δύο καθιστούν την κανονική κατανομή έναν ιδιαίτερα ελκυστικό υποψήφιο για τυχαία αναζήτηση.

Οι τυχαίοι αλγόριθμοι αναζήτησης θεωρούνται ισχυροί σε ένα «τραχύ» τοπίο αναζήτησης, που μπορεί να περιλαμβάνει ασυνέχειες, (αιχμηρές) κορυφογραμμές ή τοπική μέγιστα. Ειδικότερα, η προσαρμογή μήτρας συνδιακύμανσης (CMA) έχει σχεδιαστεί για να αντιμετωπίσει, επιπρόσθετα, τα προβλήματα που έχουν υποστεί δυσκολίες και τα μη διαχωρίσιμα προβλήματα.

2.4 Εσσιανός Πίνακας και Πίνακας Συνδιακύμανσης

Θεωρούμε την κυρτή-τετραγωνική αντισυμμετρική συνάρτηση $f_H: x \mapsto \frac{1}{2}x^T H x$, όπου Εσσιανός πίνακας $H \in \mathbb{R}^{n \times n}$ είναι ένας θετικά ορισμένος πίνακας. Σύμφωνα με τον Hansen [7] σε μία δοσμένη κατανομή αναζήτησης $N(m, C)$ υπάρχει μία αρκετά μεγάλη σχέση μεταξύ H και C : Θέτοντας $C = H^{-1}$ σε μία f_H είναι ισοδύναμο με την βελτιστοποίηση της ισοτροπικής συνάρτησης $f_{sphere}(x) = \frac{1}{2}x^T x = \frac{1}{2}\sum_i x_i^2$ (όπου $H = I$) με $C = I$. Να τονίσουμε εδώ ότι **ισοτροπική** (isotropic) συνάρτηση λέγεται μια συνάρτηση όταν η τιμή της δεν εξαρτάται από την κατεύθυνση της συνάρτησης. Έτσι, έχουμε ότι στις κυρτές -τετραγωνικές αντικειμενικές συναρτήσεις, ο καθορισμός του πίνακα συνδιακύμανσης της κατανομής αναζήτησης προς το αντίστροφο Εσσιανό πίνακα είναι ισοδύναμος με την αναδιάταξη της λειτουργίας ελλειψοειδούς σε σφαιρική. Συνεπώς, υποθέτουμε ότι ο βέλτιστος πίνακας συνδιακύμανσης ισούται με

το αντίστροφο **Εσσιανό** πίνακα, μέχρι έναν σταθερό παράγοντα. Επιπλέον, η επιλογή ενός πίνακα συνδιακύμανσης ή η επιλογή ενός αντίστοιχου γραμμικού μετασχηματισμού συναρτήσει του χώρου αναζήτησης (δηλαδή του x) είναι ισοδύναμος επειδή για κάθε πλήρους βαθμού $n \times n$ -πίνακα A βρίσκουμε ένα θετικό καθορισμένο Εσσιανό τέτοιο ώστε $\frac{1}{2}(Ax)^T Ax = \frac{1}{2}x^T A^T Ax = \frac{1}{2}x^T Hx$

Ο τελικός στόχος της προσαρμογής του πίνακα συνδιακύμανσης είναι η όσο το δυνατόν κοντινότερη προσέγγιση των γραμμών περιγράμματος Contour της αντικειμενικής συνάρτησης f . Στις κυρτές-τετραγωνικές συναρτήσεις αυτό ισοδυναμεί με την προσέγγιση του αντιστρόφου Εσσιανού πίνακα όπως δηλαδή με μια μέθοδο quasi-Newton.

Στο Σχήμα 1 η κατανομή στερεάς γραμμής στο δεξιό σχήμα ακολουθεί τα περιγράμματα της αντικειμενικής λειτουργίας κατά τον καλύτερο τρόπο και είναι εύκολο να προβλεφθεί ότι θα βοηθήσει να προσεγγίσει το βέλτιστο περισσότερο.

Ο δείκτης κατάστασης ενός θετικά ορισμένου πίνακα A είναι ορισμένος από την Ευκλείδεια νόρμα που είναι: $cond(A) \stackrel{\text{def}}{=} \|A\| \times \|A^{-1}\|$, όπου $\|A\| = \sup_{\|x\|=1} \|Ax\|$. Για έναν θετικά ορισμένο πίνακα A έχουμε ότι $\|A\| = \lambda$ και $cond(A) = \frac{\lambda_{max}}{\lambda_{min}} \geq 1$, όπου λ_{max} και λ_{min} είναι οι μικρότερες και οι μεγαλύτερες ιδιοτιμές του A

Κεφάλαιο 3: Ανάλυση του αλγορίθμου

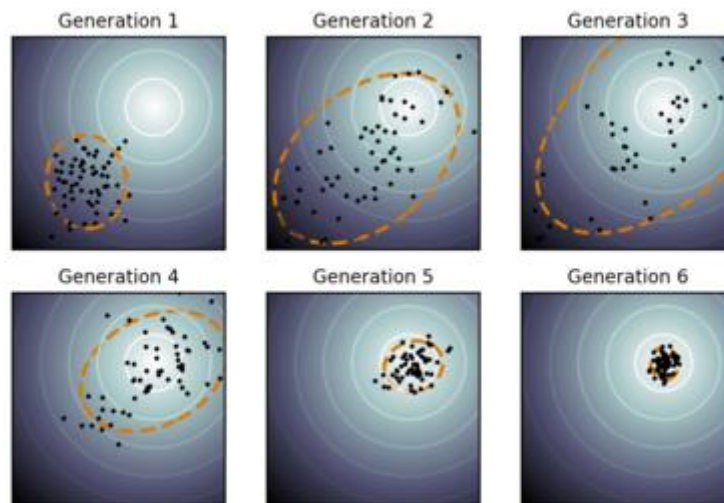
Από την αρχή της δημιουργίας των εξελικτικών αλγορίθμων και βλέποντας την χρησιμότητά τους σε πολλά και πολύπλοκα προβλήματα, άρχισαν να αναπτύσσονται όλο και περισσότεροι για διάφορα προβλήματα. Ένας από αυτούς, ο οποίος επιχείρησε να λύσει προβλήματα βελτιστοποίησης πολλών μεταβλητών σε πολλές διαστάσεις είναι ο Covariance Matrix Adaptation (CMA), τον οποίο και θα επιχειρήσουμε να μελετήσουμε. Όπως κάθε εξελικτικός αλγόριθμος, βασίζεται ευρέως στην αρχή της βιολογικής εξέλιξης, δηλαδή στην επαναλαμβανόμενη αλληλεπίδραση της μεταβολής (μέσω ανασυνδυασμού και μετάλλαξης) και της επιλογής: σε κάθε γενιά (στην περίπτωση του αλγορίθμου μας, επανάληψη) νέα άτομα (υποψήφια λύσεις, που υποδηλώνονται ως x) παράγονται με παραλλαγή, συνήθως με στοχαστικό τρόπο, των σημερινών γονικών ατόμων. Στη συνέχεια, ορισμένα άτομα επιλέγονται για να γίνουν οι γονείς στην επόμενη γενιά βάσει της φυσικής τους αξίας ή της αντικειμενικής τους αξίας $f(x)$. Έτσι, πάνω από τη γενετική ακολουθία, παράγονται άτομα με καλύτερες και καλύτερες τιμές f .

Η CMA αποτελεί Evolutionary Strategy οι οποίες, όπως προαναφέρθηκε, είναι στοχαστικές μέθοδοι για αριθμητική βελτιστοποίηση μη γραμμικών ή μη κυρτών προβλημάτων συνεχούς βελτιστοποίησης.

Συγκεκριμένα, η CMA, μέσα από μια στρατηγική εξέλιξης, υποβάλλει σε δειγματοληψία νέες υποψήφια λύσεις σύμφωνα με μια πολυμεταβλητή κανονική κατανομή στον \mathbb{R}^n . Ο ανασυνδυασμός ισοδυναμεί με την επιλογή μιας νέας μέσης τιμής για τη διανομή ενώ η μετάλλαξη ισοδυναμεί με την προσθήκη ενός τυχαίου διανύσματος, μιας διαταραχής με μηδενική μέση τιμή. Οι αντιστοιχίες εξαρτήσεις μεταξύ των μεταβλητών στην κατανομή αντιπροσωπεύονται από ένα πίνακα συνδιακύμανσης. Η προσαρμογή μήτρας συνδιακύμανσης (CMA) είναι μια μέθοδος για την ενημέρωση της μήτρας συνδιακύμανσης αυτής της κατανομής. Αυτό είναι ιδιαίτερα χρήσιμο αν η συνάρτηση f είναι ακατάλληλη.

Η προσαρμογή της μήτρας συνδιακύμανσης ισοδυναμεί με την εκμάθηση ενός μοντέλου δεύτερης τάξης της βασικής αντικειμενικής συνάρτησης παρόμοιας με την προσέγγιση του αντιστρόφου Εσσιανού πίνακα στη μέθοδο Quasi-Newton στην κλασσική βελτιστοποίηση. Σε αντίθεση με τις περισσότερες κλασσικές μεθόδους,

γίνονται λιγότερες υποθέσεις σχετικά με τη φύση της υποκείμενης αντικειμενικής λειτουργίας. Ουσιαστικά μόνο η κατάταξη μεταξύ υποψηφίων λύσεων αξιοποιείται για την εκμάθηση της κατανομής δειγμάτων και δεν απαιτούνται ούτε από τα ίδια τα παράγωγα ούτε από τις ίδιες οι αξίες λειτουργίας. Χαρακτηριστικά, μπορούμε να δούμε γραφικά ένα τυχαίο πρόβλημα βελτιστοποίησης 2 διαστάσεων μέσα από 6 επαναλήψεις:



Σχήμα 9: Πηγή: Wikipedia

Οι βασικές αρχές που αξιοποιούνται στον αλγόριθμο CMA-ES για την προσαρμογή των παραμέτρων της κατανομής αναζήτησης είναι οι εξής:

1) Μια αρχή μέγιστης πιθανότητας, η οποία βασίζεται στην αύξηση της πιθανότητα επιτυχημένων υποψηφίων λύσεων και βημάτων αναζήτησης. Πιο συγκεκριμένα, ο μέσος όρος της κατανομής ενημερώνεται έτσι ώστε να μεγιστοποιείται η πιθανότητα των επιτυχημένων υποψηφίων λύσεων ενώ, ταυτόχρονα, ο πίνακας συνδιακύμανσης της κατανομής ενημερώνεται (διαδοχικά) έτσι ώστε να αυξάνεται η πιθανότητα προηγούμενων επιτυχημένων βημάτων αναζήτησης. Και οι δύο ενημερώσεις μπορούν να ερμηνευτούν ως φυσική κάθοδος κλίσης. Επίσης, κατά συνέπεια, η CMA διεξάγει μια επανεξεταζόμενη βασική ανάλυση των επιτυχημένων βημάτων αναζήτησης, διατηρώντας παράλληλα όλους τους κύριους άξονες. Η εκτίμηση των αλγορίθμων κατανομής και η μέθοδος Cross-Entropy βασίζονται σε πολύ παρόμοιες ιδέες, αλλά υπολογίζουν (μη διαδοχικά) τον πίνακα συνδιακύμανσης μεγιστοποιώντας την πιθανότητα επιτυχίας των σημείων λύσης αντί των επιτυχημένων βημάτων αναζήτησης.

2) Καταγράφονται δύο διαδρομές της χρονικής εξέλιξης του μέσου διανομής της στρατηγικής, που ονομάζονται διαδρομές αναζήτησης ή μονοπάτια εξέλιξης. Αυτές οι διαδρομές περιέχουν σημαντικές πληροφορίες σχετικά με τη συσχέτιση μεταξύ διαδοχικών βημάτων. Συγκεκριμένα, αν ακολουθούνται διαδοχικά βήματα προς μια παρόμοια κατεύθυνση, οι διαδρομές εξέλιξης καθίστανται μεγάλες. Οι διαδρομές εξέλιξης είναι εκμεταλλεύσιμες με δύο τρόπους. Η μία διαδρομή χρησιμοποιείται για τη διαδικασία προσαρμογής μήτρας συνδιακύμανσης αντί για μεμονωμένα επιτυχημένα βήματα αναζήτησης και διευκολύνει μια πιθανώς πολύ ταχύτερη αύξηση διακύμανσης ευνοϊκών διευθύνσεων. Η άλλη διαδρομή χρησιμοποιείται για τη διεξαγωγή πρόσθετου ελέγχου μεγέθους βημάτων. Αυτός ο έλεγχος μεγέθους βημάτων στοχεύει να καταστήσει τις επακόλουθες κινήσεις της μέσης κατανομής ορθογώνιες σε προσδοκία. Ο έλεγχος μεγέθους βημάτων εμποδίζει αποτελεσματικά την πρόωρη σύγκλιση, επιτρέποντας ωστόσο τη γρήγορη σύγκλιση σε ένα βέλτιστο επίπεδο.

3.1 Η Βασική Εξίσωση

Στην CMA, όπως τονίζει και ο Hansen [7] οι απόγονοι κάθε γενιάς ακολουθούν την εξής εξίσωση:

$$x_k^{(g+1)} \sim m^{(g)} + \sigma^{(g)} \mathcal{N}(0, C^{(g)}) \quad \text{για } k = 1, \dots, \lambda \quad (5)$$

Όπου,

$x_k^{(g+1)}$, ο k -ος απόγονος της $g + 1$ γενιάς

$m^{(g)}$, μέση τιμή της κατανομής στην γενιά g

$\sigma^{(g)} \in \mathbb{R}_{>0}$, ολική τυπική απόκλιση για την γενιά g

$C^{(g)} \in \mathbb{R}^{n \times n}$, πίνακας συνδιακύμανσης για την γενιά g

$\lambda \geq 2$, αριθμός πληθυσμού

Έτσι, είναι φανερό ότι για να μπορέσουμε να συνεχίσουμε τα βήματα της επαναληπτικής διαδικασίας μας θα πρέπει να υπολογίσουμε όλα τα επιμέρους στοιχεία της εξίσωσης

3.2 Μέση Τιμή m

Όπως έχουμε τονίσει, η διαδικασία μετά από κάθε βήμα επανάληψης της θα προσαρμόζεται στα δεδομένα που έχουν προκύψει. Έτσι, η μέση τιμή $m^{(g+1)}$ θα αλλάζει και αυτή και, πιο συγκεκριμένα, θα αποτελεί τον σταθμισμένο μέσο μ σημείων του δείγματος του δείγματος $x_1^{(g+1)}, x_2^{(g+1)}, \dots, x_\lambda^{(g+1)}$:

$$m^{(g+1)} = \sum_{i=1}^{\mu} w_i x_{i:\lambda}^{(g+1)} \quad (6)$$

Για τα βάρη των οποίων ισχύει ότι:

- 1) $\sum_{i=1}^{\mu} w_i = 1$
- 2) $w_1 \geq w_2 \geq \dots \geq w_\mu \geq 0$

Και όπου,

$\mu \leq \lambda$ είναι ο αρχικός μας πληθυσμός και στην περίπτωση μας ο αριθμός των σημείων με τον οποίο θα γίνει η διαδικασία

$w_{i=1, \dots, \mu} \in \mathbb{R}_{>0}$, είναι θετικά βάρη. Να τονίσουμε εδώ ότι για $w_{i=1, \dots, \mu} = \frac{1}{\mu}$ μέσω της

(6) θα μπορούσαμε να υπολογίσουμε την μέση τιμή των μ σημείων μας

$x_{i:\lambda}^{(g+1)}$, είναι το i -οστό καλύτερο σωματίδιο εκ των $x_1^{(g+1)}, \dots, x_\lambda^{(g+1)}$

Η εξίσωση (6) υλοποιεί την **επιλογή περικοπών** (truncation selection) επιλέγοντας μ από τα συνολικά λ σημεία απογόνων. Η εκχώρηση διαφορετικών βαρών w_i θα πρέπει επίσης να ερμηνευτεί ως μηχανισμός επιλογής. Η εξίσωση (6) εφαρμόζει **σταθμισμένο ενδιάμεσο ανασυνδυασμό** (weighted intermediate recombination) λαμβάνοντας υπόψη μ άτομα για έναν σταθμισμένο μέσο όρο.

Από εδώ και πέρα, συχνά θα χρησιμοποιούμε την **μάζα διαφοράς αποτελεσματικής επιλογής** (variance effective selection mass), μ_{eff} , για το οποίο έχουμε:

$$\mu_{eff} = \left(\frac{\|w\|_1}{\|w\|_2} \right)^2 = \frac{\|w\|_1^2}{\|w\|_2^2} = \frac{(\sum_{i=1}^{\mu} w_i)^2}{\sum_{i=1}^{\mu} w_i^2} = \frac{1}{\sum_{i=1}^{\mu} w_i^2} \quad (8)$$

Με βάση την εξίσωση μας προκύπτει ότι $1 \leq \mu_{eff} \leq \mu$. Ισότητα έχουμε όταν τα βάρη ανασυνδυασμού ισούνται.

Όπως τονίζει ο Hansen [5] μια συνηθισμένη τιμή όταν έχουμε κανονικά βάρη ανασυνδυασμού για το μ_{eff} είναι το $\lambda/4$ ενώ μία απλή και λογική ρύθμιση θα μπορούσε να είναι $w_1 \propto \mu - i + 1$ και $\mu \approx \lambda/2$.

Με βάση την (8) προκύπτει, λοιπόν, μια αναβάθμιση για το m η οποία είναι

$$m^{(g+1)} = m^{(g)} + c_m \sum_{t=1}^{\mu} w_i \left(x_{i:\lambda}^{(g+1)} - m^{(g)} \right) \quad (9)$$

όπου,

c_m , ο ρυθμός εκμάθησης (learning rate) ο οποίος συνήθως παίρνει την τιμή 1

Παρατηρούμε, λοιπόν, πως η εξίσωση (9) αποτελεί γενίκευση της εξίσωσης (6), οι οποίες ισούνται για $c_m \sum_{i=1}^{\mu} w_i$. Συνήθως διαλέγουμε $c_m < 1$ όταν έχουμε Black Box συνάρτηση.

3.3 Προσαρμόζοντας τον πίνακα συνδιακύμανσης

Θα μελετήσουμε τώρα την πορεία του πίνακα συνδιακύμανσης (Covariance Matrix) στην διαδικασία μας. Είναι σημαντικό να τονίσουμε εδώ ότι όταν ο πληθυσμός είναι μικρός πολλές φορές ο υπολογισμός του πίνακα είναι αχρείαστος και ως εκ τούτου χρησιμοποιείται μία διαφορετική προσέγγιση την οποία και θα αναλύσουμε. Συγκεκριμένα, χρησιμοποιείται η αναβάθμιση μ -βαθμού (rank μ -update) ενώ στην ιδιαίτερη περίπτωση που ο πληθυσμός έχει μόνο ένα άτομο έχουμε αναβάθμιση 1^{ου} βαθμού.

3.3.1 Υπολογίζοντας τον πίνακα συνδιακύμανσης C

Έστω ότι το νούμερο του πληθυσμού μας είναι επαρκές σε πληροφορίες για τον υπολογισμό του πίνακα μας. Για ευκολία στους υπολογισμούς θέτουμε $\sigma^{(g)} = 1$ στο κεφάλαιο αυτό.

Έχουμε λοιπόν άλλη μία εκτίμηση του πίνακα συνδιακύμανσης $C^{(g)}$ χρησιμοποιώντας το δείγμα του πληθυσμού από την (4), $x_1^{(g+1)}, \dots, x_\lambda^{(g+1)}$

$$C_{emp}^{(g+1)} = \frac{1}{\lambda-1} \sum_{i=1}^{\lambda} \left(x_i^{(g+1)} - \frac{1}{\lambda} \sum_{j=1}^{\lambda} x_j^{(g+1)} \right) \left(x_i^{(g+1)} - \frac{1}{\lambda} \sum_{j=1}^{\lambda} x_j^{(g+1)} \right)^T \quad (10)$$

Ο εμπειρικός πίνακας συνδιακύμανσης $C_{emp}^{(g+1)}$ είναι ένας αμερόληπτος εκτιμητής του $C^{(g)}$: υποθέτουμε ότι $x_i^{(g+1)}, i = 1, \dots, \lambda$ να είναι τυχαίες μεταβλητές και έτσι προκύπτει ότι $E[C_{emp}^{(g+1)} | C^{(g)}] = C^{(g)}$. Θα πάρουμε τώρα μια ελαφρώς διαφορετική προσέγγιση ούτως ώστε να πάρουμε έναν εκτιμητή για το $C^{(g)}$. Έτσι έχουμε:

$$C_{\lambda}^{(g+1)} = \frac{1}{\lambda} \sum_{i=1}^{\lambda} \left(x_i^{(g+1)} - m^{(g)} \right) \left(x_i^{(g+1)} - m^{(g)} \right)^T \quad (11)$$

Επίσης ο πίνακας $C_{\lambda}^{(g+1)}$ είναι ένας αμερόληπτος εκτιμητής του $C^{(g)}$. Η αξιοσημείωτη διαφορά μεταξύ (10) και (11) είναι η μέση τιμή αναφοράς. Για το $C_{emp}^{(g+1)}$ είναι ο μέσος όρος του πραγματικού αντιληπτού δείγματος. Για το $C_{\lambda}^{(g+1)}$ είναι η πραγματική μέση τιμή, $m^{(g)}$, της κατανομής του δείγματος (βλ. (5)). Επομένως, οι εκτιμητές $C_{emp}^{(g+1)}$ και $C_{\lambda}^{(g+1)}$ μπορεί να ερμηνευθούν διαφορετικά: ενώ το $C_{emp}^{(g+1)}$ υπολογίζει τη διακύμανση κατανομής στα σημεία δειγματοληψίας, $C_{\lambda}^{(g+1)}$ εκτιμά τις διακυμάνσεις της δείγματα βημάτων, $x_{i:\lambda}^{(g+1)} - m^{(g)}$.

Μια μικρή διαφορά μεταξύ των εξισώσεων(10) και (11) είναι οι παράγοντες κανονικοποίησης που ακολουθούν, οι οποίοι είναι $\frac{1}{\lambda-1}$ και $\frac{1}{\lambda}$ αντίστοιχα, προκειμένου να πάρουμε έναν αμερόληπτο εκτιμητή. Συγκεκριμένα, στην περίπτωση της (10) έχουμε ήδη έναν βαθμό ελευθερίας μέσα στο εσωτερικό του αθροίσματος και για αυτό και υπάρχει η διαφορά στους παράγοντες.

Είδαμε ήδη μέσω της (11) μια επανεκτίμηση του πίνακα συνδιακύμανσης. Προκειμένου λοιπόν να επιτευχθεί μία ακόμα καλύτερη χρησιμοποιούμε τον ίδιο μηχανισμό επιλογής βαρών όποτε η (6) γίνεται:

$$C_{\mu}^{(g+1)} = \sum_{i=1}^{\mu} w_i \left(x_{i:\lambda}^{(g+1)} - m^{(g)} \right) \left(x_{i:\lambda}^{(g+1)} - m^{(g)} \right)^T \quad (12)$$

Ο πίνακας συνδιακύμανσης $C_{\mu}^{(g+1)}$ είναι ένας εκτιμητής της κατανομής των επιλεγμένων βημάτων, ακριβώς όπως και ο $C_{\lambda}^{(g+1)}$ είναι ένας εκτιμητής της αρχικής κατανομής των βημάτων πριν από την διαδικασία της επιλογή. Η δειγματοληψία από

του $C_\mu^{(g+1)}$ τείνει να αναπαράγει επιλεγμένα, δηλαδή επιτυχή βήματα, δίνοντας μια αιτιολόγηση για το τι σημαίνει ένας "καλύτερος" πίνακας συνδιακύμανσης.

Έπειτα ακολουθώντας την διαδικασία του Hansen [7], συγκρίνουμε την (12) με την Εκτίμηση του Πολυμεταβλητού Κανονικού Αλγορίθμου $EMNA_{global}$. Έτσι, έχουμε:

$$C_{EMNA}^{(g+1)} = \frac{1}{\mu} \sum_{i=1}^{\lambda} \left(x_{i:\lambda}^{(g+1)} - m^{(g+1)} \right) \left(x_{i:\lambda}^{(g+1)} - m^{(g+1)} \right)^T \quad (13)$$

όπου, $m^{(g+1)} = \frac{1}{\mu} \sum_{i=1}^{\mu} x_{i:\lambda}^{(g+1)}$. Αντίστοιχα, εφαρμόζοντας την μέθοδο Cross-entropy για βελτιστοποίηση σε συνεχείς χώρους έχουμε τον πίνακα συνδιακύμανσης που γίνεται $\frac{\mu}{\mu-1} C_{EMNA_{global}}^{(g+1)}$. Και στις δύο περιπτώσεις η λεπτή αλλά σημαντική διαφορά στο (12) είναι και πάλι η επιλογή της μέσης τιμής αναφοράς. Η εξίσωση (13) υπολογίζει τη διακύμανση εντός του επιλεγμένου πληθυσμού ενώ η (12) κάνει εκτίμηση για τα επιλεγμένα βήματα. Η εξίσωση (13) αποκαλύπτει πάντοτε μικρότερες διακυμάνσεις από (12), επειδή η μέση τιμή αναφοράς είναι ο ελαχιστοποιητής για τις διακυμάνσεις. Επιπλέον, στις περισσότερες περιπτώσεις πιθανής επιλογής (13) μειώνονται οι διακυμάνσεις σε σχέση με τον πίνακα συνδιακύμανσης $C^{(g)}$.

Στο Σχήμα 10 φαίνονται τα αποτελέσματα της εκτίμησης σε μια γραμμική αντικειμενική συνάρτηση όπου $\lambda = 150$, $\mu = 50$ και $w_i = 1/\mu$. Η εξίσωση (12) αυξάνει γεωμετρικά την αναμενόμενη διακύμανση στην κατεύθυνση της κλίσης (όπου γίνεται η επιλογή. Στην συγκεκριμένη περίπτωση έχουμε την διαγώνια), λαμβάνοντας υπόψη τις συνήθεις ρυθμίσεις για το γονικό αριθμό μ και τα βάρη ανασυνδυασμού w_1, \dots, w_i . Εν αντιθέσει, η εξίσωση (13) μειώνει πάντα τη γεωμετρική ταχύτητα της διακύμανσης της καμπύλης κλίσης. Ως εκ τούτου, η (13) είναι εξαιρετικά ευάλωτη σε πρόωρη σύγκλιση, ιδίως άμα έχει μικρούς γονικούς πληθυσμούς, εκεί όπου ο πληθυσμός δεν μπορεί να αναμένεται να υποστηρίξει το βέλτιστο ανά πάσα στιγμή. Ωστόσο, για μεγάλες τιμές μ σε μεγάλους πληθυσμούς με μεγάλες αρχικές διακυμάνσεις, ο αντίκτυπος της διαφορετικής μέσης τιμής αναφοράς μπορεί να γίνει οριακός.

Προκειμένου να διασφαλίσουμε μέσω των εξισώσεων (5), (6) και (12) ότι ο $C_\mu^{(g+1)}$ είναι ένας αξιόπιστος εκτιμητής, η αποτελεσματική μάζα επιλογής της διακύμανσης μ_{eff} πρέπει να είναι αρκετά μεγάλη αφού προκειμένου να έχουμε τα νούμερα που

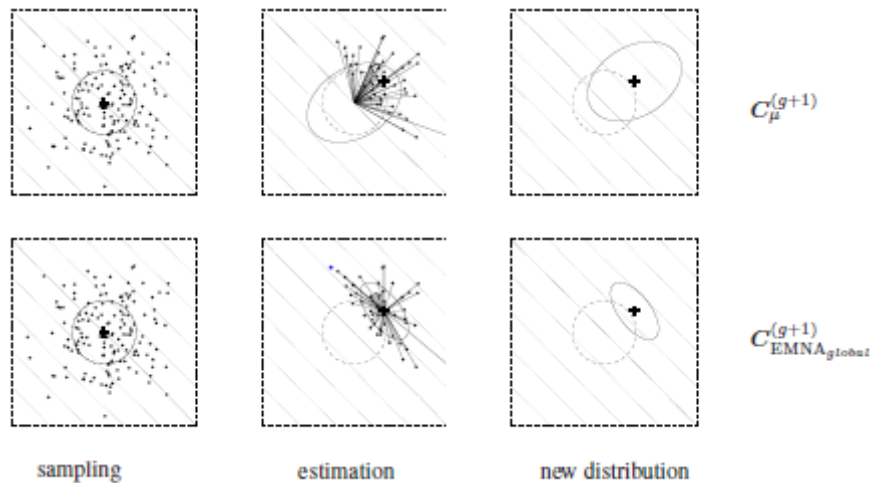
απαιτούνται από τις συνθήκες μας μικρότερα από 10 για τον $C_\mu^{(g)}$ πάνω στην $f_{sphere} = \sum_{t=1}^n x_t^2$, πρέπει να ισχύει ότι $\mu_{eff} \approx 10n$. Το επόμενο βήμα είναι να γίνει παράκαμψη του περιορισμού αυτού στο μ_{eff} .

3.3.2 Αναβάθμιση μ-βαθμού

Για την επίτευξη μιας «γρήγορης» δημιουργίας του πίνακα συνδιακύμανσης για ένα πρόβλημα, έχουμε ότι για την την σφαίρα $f_{sphere} : x \mapsto \sum x_t^2$ πρέπει να έχουμε μικρό πληθυσμό. Τυπικά, και ιδανικά, έχουμε ότι $\mu_{eff} \approx \lambda/4$ και έτσι αφού θέλουμε να είναι μικρό πληθυσμό τότε θα πρέπει και το μ_{eff} να είναι μικρό δημιουργώντας όμως πρόβλημα αναφορικά με τον πίνακα συνδιακύμανσης αφού δεν θα μπορούμε να βρούμε έναν καλό εκτιμητή. Για να λυθεί αυτό το πρόβλημα, χρησιμοποιούνται πληροφορίες από προηγούμενες γενιές. Ένα παράδειγμα, σύμφωνα με τον Hansen [7], τέτοιου πίνακα συνδιακύμανσης είναι ο:

$$C^{(g+1)} = \frac{1}{g+1} \sum_{i=0}^g \frac{1}{\sigma^{(i)^2}} C_\mu^{(t+1)} \quad (14)$$

Που προκύπτει ότι



Σχήμα 10: Πηγή: Γραμμές Contour οι οποίες μας δείχνουν την συγκέντρωση των σημείων μέσα στην διαδικασία. Στην πάνω γραμμή έχουμε τον εκτιμητή του $C^{(g+1)}$ με $w_i = 1/\mu$ ενώ στην κάτω εκτίμηση του $C_{EMNA}^{(g+1)}$. Η αριστερή στήλη περιέχει δείγμα με $\lambda = 150N(0, I)$ κατανεμημένων σημείων, η μεσαία 50 σημεία που καθορίζουν το ποια θα συμμετάσχουν στην εκτίμηση και η δεξιά την κατανομή αναζήτησης.

Στον εκτιμητή του προβλήματος τα βήματα μεταξύ των γενεών έχουν ίδιο βάρος και προκειμένου να έχουμε μεγαλύτερο βάρος θα επιχειρήσουμε να την κάνουμε εκθετική εξομάλυνση (exponential smoothing).

Για τον λόγο αυτό, διαλέγουμε $C^{(0)} = I$ για να είναι ο αρχικός μας πίνακας και $0 < c_\mu \leq 1$ και έτσι για τον πίνακα της γενιάς $g + 1$ έχουμε:

$$\begin{aligned} C^{(g+1)} &= (1 - c_\mu)C^{(g)} + c_\mu \frac{1}{\sigma^{(g)^2}} C_\mu^{(g+1)} \\ &= (1 - c_\mu)C^{(g)} + c_\mu \sum_{i=1}^{\mu} w_i y_{i:\lambda}^{(g+1)} y_{i:\lambda}^{(g+1)T} \end{aligned} \quad (15)$$

όπου,

$c_\mu \leq 1$ αφού αν ήταν 1 δεν θα διατηρηθεί καμία προηγούμενη πληροφορία και άμα 0 δεν θα έχουμε εισαγωγή καμίας άλλης

$$w_{i\dots\mu} \in \mathbb{R} \text{ με } \sum_i w_i = 1$$

$$y_{i:\lambda}^{(g+1)} = (x_{i:\lambda}^{(g+1)} - m^{(g)}) / \sigma^{(g)}$$

Έτσι προκύπτει ο πίνακας μ-βαθμού αναβάθμισης που ονομάζεται έτσι, επειδή το άθροισμα των εξωτερικών προ είναι βαθμού $\min(\mu, n)$ με πιθανότητα 1 (δεδομένου ότι μ έχει μη μηδενικά βάρη). Αν $\mu = 1$ τότε αποτελείται από έναν όρο.

$$\begin{aligned} C^{(g+1)} &= \left(1 - c_\mu \sum w_i\right) C^{(g)} + c_\mu \sum_{i=1}^{\mu} w_i y_{i:\lambda}^{(g+1)} y_{i:\lambda}^{(g+1)T} \\ &= C^{(g)1/2} \left(1 + c_\mu \sum_{i=1}^{\lambda} w_i \left(z_{i:\lambda}^{(g+1)} z_{i:\lambda}^{(g+1)T} - I\right)\right) C^{(g)1/2} \end{aligned} \quad (16)$$

όπου,

$$z_{i:\lambda}^{(g+1)} = C^{(g)-1/2} y_{i:\lambda}^{(g+1)}$$

$$w_{i\dots\lambda} \in \mathbb{R} \sum_{t=1}^{\lambda} w_i \approx 0$$

$$\sum w_i = \sum_{t=1}^{\lambda} w_i$$

Η δεύτερη γραμμή της (16) εκφράζει την αναβάθμιση στο φυσικό σύστημα συντεταγμένων, μια ιδέα που έχει ήδη εξεταστεί στο [8]. Ο ταυτοτικός πίνακας συνδιακύμανσης ενημερώνεται και μεταγενέστερα εφαρμόζεται μετασχηματισμός συστήματος συντεταγμένων με πολλαπλασιασμό με $C^{(g)1/2}$ και στις δύο πλευρές. Η

εξίσωση (16) χρησιμοποιεί βάρη, w_i , εκ των οποίων τα μισά είναι αρνητικά. Εάν τα βάρη επιλέγονται έτσι ώστε $\sum w_i = 0$, η αποσύνθεση του $C^{(g)}$ εξαφανίζεται και οι αλλαγές γίνονται μόνο κατά μήκος των αξόνων στους οποίους πραγματοποιούνται τα δείγματα.

Ο αριθμός $1/c_\mu$ είναι ο χρονικός ορίζοντας προς τα πίσω που προσφέρει περίπου το 63% του συνόλου πληροφορίες.

Συνεχίζοντας την (16) λοιπόν έχουμε για το σταθμισμένο άθροισμα:

$$C^{(g+1)} = (1 - c_\mu)^{g+1} C^{(0)} + c_\mu \sum_{i=0}^g (1 - c_\mu)^{g-i} \frac{1}{\sigma(i)^2} C_\mu^{(i+1)} \quad (17)$$

ο ορίζοντας του χρόνου του παρελθόντος, Δ_g , όπου 63% του συνολικού βάρους είναι αθροισμένο και έτσι ορίζεται ως

$$c_\mu \sum_{i=g+1-\Delta_g}^g (1 - c_\mu)^{g-i} \approx 0.63 \approx 1 - \frac{1}{e} \quad (18)$$

Και επιλύοντας για το άθροισμα έχουμε:

$$(1 - c_\mu)^{\Delta_g} \approx \frac{1}{e} \quad (19)$$

Ενώ επιλύοντας για το Δ_g , χρησιμοποιώντας της λογαριθμική προσέγγιση Taylor έχουμε:

$$\Delta_g \approx \frac{1}{c_\mu} \quad (20)$$

Δηλαδή, περίπου το 37% των πληροφοριών στον πίνακα $C^{(g)}$ είναι παλαιότερες από $1/c_\mu$ γενεές, και, σύμφωνα με το (19), το αρχικό βάρος μειώνεται κατά ένα συντελεστή 0,37 μετά από περίπου $1/c_\mu$ γενεές.

Η επιλογή του c_μ είναι κρίσιμη. Οι μικρές τιμές οδηγούν σε αργή εκμάθηση, πολύ μεγάλες τιμές οδηγούν σε αποτυχία, επειδή ο πίνακας συνδιακύμανσης εκφυλίζεται. Ευτυχώς, μια καλή ρύθμιση φαίνεται να είναι σε μεγάλο βαθμό ανεξάρτητη από τη λειτουργία που πρέπει να βελτιστοποιηθεί. Μια προσέγγιση πρώτης τάξης για μια καλή επιλογή είναι $c_\mu \approx \mu_{eff}/n^2$. Επομένως, ο χαρακτηριστικός χρονικός ορίζοντας για το (16) είναι περίπου n^2/μ_{eff} .

Τα πειράματα σύμφωνα με τον Hansen υποδεικνύουν ότι $c_\mu \approx \mu_{eff}/n^2$ είναι μάλλον συντηρητική ρύθμιση για μεγάλες τιμές του n , ενώ το $\mu_{eff}/n^{1.5}$ φαίνεται να είναι

ελαφρώς πέρα από το όριο σταθερότητας. Η καλύτερη, αλλά παράλληλα και στιβαρή επιλογή του εκθέτη παραμένει ανοιχτό.

Ακόμη και για τον ρυθμό εκμάθησης $c_\mu = 1$, η προσαρμογή του πίνακα συνδιακύμανσης δεν μπορεί να επιτευχθεί μέσα σε μία γενιά. Η επίδραση της αρχικής κατανομής δειγμάτων (πληθυσμού) δεν εξαφανίζεται μέχρι να υπάρξει επαρκής αριθμός γενεών. Υποθέτοντας ένα σταθερό κόστος αναζήτησης (αριθμός αξιολογήσεων λειτουργίας), ένα μικρό μέγεθος πληθυσμού λ επιτρέπει έναν μεγαλύτερο αριθμό γενεών και επομένως συνήθως οδηγεί σε ταχύτερη προσαρμογή του πίνακα συνδιακύμανσης.

3.3.2 Αναβάθμιση 1-βαθμού

Για την ανάβαθμιση θεωρούμε μία συγκεκριμένη μέθοδο παραγωγής n -διαστάσεων κανονικών κατανομών με μηδενική μέση τιμή. Έστω διανύσματα $y_1, \dots, y_{g_0} \in \mathbb{R}^n$, $g_0 \geq n$ και $\mathcal{N}(0,1)$ η κανονική κατανομή με μέση τιμή 0 και διασπορά 1. Τότε ο όρος

$$\mathcal{N}(0,1)y_1 + \dots + \mathcal{N}(0,1)y_{g_0} \sim \mathcal{N}\left(0, \sum_{i=1}^{g_0} y_i y_i^T\right) \quad (21)$$

είναι ένα τυχαίο κατανεμημένο διάνυσμα με μέση τιμή 0 και πίνακα συνδιακύμανσης $\sum_{i=1}^{g_0} y_i y_i^T$. Ουσιαστικά, το τυχαίο αυτό διάνυσμα παράγεται μέσω πρόσθεσης «γραμμικών-κατανομών» $\mathcal{N}(0,1)y_i$. Η κατανομή $\mathcal{N}(0,1)y_i \sim \mathcal{N}(0, y_i y_i^T)$ παράγει το διάνυσμα με μέγιστη πιθανότητα λόγω των ιδιοτήτων της κατανομής.

Με βάση λοιπόν την (21) και την (16) θα επιχειρήσουμε να δούμε την προσαρμογή του πίνακα συνδιακύμανσης. Έστω ότι το άθροισμα της (16) αποτελείται από μόνο ένα μόνο μέλος και έστω ότι $y_{g+1} = \frac{x_{i,\lambda}^{(g+1)} - m^{(g)}}{\sigma^{(g)}}$. Τότε, για την αναβάθμιση 1-βαθμού του πίνακα συνδιακύμανσης έχουμε:

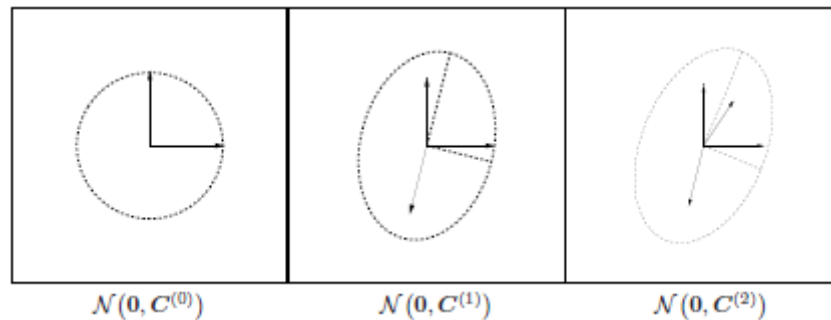
$$\mathcal{C}^{(g+1)} = (1 - c_1)\mathcal{C}^{(g)} + c_1 y_{g+1} y_{g+1}^T \quad (22)$$

Το άθροισμα του δεξιού μέλους της εξίσωσής μας είναι πρώτης βαθμίδας και προσθέτει τον μέγιστο όρο πιθανότητας για το y_{g+1} στον πίνακα συνδιακύμανσης $\mathcal{C}^{(g)}$. Επομένως, η πιθανότητα να παραχθεί y_{g+1} στην επόμενη γενιά αυξάνεται.

Ένα παράδειγμα που μπορεί να δει κανείς στο Σχήμα 4 των δύο πρώτων βημάτων επανάληψης του (22) φαίνεται στο Σχήμα 4. Η κατανομή $\mathcal{N}(0, C^{(1)})$ τείνει να αναπαράγει το y_1 με μεγαλύτερη πιθανότητα από την αρχική κατανομή $\mathcal{N}(0, I)$ ενώ η κατανομή $\mathcal{N}(0, C^{(2)})$ τείνει να αναπαράγει το y_2 με μεγαλύτερη πιθανότητα από το $\mathcal{N}(0, C^{(1)})$, και ούτω καθεξής. Όταν y_1, \dots, y_g υποδηλώνουν τα προηγούμεως επιλεγμένα, ευνοϊκά βήματα, $\mathcal{N}(0, C^{(g)})$ τείνει να αναπαράγει αυτά τα βήματα. Η διαδικασία οδηγεί σε ευθυγράμμιση της διανομής αναζήτησης $\mathcal{N}(0, C^{(g)})$ στην κατανομή των επιλεγμένων βαθμίδων. Αν και οι δύο κατανομές γίνουν παρόμοιες, όπως και κάτω από την τυχαία επιλογή, όπως αναμένεται δεν πραγματοποιείται καμία άλλη αλλαγή του πίνακα συνδιακύμανσης.

3.3.3 Cumulation: Χρησιμοποιώντας την διαδρομή της εξέλιξης

Στην μέχρι τώρα διαδικασία μας έχουμε χρησιμοποιήσει ως βήματα αναβάθμισης του πίνακα συνδιακύμανσης



Σχήμα 11: Πορεία αναβάθμισης πίνακα συνδιακύμανσης

$$\frac{m^{(g+1)} - m^{(g)}}{\sigma^{(g)}} + \frac{m^{(g)} - m^{(g-1)}}{\sigma^{(g-1)}} + \frac{m^{(g-1)} - m^{(g-2)}}{\sigma^{(g-2)}} \quad (23)$$

$$p_c^{(g+1)} = (1 - c_c)p_c^{(g)} + \sqrt{c_c(2 - c_c)}\mu_{eff} \frac{m^{(g+1)} - m^{(g)}}{\sigma^{(g)}} \quad (24)$$

Όπου,

$p_c^{(g)} \in \mathbb{R}^n$, μονοπάτι εξέλιξης στην γενιά g

$c_c \leq 1$. Ισχύει ότι $1/c_c$ είναι ο χρονικός ορίζοντας που πάει προς τα πίσω στο μονοπάτι της εξέλιξης p_c και περιέχει περίπου το 63% του συνολικού βάρους. Αποτελεσματικός χρονικός ορίζοντας είναι μεταξύ \sqrt{n} και n .

Ο παράγοντας $\sqrt{c_c(2 - c_c)\mu_{eff}}$ είναι μια σταθερά κανονικοποίησης για το p_c . Για $c_c = 1$ και $\mu_{eff} = 1$ ο παράγοντας μειώνεται σε έναν και έτσι $p_c^{(g+1)} = (x_{1:\lambda}^{(g+1)})/\sigma^{(g)}$. Τον παράγοντα $\sqrt{c_c(2 - c_c)\mu_{eff}}$ τον διαλέγουμε έτσι ώστε να ισχύει

$$p_c^{(g+1)} \sim \mathcal{N}(0, C) \quad (25)$$

αν

$$p_c^{(g)} \sim \frac{x_{i:\lambda}^{(g+1)} - m^{(g)}}{\sigma^{(g)}} \sim \mathcal{N}(0, C) \quad \text{για κάθε } i = 1, \dots, \mu \quad (26)$$

Έτσι, η αναβάθμιση του πίνακα συνδιακύμανσης μέσω του μονοπατιού εξέλιξης είναι:

$$C^{(g)} = (1 - c_1)C^{(g)} + c_1 p_c^{(g+1)} p_c^{(g+1)T} \quad (27)$$

Μια εμπειρικά έγκυρη επιλογή ρυθμού εκμάθησης c_1 είναι το $c_1 \approx 2/n^2$. Για $c_c = 1$ και $\mu = 1$ οι εξισώσεις (27), (22) και (16) είναι ίδιες.

Η χρήση του μονοπατιού εξέλιξης στην αναβάθμιση του πίνακα συνδιακύμανσης C βελτιώνει σημαντικά την διαδικασία για μικρό μ_{eff} επειδή εκμεταλλευόμαστε τις συσχετίσεις μεταξύ των διαδοχικών βημάτων.

Θεωρούμε δύο ακραίες καταστάσεις με απολύτως εξαρτώμενα βήματα και με πλήρως μη εξαρτώμενα βήματα. Για τις θετικές εξαρτήσεις έχουμε:

$$\sum_{i=0}^g (1 - c_c)^i \rightarrow \frac{1}{c_c} \text{ όταν το } g \rightarrow \infty$$

ενώ για τις αρνητικές

$$\begin{aligned} \sum_{i=0}^g (-1)^i (1 - c_c)^i &= \sum_{i=0}^{\lfloor g/2 \rfloor} (1 - c_c)^{2i} - \sum_{i=0}^{\lfloor g/2 \rfloor} (1 - c_c)^{2i+1} \\ &= \sum_{i=0}^{\lfloor g/2 \rfloor} (1 - c_c)^{2i} - (1 - c_c) \sum_{i=0}^{(g-1)/2} (1 - c_c)^{2i} \\ &= c_c \sum_{i=0}^{\lfloor g/2 \rfloor} ((1 - c_c)^2)^i + (1 - c_c)^g ((g+1) \bmod 2) \\ &\rightarrow \frac{c_c}{1 - (1 - c_c)^2} = \frac{1}{2 - c_c} \text{ (για } g \rightarrow \infty) \end{aligned}$$

Πολλαπλασιάζοντας λοιπόν και τα δύο μέλη με $\sqrt{c_c(1-c_c)}$ τότε έχουμε ότι το μονοπάτι της εξέλιξης ρυθμίζεται από τον παράγοντα μέχρι το

$$\sqrt{\frac{2-c_c}{c_c}} \approx \frac{1}{\sqrt{c_c}} \quad (28)$$

λόγω των θετικών του συσχετίσεων και ή, αντίστοιχα, ο αντίθετος, λόγω των αρνητικών συσχετίσεων.

Με το $\sqrt{n} \leq 1/c_c$ ο αριθμός των αξιολογήσεων της συνάρτησης που απαιτούνται για την προσαρμογή ενός σχεδόν βέλτιστης πίνακα συνδιακύμανσης σε αντικειμενικές συναρτήσεις τύπου CIGAR (Concurrent and Interleaving Goal and Activity Recognition) γίνεται $O(n)$ παρά τον ρυθμό εκμάθησης που είναι $c_1 \approx 2/n^2$. Μια ερμηνεία αυτού του αποτελέσματος είναι διπλή. Πρώτον, ο επιθυμητός άξονας αναπαρίσταται στην πορεία με πολύ μεγαλύτερη ακρίβεια απ'ότι σε μεμονωμένα βήματα. Δεύτερον, ο ρυθμός εκμάθησης c_1 είναι διαμορφωμένος έτσι ώστε το αυξημένο μήκος της διαδρομής εξέλιξης όπως υπολογίζεται στο (29) ενεργεί στην πραγματικότητα παρόμοιο με ένα αυξημένο ρυθμό εκμάθησης με συντελεστή μέχρι το $c_c^{-1/2}$.

3.4 Συνδυάζοντας την αναβάθμιση μ-βαθμού και την cumulation

Σύμφωνα με τον (πηγή) το τελευταίο βήμα για την αναβάθμιση του πίνακα συνδιακύμανσης στην CMA είναι ο συνδυασμός των εξισώσεων (16) και (27) με βάση τον οποίο έχουμε:

$$C^{(g+1)} = \underbrace{\left(1 - c_1 - c_\mu \sum w_i\right)}_{0 \text{ ή κοντά στο } 0} C^{(g)} + c_1 \underbrace{p_c^{(g+1)} p_c^{(g+1)T}}_{\text{αναβάθμιση } 1\text{-βαθμού}} + c_\mu \underbrace{\sum_{i=1}^{\lambda} w_i y_{i:\lambda}^{(g+1)} \left(y_{i:\lambda}^{(g+1)}\right)^T}_{\text{αναβάθμιση } n\text{-βαθμού}} \quad (29)$$

όπου,

$$c_1 \approx 2/n^2$$

$$c_\mu \approx \min\left(\frac{\mu_{eff}}{n^2}, 1 - c_1\right)$$

$$y_{i:\lambda}^{(g+1)} \approx (x_{i:\lambda}^{(g+1)} - m^{(g)})/\sigma^{(g)}$$

$$\sum_{t=1}^{\lambda} w_t \approx -c_1/c_\mu$$

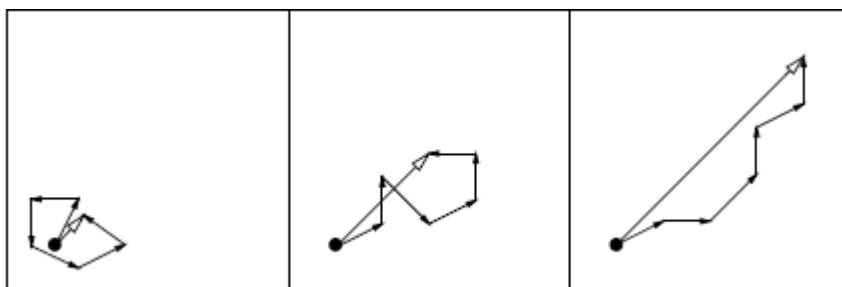
Είναι προφανές ότι για $c_1 = 0$ στην εξίσωση (29) προκύπτει η εξίσωση (16) και για $c_\mu = 0$ προκύπτει η (27). Επίσης, με βάση την μορφή της εξίσωσης μας παρατηρούμε ότι αποτελεί έναν συνδυασμό των αναβαθμίσεων 1^{00} και μ -ου βαθμού. Συγκεκριμένα, από την αναβάθμιση μ -ου βαθμού παίρνουμε πληροφορίες από όλο τον πληθυσμό ενώ από το μονοπάτι εξέλιξης από την αναβάθμιση 1^{00} βαθμού παίρνουμε τις πληροφορίες σχετικά με τις συσχετίσεις μεταξύ των γενεών. Όπως έχει ήδη ειπωθεί η 1^1 αναβάθμιση αποτελεί πολύ σημαντικό παράγοντα στους μεγάλους πληθυσμούς ενώ η 2^1 για τους μικρούς.

3.5 Έλεγχος βήματος

Ένα από τα πιο σημαντικά κομμάτια του κώδικα αφορά το κομμάτι του έλεγχου των βημάτων σε κάθε επανάληψη.

Η προσαρμογή του πίνακα συνδιακύμανσης, που συζητήθηκε στο τελευταίο τμήμα, δεν ελέγχει ρητά τη "συνολική κλίμακα" της διανομής που είναι ο έλεγχος βήμα. Η προσαρμογή πίνακα συνδιακύμανσης αυξάνει ή μειώνει την κλίμακα μόνο σε μία μόνο κατεύθυνση για κάθε επιλεγμένο βήμα - ή μειώνει την κλίμακα εξαφανίζοντας παλιές πληροφορίες από έναν δεδομένο, μη προσαρμοστικό παράγοντα. Λιγότερο ανεπίσημα, έχουμε δύο ειδικούς λόγους για να εισαγάγουμε έναν έλεγχο μεγέθους βημάτων επιπλέον του κανόνα προσαρμογής (29) για το $C^{(g)}$.

1. Το βέλτιστο συνολικό μήκος βήματος δεν μπορεί να προσεγγιστεί με (30), ιδιαίτερα εάν το meff επιλέγεται μεγαλύτερο από ένα.



Σχήμα 12: Τρεις διαδρομές εξέλιξης αντίστοιχα έξι βημάτων από διαφορετικές καταστάσεις επιλογής (ιδεαλισμένη). Τα μήκη των μεμονωμένων βημάτων είναι όλα συγκρίσιμα. Το μήκος της εξέλιξης οι διαδρομές (άθροισμα των βημάτων) είναι αξιοσημείωτα διαφορετικές και εκμεταλλεύονται για έλεγχο μεγέθους βημάτων

Για παράδειγμα, στο $f_{sphere}(x) = \sum_{i=1}^n x_i^2$, δεδομένου ότι $C^{(g)} = I$ και $\lambda \leq n$, το βέλτιστο μέγεθος βήματος σ ισούται προσεγγιστικά με $\mu\sqrt{f_{sphere}(x)/n}$ με ίσο ανασυνδυασμό βάρη [4, 29], και $1,25\mu\sqrt{f_{sphere}(x)/n}$ με βέλτιστα βάρη ανασυνδυασμού [2]. Αυτή η εξάρτηση από μ ή μ_{eff} δεν μπορεί να πραγματοποιηθεί με (16) ή (30).

2. Ο μεγαλύτερος αξιόπιστος ρυθμός εκμάθησης για την αναβάθμιση του πίνακα συντεταγμένων στο (30) είναι πολύ αργός για την επίτευξη ανταγωνιστικών ποσοστών αλλαγής για το συνολικό μήκος βήματος.

Για να πετύχουμε την βέλτιστη απόδοση σε f_{sphere} με μία στρατηγική εξέλιξης με σταθμισμένο ανασυνδυασμό, το συνολικό μήκος βήματος πρέπει να μειωθεί κατά ένα συντελεστή περίπου $\exp(0,25) \approx 1.28$ σε n λειτουργίες αξιολόγησης. Δηλαδή, ο χρονικός ορίζοντας για την αλλαγή βήματος πρέπει να είναι ανάλογος σε n ή μικρότερο. Από τους ρυθμούς εκμάθησης c_1 και c_μ στο (30) ακολουθεί ότι το η προσαρμογή είναι πολύ αργή για να γίνει ανταγωνιστική σε κάθε περίπτωση που ισχύει ότι $\mu_{eff} \ll n$. Αυτό μπορεί να επικυρωθεί με προσομοιώσεις ακόμη και για μέτριες διαστάσεις, $n \geq 10$ και μικρό $\mu_{eff} \leq 1 + \ln n$.

Για να ελέγξουμε το μέγεθος του βήματος $\sigma^{(g)}$ χρησιμοποιούμε μια διαδρομή εξέλιξης, δηλαδή ένα άθροισμα των διαδοχικών βημάτων που αποτελεί μέθοδο μπορεί να εφαρμοστεί ανεξάρτητα από την αναβάθμιση του πίνακα συνδιακύμανσης και υποδηλώνεται ως σωρευτική προσαρμογή μήκους βήματος (ή Cumulative step length adaptation CSA). Σύμφωνα με αυτήν, αξιοποιείται το μήκος της διαδρομής εξέλιξης, ακολουθώντας την εξής συλλογιστική (βλ. Σχήμα 5).

- Όταν η διαδρομή εξέλιξης είναι σύντομη, τα μεμονωμένα βήματα ακυρώνουν το ένα το άλλο (Εικ. 5, αριστερά) και έχουν αρνητική συσχέτιση.
- Όταν η διαδρομή εξέλιξης είναι μεγάλη, τα μεμονωμένα βήματα δείχνουν προς παρόμοιες κατευθύνσεις (Εικ. 5, δεξιά) και έχουν συσχέτιση.
- Στην επιθυμητή κατάσταση τα βήματα είναι (περίπου) αντίθετα σε προσδοκία και συνεπώς ασυσχέτιστα (Σχήμα 5, μέση).

Για να αποφασίσουμε αν η διαδρομή εξέλιξης είναι "μεγάλη" ή "σύντομη", συγκρίνουμε το μήκος της διαδρομής με το αναμενόμενο μήκος του κάτω από την τυχαία επιλογή, όπου τα διαδοχικά βήματα είναι ανεξάρτητα και ως εκ τούτου ασύνδετα (μη συσχετισμένα βήματα είναι η επιθυμητή κατάσταση). Εάν η επιλογή προκαθοριστεί η πορεία της εξέλιξης να είναι πλέον αναμενόμενη, αυξάνεται, και, αντιστρόφως, εάν η επιλογή υποβαθμίζει η πορεία της εξέλιξης να είναι μικρότερη από την αναμενόμενη; μειώνεται. Στην ιδανική κατάσταση, επιλογή δεν επηρεάζει το μήκος της διαδρομής εξέλιξης και το μήκος είναι ίσο με το αναμενόμενο μήκος κάτω από τυχαία επιλογή.

Στην πράξη, για να κατασκευαστεί η πορεία εξέλιξης, p_σ , εφαρμόζονται οι ίδιες τεχνικές όπως στο (24). Σε αντίθεση με το (24), κατασκευάζεται μια διαδρομή συζευγμένης εξέλιξης, επειδή το αναμενόμενο μήκος της πορείας εξέλιξης p_c από την (24) εξαρτάται από την κατεύθυνσή του (σύγκριση (25)), την οποία αρχικοποιούμε με $p_\sigma^{(0)} = 0$, και έτσι η διαδρομή συσχέτισης εξέλιξης είναι η εξής, σύμφωνα με τον Hansen [7]:

$$p_\sigma^{(g+1)} = (1 - c_\sigma)p_\sigma^{(g)} + \sqrt{c_\sigma(2 - c_\sigma)\mu_{eff}}C^{(g)^{-\frac{1}{2}}}\frac{m^{(g+1)} - m^{(g)}}{\sigma^{(g)}} \quad (30)$$

όπου

$$p_\sigma^{(g)} \in \mathbb{R}^n$$

$$c_\sigma < 1$$

$\sqrt{c_\sigma(2 - c_\sigma)\mu_{eff}}$ είναι μία σταθερά κανονικοποίησης

$$C^{(g)^{-\frac{1}{2}}} = B^{(g)}D^{(g)^{-1}}B^{(g)T}$$

Προκειμένου να αναβαθμίσουμε το $\sigma^{(g)}$, συγκρίνουμε το $\|p_\sigma^{(g+1)}\|$ με το $E\|\mathcal{N}(0, I)\|$ μέσω της:

$$\ln\sigma^{(g+1)} = \ln\sigma^{(g)} + \frac{c_\sigma}{d_\sigma} \left(\frac{\|p_\sigma^{(g+1)}\|}{E\|\mathcal{N}(0, I)\|} - 1 \right) \quad (31)$$

Όπου,

$d_\sigma \approx 1$, παράμετρος απόρριψης

$$E\|\mathcal{N}(0, I)\| = \sqrt{2}\Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)$$

Εναλλακτικά θα μπορούσαμε να χρησιμοποιήσουμε την νόρμα του $p_\sigma^{(g+1)}$ στο τετράγωνο στην (31) και έτσι να προέκυπτε:

$$\ln\sigma^{(g+1)} = \ln\sigma^{(g)} + \frac{c_\sigma}{2d_\sigma} \left(\frac{\|p_\sigma^{(g+1)}\|^2}{n} - 1 \right) \quad (32)$$

Αυτή η αναβάθμιση έχει παρόμοια αποτελέσματα με την (31), ενώ προφανώς οδηγεί σε ταχύτερα βήματα μεγέθους βημάτων και βραδύτερες μειώσεις μεγέθους βημάτων.

Η μεταβολή του βήματος μεγέθους είναι αμερόληπτη στην κλίμακα καταγραφής, επειδή $E[\ln\sigma^{(g+1)}|\sigma^{(g)}] = \ln\sigma^{(g+1)}$ για $p_\sigma^{(g+1)} \sim N(0, I)$.

Προκειμένου να δείξουμε ότι τα επιτυχημένα βήματα είναι προσεγγιστικά αναστροφοσυζυγής με το $C^{(g)-1}$ τονίζουμε ότι ισχύουν οι σχέσεις (30) και (31) και υιοθετούμε σ τέτοιο ώστε το μήκος του $p_\sigma^{(g+1)}$ να ισούται με το $E\|\mathcal{N}(0, I)\|$. Έχοντας ότι $(E\|\mathcal{N}(0, I)\|)^2 \approx \|p_\sigma^{(g+1)}\|^2 = p_\sigma^{(g)T} p_\sigma^{(g+1)} = RHS^T RHS$ από την (31) υποθέτουμε ότι η αναμενόμενη ρίζα του μήκους του $C^{(g)-\frac{1}{2}}(m^{(g+1)} - m^{(g)})$ είναι

$$p_\sigma^{(g)T} C^{(g)-\frac{1}{2}}(m^{(g+1)} + m^{(g)}) \approx 0 \quad (33)$$

και

$$\left(C^{(g)\frac{1}{2}} p_\sigma^{(g)} \right)^T C^{(g)-1} (m^{(g+1)} - m^{(g)}) \approx 0 \quad (34)$$

Δεδομένου ότι $1/(c_1 + c_2) \gg 1$ λόγω της σχέσης (34) υποθέτουμε ακόμα ότι $p_\sigma^{(g)T} C^{(g)-\frac{1}{2}}(m^{(g+1)} + m^{(g)}) \approx 0$ και έτσι έχουμε ότι:

$$(m^{(g+1)} - m^{(g)})^T C^{(g)-1} (m^{(g+1)} - m^{(g)}) \approx 0 \quad (35)$$

Και έτσι το βήμα που πήραμε από την μέση τιμή της κατανομή είναι σχεδόν αναστροφοσυζυγές με το $C^{(g)-1}$.

Τέλος, επειδή το $\sigma^{(g)} > 0$, η (32) είναι ισοδύναμη με την

$$\sigma^{(g+1)} = \sigma^{(g)} \exp \left(\frac{c_\sigma}{d_\sigma} \left(\frac{\|p_\sigma^{(g+1)}\|}{E\|\mathcal{N}(0, I)\|} - 1 \right) \right) \quad (36)$$

Το μήκος της διαδρομής εξέλιξης είναι ένα διαισθητικό και εμπειρικά αποδεκτό μέτρο καλής συμπεριφοράς για το συνολικό μήκος βήματος. Για $\mu_{eff} > 1$ είναι το καλύτερο μέτρο που γνωρίζουμε.

Παρ' όλα αυτά, δεν προσαρμόζει σχεδόν βέλτιστα μεγέθη βημάτων σε πολύ noisy αντικειμενικές συναρτήσεις, δηλαδή σε συναρτήσεις στις οποίες γνωρίζουμε μόνο σημεία τους.

3.6 Βήματα του Αλγορίθμου

Βασιζόμενος λοιπόν στην παραπάνω μαθηματική σκέψη, δημιουργήθηκε η CMA με σκοπό την μέγιστη απόδοση στην βελτιστοποίηση και ακολουθήθηκαν τα εξής βήματα προκειμένου να κατασκευαστεί:

Βήμα 1^ο: Ορισμός της συνάρτησης και αρχικοποίηση των μεταβλητών

Αρχικά, ο Hansen φροντίζει να ορίσει την συνάρτηση, μαζί με τις διαστάσεις τις, που θα βελτιστοποιήσει ενώ επιπλέον αρχικοποιεί και κάποιες μεταβλητές εκ των οποίων άλλες θα παραμείνουν σταθερές καθ' όλη την διάρκεια της διαδικασίας (π.χ. **stopfitness**) ενώ άλλες θα μεταβληθούν (π.χ. **xmean**). Σημαντικό σημείο του κώδικα είναι η ανάθεση των μεταβλητών **stopfitness** και **stopeval** αφού αυτές θα καθορίσουν το μέγεθος των επαναλήψεων της διαδικασίας βελτιστοποίησης. Συγκεκριμένα, η μεταβλητή **stopfitness** καθορίζει το κατά πόσο έχουμε πλησιάσει στο σωστό αποτέλεσμα, ενώ η **stopeval** δίνει τον μέγιστο αριθμό επαναλήψεων στην διαδικασία μας. Έτσι, η επανάληψη σταματάει είτε άμα προσεγγίσουμε το αποτέλεσμα μας πολύ καλά, είτε άμα περάσει ένας συγκεκριμένος αριθμός επαναλήψεων.

Βήμα 2^ο:

Μετά τον ορισμό των μεταβλητών μας, περνάμε στο κύριο μέρος στον οποίο ουσιαστικά γίνεται ο υπολογισμός των αποτελεσμάτων μας. Αρχικά, γίνεται η διαδικασία με την οποία δημιουργείται ο απόγονος και μέσω της **arfitness** αξιολογείται. Να τονίσουμε εδώ πως ο απόγονος προκύπτει τυχαία αλλά έχοντας πάντα σημείο αναφοράς τους προγόνους του με βάση την κανονική κατανομή γύρω από την **xmean**. Έτσι, αφού αξιολογηθούν οι απόγονοι, θα τοποθετηθούν σε φθίνουσα σειρά και θα προκύψει νέα **xmean**. Εν συνεχεία, θα αναβαθμιστούν τα μονοπάτια της

εξέλιξης ούτως ώστε να προκύψει ο νέος πίνακας συνδιακύμανσης C καθώς και το νέο βήμα sigma.

Βήμα 3^ο:

Τελευταία ενέργεια στον CMAES είναι η παράθεση των αποτελεσμάτων σε αριθμητική μορφή καθώς επίσης και σε μορφή διαγραμμάτων. Συγκεκριμένα, όσον αφορά τα αριθμητικά αποτελέσματα παρατίθενται πίνακες με τα αποτελέσματα που έχουμε ενώ γραφικά έχουμε στο ένα διάγραμμα την fitness function το sigma (σ) και τις ιδιοτιμές ενώ στο άλλο έχουμε την πορεία της xmean σε σχέση με τις επαναλήψεις. Τα τελευταία δεδομένα μας βοηθούν ιδιαίτερα αφού μπορούμε να βλέπουμε με τον ρυθμό των επαναλήψεων την πορεία των δεδομένων και γραφικά, αποφεύγοντας να παίρνουμε σαν σημείο αναφοράς τον τεράστιο όγκο δεδομένων που έχουμε σαν αριθμητικά αποτελέσματα.

Κεφάλαιο 4: Επίλυση Προβλημάτων μέσω του CMAES

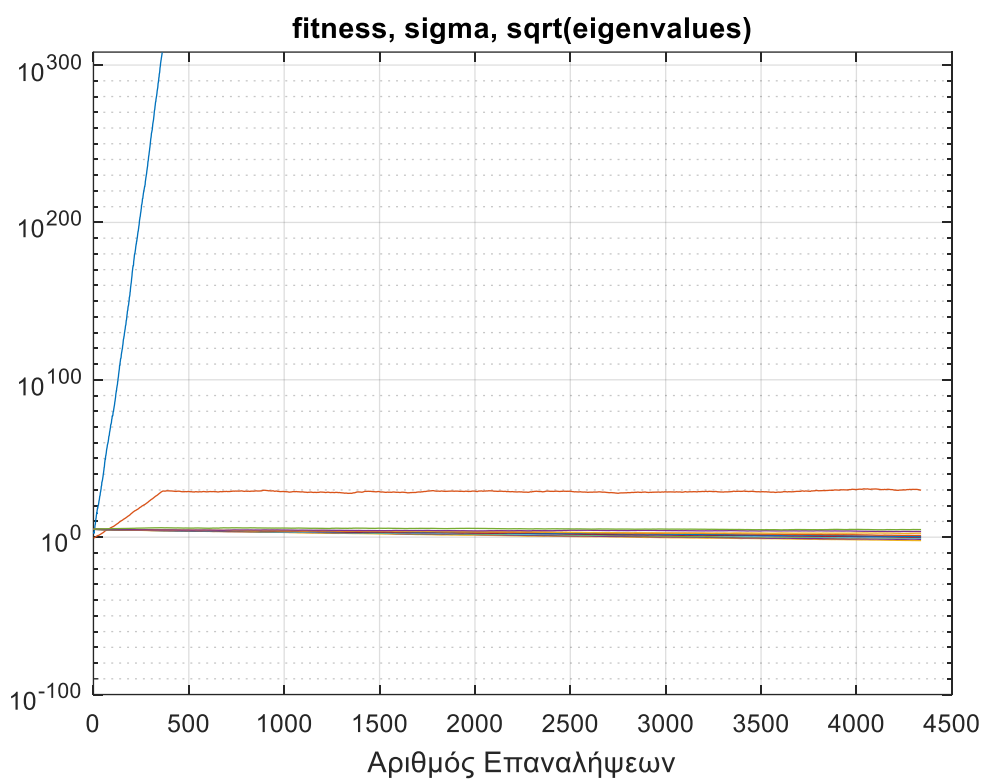
Με την χρήση του αλγορίθμου επιχειρήσαμε να επιλύσουμε πολλά και διαφορετικά προβλήματα ούτως ώστε να είμαστε βέβαιοι για την εγκυρότητά του. Πιο συγκεκριμένα, επιχειρήσαμε να λύσουμε 13 προβλήματα τα οποία και είχαν τεθεί από τους Thomas Runarsson and Xin Yao [16] ούτως ώστε να υπάρχει δυνατότητα ελέγχου της CMA πάνω σε πολύπλοκα προβλήματα πολλών διαστάσεων. Βέβαια, πρέπει να τονιστεί εδώ ότι στην διαδικασία της επίλυσης δεν λαμβάνονται υπ' όψιν οι περιορισμοί του κάθε πρόβληματος.

Έτσι, λαμβάνοντας δεδομένου ότι ο αλγόριθμος είναι στοχαστικός, τρέξαμε τα προβλήματα 6 φορές προκειμένου να εξάγουμε επιπλέον συμπεράσματα για την σύγκλιση των συναρτήσεων μας. Θυμίζουμε ότι ο αλγόριθμος αυτός αποτελεί μια στοχαστική διαδικασία γεγονός που θα πει ότι τα αποτελέσματά μας σε διάφορες εκτελέσεις του μπορεί να διαφέρουν. Έτσι, θα παρουσιάσουμε τα αποτελέσματά μας αναλύοντας στατιστικά τα διάφορα αποτελέσματα του τελευταίου σωματιδίου με μεγαλύτερη fitness function από όλες τις φορές που τρέξαμε τον αλγόριθμο και παρουσιάζοντας τα διαγράμματα μίας εκτέλεσης του αλγορίθμου μαζί με έναν σχολιασμό τους.

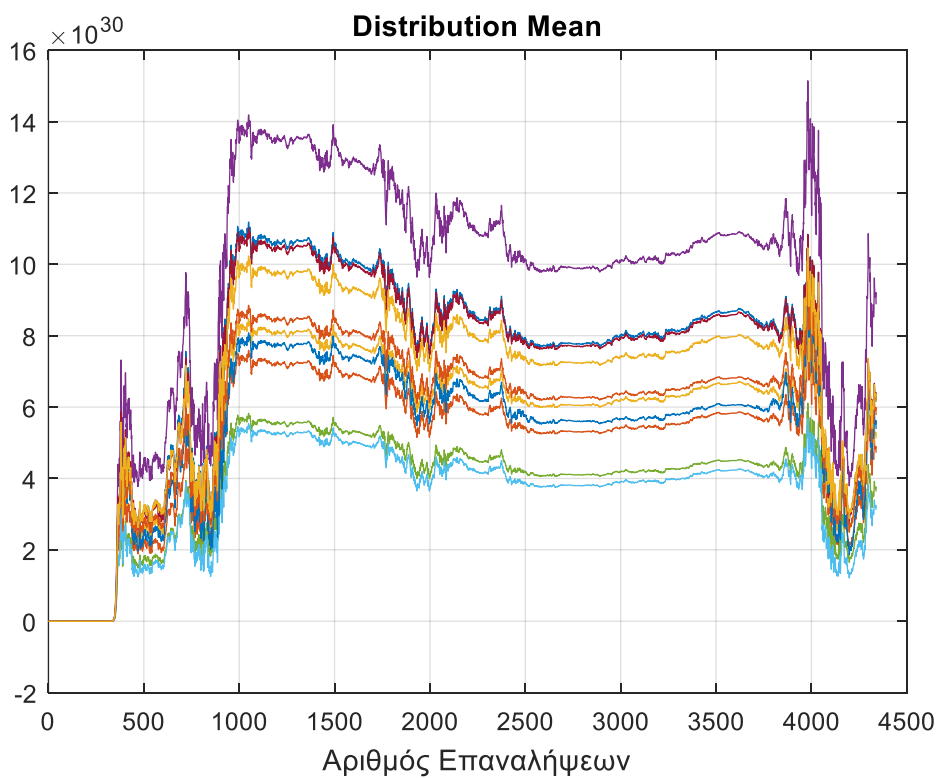
Κατ' αυτόν τον τρόπο θα μπορέσουμε να δούμε και το κατά πόσο διαφέρουν τα διάφορα αποτελέσματα, και να δούμε σχετικά διαγράμματα από την τελευταία εκτέλεση του αλγορίθμου μας. Προφανώς ευελπιστούμε σε όσο δυνατόν μικρότερες αποκλίσεις κάθε φορά.

4.1 Προσέγγιση Λύσεων Προβλημάτων

Πρόβλημα 1: $\min f(x)$ με $f(x) = \sqrt{n} \prod_{i=1}^n x_i$ για $n = 10$



Διάγραμμα 1



Διάγραμμα 2

Μέση Τιμή	<i>inf</i>
Ελάχιστο	-1.670×10^{308}
Μέγιστο	-7.311×10^{307}
Τυπική Απόκλιση	<i>inf</i>

Πίνακας 1: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

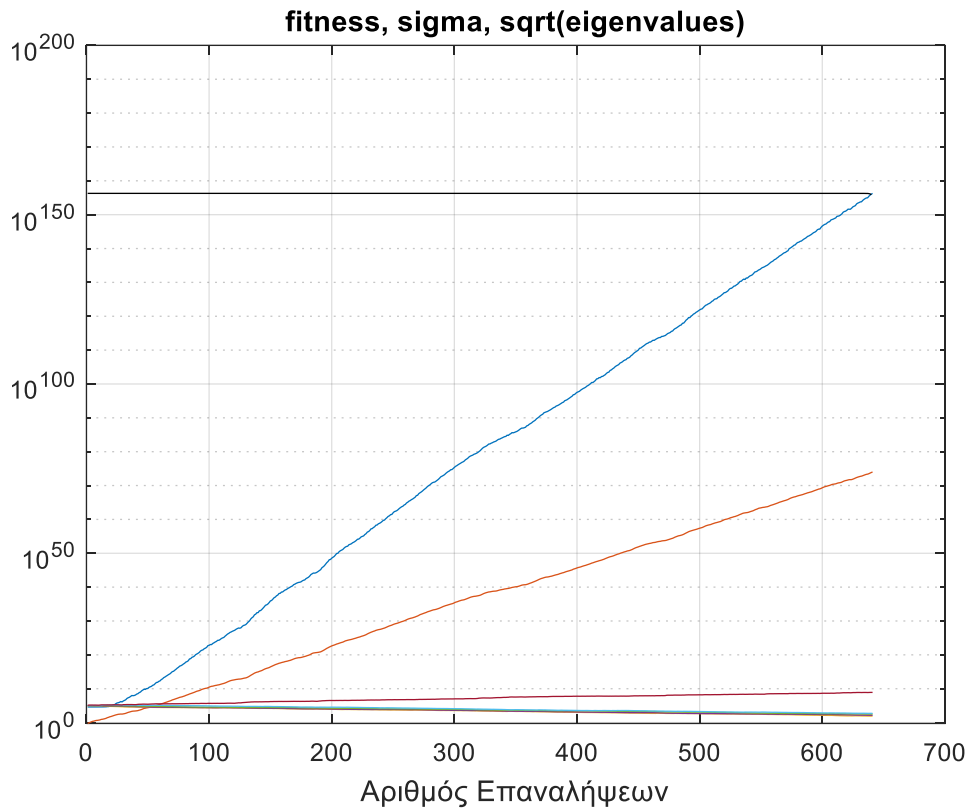
Σχολιασμός: Όπως προαναφέραμε, το γεγονός ότι η μέση τιμή και η τυπική απόκλιση παίρνουν τιμές στο άπειρο, σημαίνει ότι ο υπολογισμός γίνεται για πολύ μεγάλους αριθμούς για το matlab, το οποίο είναι λογικό άμα παρατηρήσει κανείς το μέγιστο και το ελάχιστο στοιχείο που προκύπτει από την στατιστική μας ανάλυση για το τελευταίο σωματίδιο σε κάθε επανάληψη του αλγορίθμου.

Από το 1^ο διάγραμμα είναι φανερό ότι κάποια σωματίδια παρεκκλίνουν των τιμών που έχουν τα άλλα στην fitness function.

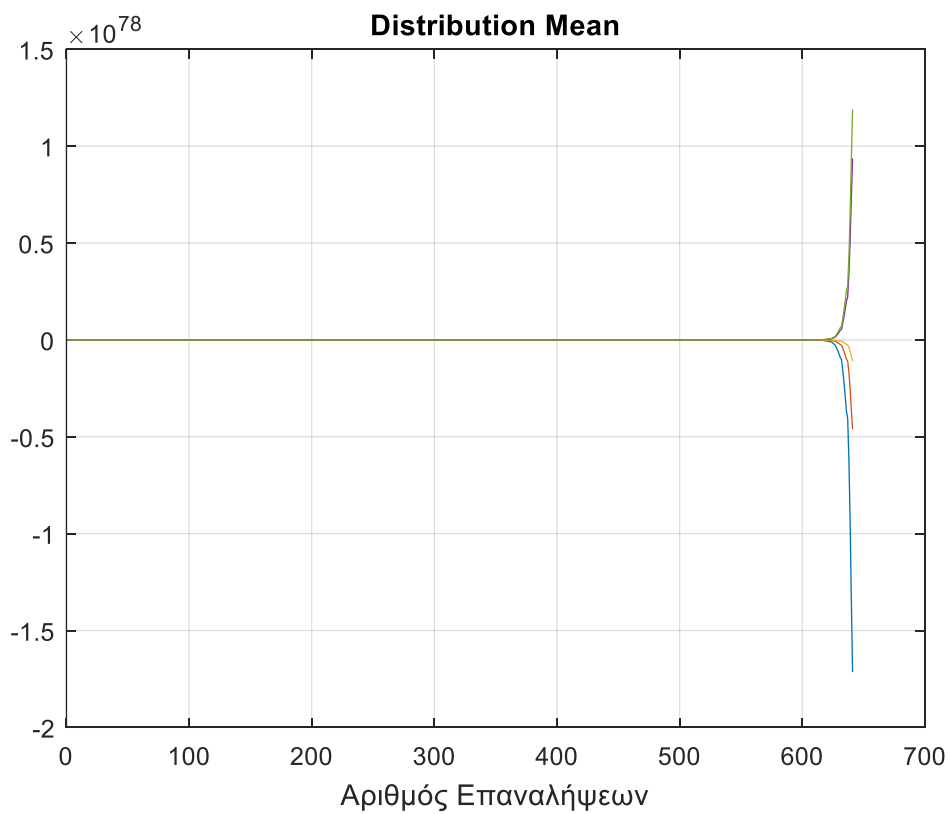
Παρόλα αυτά η πορεία της μέσης τιμής για όλα τα σωματίδια ακολουθεί παρόμοια πορεία σε όλες τις φάσεις ενώ μετά την 4000^η επανάληψη παρατηρούμε ότι το νούμερο σταθεροποιείται για όλα τα σωματίδια.

Όσον αφορά τα διαγράμματα, στο Σχήμα 5.1 παρατηρούμε ότι η fitness function έχει σταθερή κλίση (μπλε γραμμή) γεγονός που σημαίνει ότι η διαδικασία μας γίνεται σχετικά γρήγορα ενώ το sigma και το τετράγωνο των ιδιοτιμών δεν επιδέχονται μεγάλες αλλαγές ιδιαίτερα μετά την 300η, περίπου, επανάληψη ενώ παρατηρούμε κόλας από τον πίνακα των στατιστικών της fitness function ότι δέχεται πολύ μικρές τιμές ενισχύοντας έτσι το αποτέλεσμά μας.

Πρόβλημα 2: $\min f(x)$ με $f(x) = 5.3578547x_3^2 + 0.8356891x_1x_5 +$
 $+37.293239x_1 - 40792.141$



Διάγραμμα 4.3



Διάγραμμα 4.4

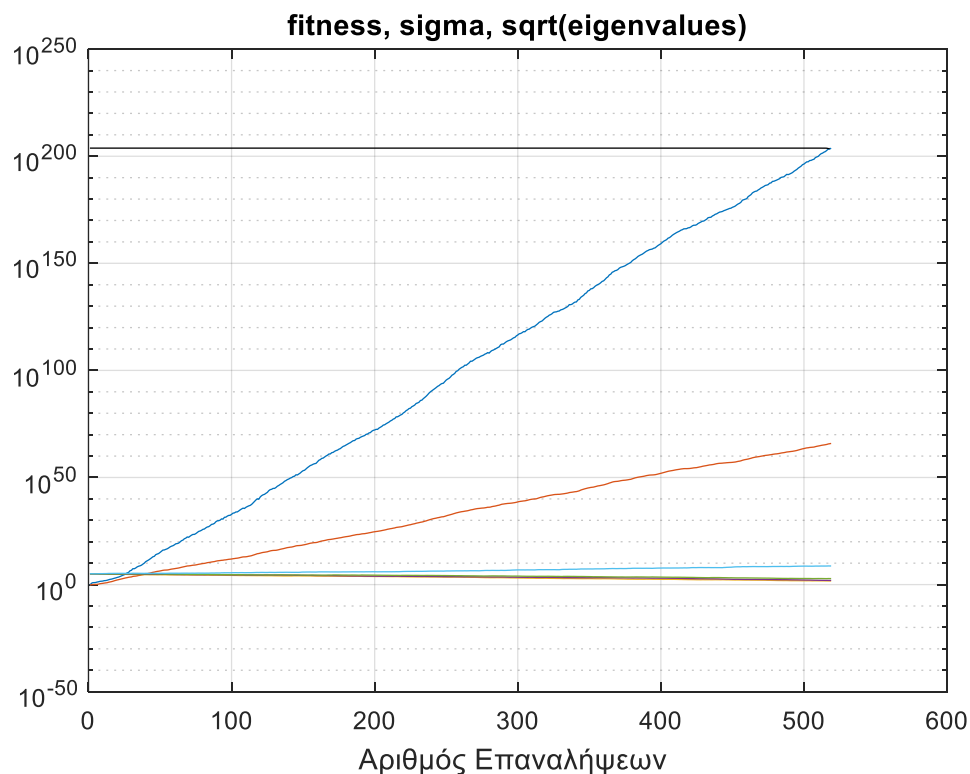
Μέση Τιμή	$-5,355 \times 10^{157}$
Ελάχιστο	$-3,194 \times 10^{158}$
Μέγιστο	$-3,220 \times 10^{123}$
Τυπική Απόκλιση	<i>inf</i>

Πίνακας 4.2: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

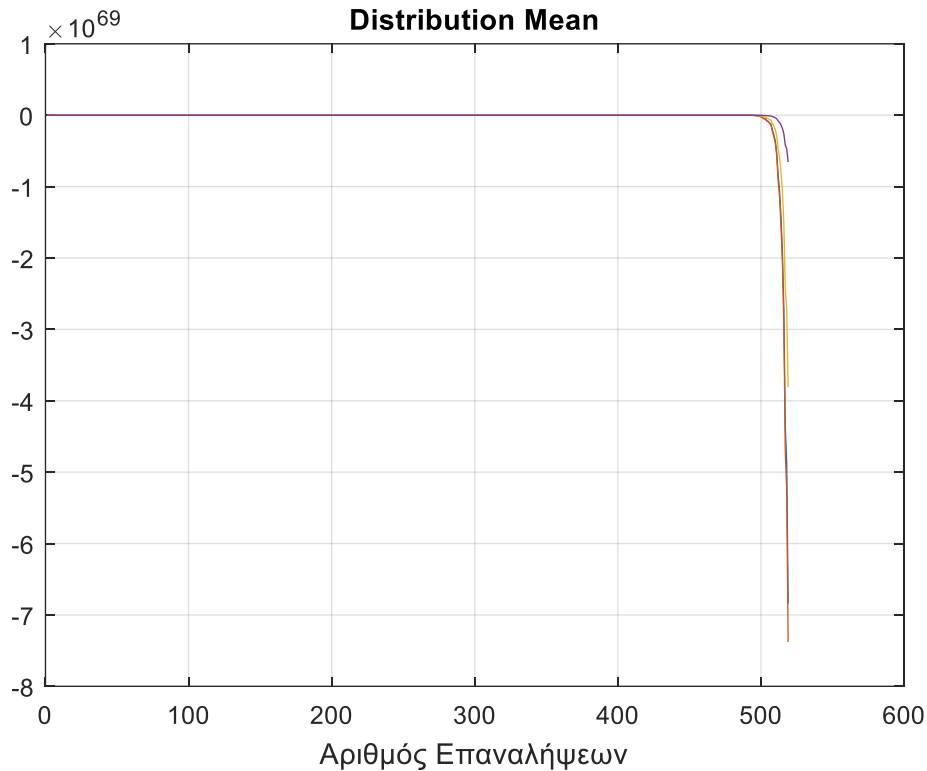
Σχολιασμός: Κατά την διάρκεια της διαδικασίας και πιο συγκεκριμένα λίγο μετά την 600^η επανάληψη βελτιστοποίησης στην συγκεκριμένη συνάρτηση παρατηρούμε απότομη αλλαγή στην μέση τιμή της εξελικτικής διαδικασίας των σωματιδίων. Όπως προαναφέραμε, το γεγονός ότι η μέση τιμή και η τυπική απόκλιση παίρνουν τιμές στο άπειρο, σημαίνει ότι ο υπολογισμός γίνεται για πολύ μεγάλους αριθμούς για το matlab, το οποίο είναι λογικό άμα παρατηρήσει κανείς το μέγιστο και το ελάχιστο στοιχείο που προκύπτει από την στατιστική μας ανάλυση.

Όσον αφορά τα διαγράμματα, στο Σχήμα 5.1 παρατηρούμε ότι η fitness function για δύο σωματίδια έχει σταθερά θετική κλίση μέχρι περίπου την 650^η επανάληψη όπου και σταθεροποιείται όπως φαίνεται και τον πίνακα όπου παρατηρούμε ότι η μέση τιμή της fitness function για τις επαναλήψεις μας δεν είναι στο άπειρο

Πρόβλημα 3: $\min f(x)$ με $f(x) = 3x_1 + 0,000001x_1^3 + (0,000002/3)x_2^3$



Διάγραμμα 4.5



Διάγραμμα 4.6

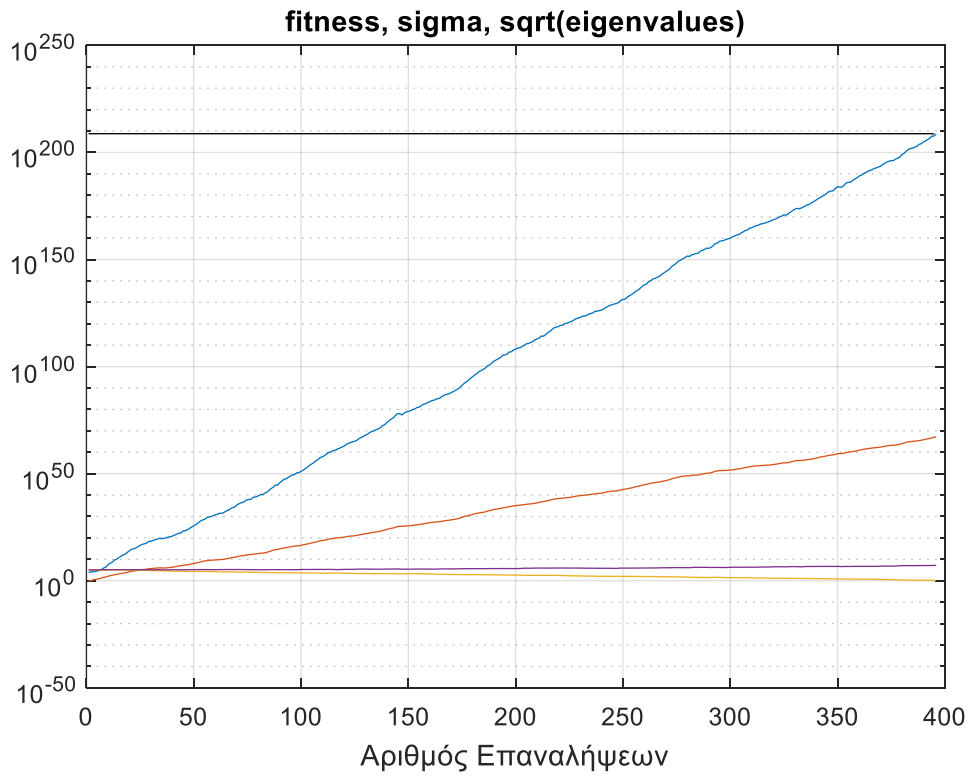
Μέση Τιμή	<i>inf</i>
Ελάχιστο	-1.814×10^{219}
Μέγιστο	-1.978×10^{189}
Τυπική Απόκλιση	<i>inf</i>

Πίνακας 4.3: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

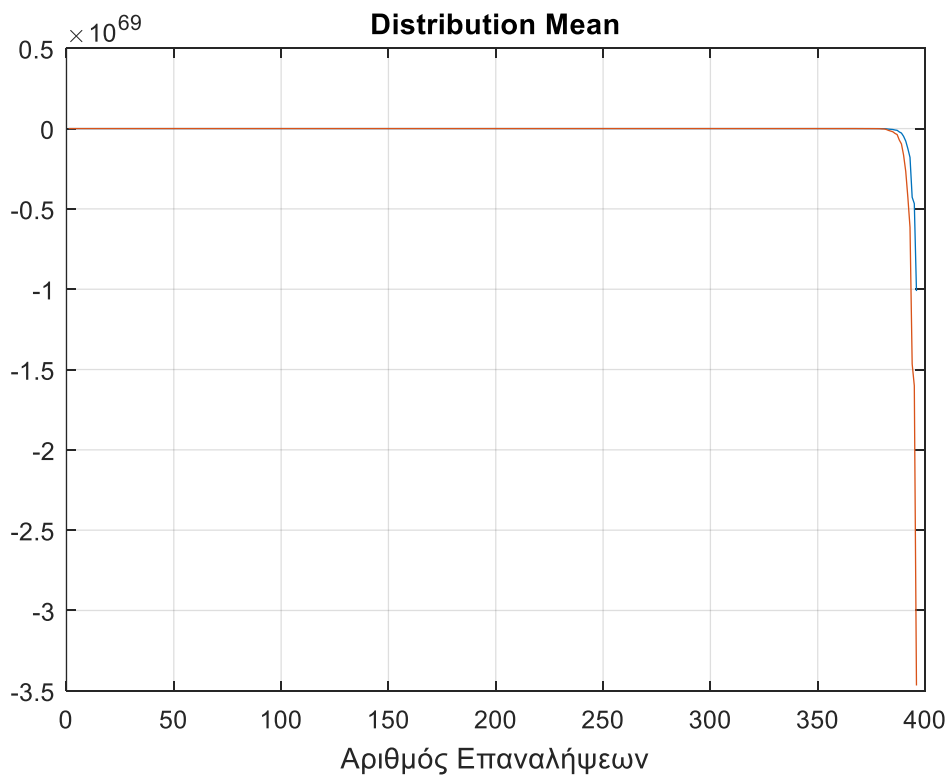
Σχολιασμός: Παρατηρούμε εδώ από το διάγραμμα 4.5 ότι η πορεία της fitness function είναι ίδια με του προβλήματος 2 αλλά έχουμε σημαντικές διαφορές στον τρόπο με τον οποίο εξελίσσεται η πορεία της μέσης τιμής της κατανομής για κάθε ένα σωματίδιο. Συγκεκριμένα, ενώ στο πρόβλημα 2 η πορεία έπειτα από μία συγκεκριμένη επανάληψη ήταν είτε προς τα θετικά είτε προς τα αρνητικά, εδώ παρατηρούμε ότι είναι μόνο προς τα αρνητικά.

Φαίνεται ότι αυτή η συμπεριφορά επηρεάζει και την fitness function όπως παρατηρούμε στον πίνακά μας αφού φαίνεται ότι απειρίζεται, γεγονός που γινόταν στο πρόβλημα 2

Πρόβλημα 4: $\min f(x)$ με $f(x) = (x_1 - 10)^3 + (x_2 - 20)^3$



Διάγραμμα 4.7



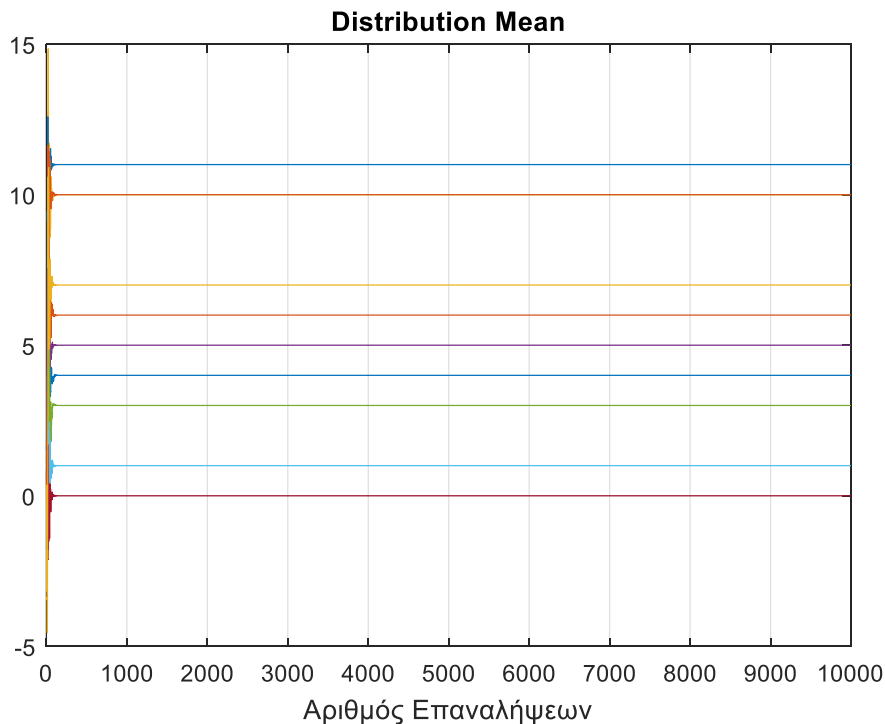
Διάγραμμα 4.8

Μέση Τιμή	<i>inf</i>
Ελάχιστο	-1.670×10^{308}
Μέγιστο	-7.311×10^{208}
Τυπική Απόκλιση	<i>inf</i>

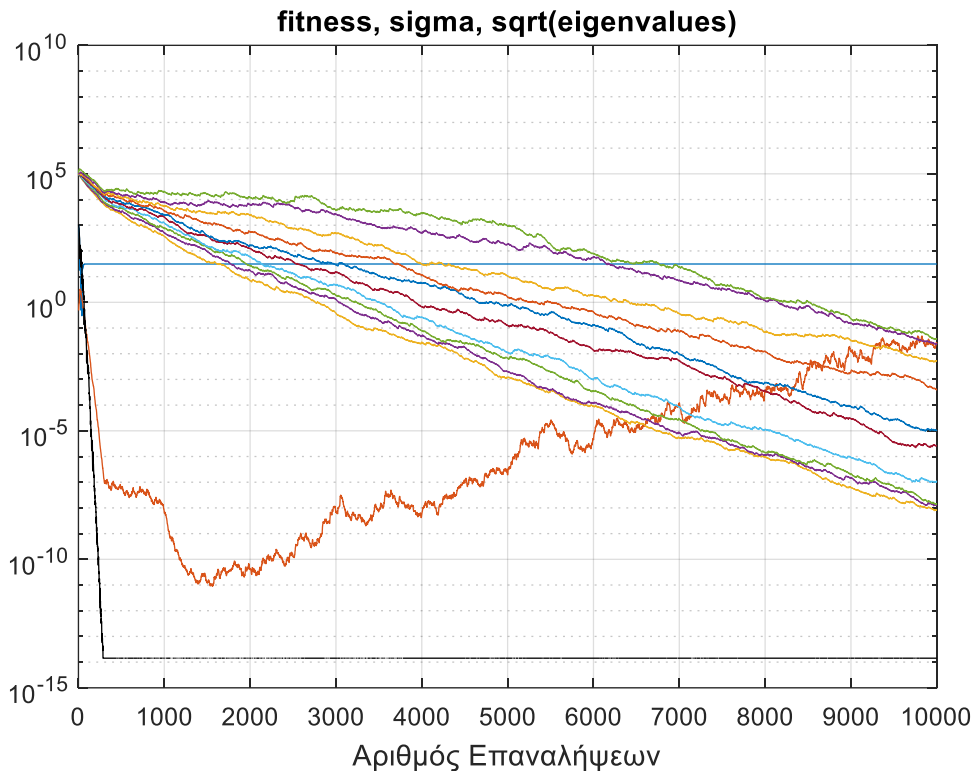
Πίνακας 4.4: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

Σχολιασμός: Με βάση τις παρατηρήσεις μας θα μπορούσαμε να βγάλουμε σχεδόν παρόμοια συμπεράσματα με το πρόβλημα 3 μόνο που σε αυτήν την περίπτωση παρατηρούμε ότι όλα τα φαινόμενα εξελίσσονται πιο γρήγορα. Συγκεκριμένα ενώ στο 3^ο πρόβλημα παρατηρούμε τα φαινόμενα αυτά στην 500^η επανάληψη, σε αυτήν την περίπτωση συμβαίνουν στην 400^η.

Πρόβλημα 5: $\min f(x)$ με $f(x) = x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45$



Διάγραμμα 4.9



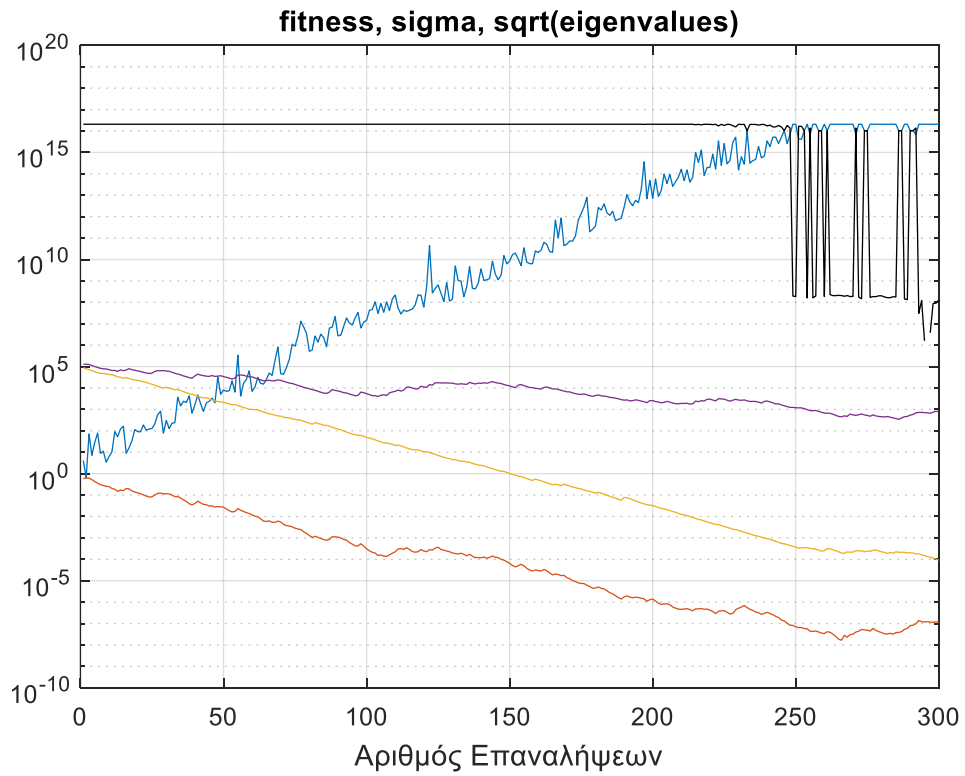
Διάγραμμα 4.10

Μέση Τιμή	-31
Ελάχιστο	-31
Μέγιστο	-31
Τυπική Απόκλιση	0

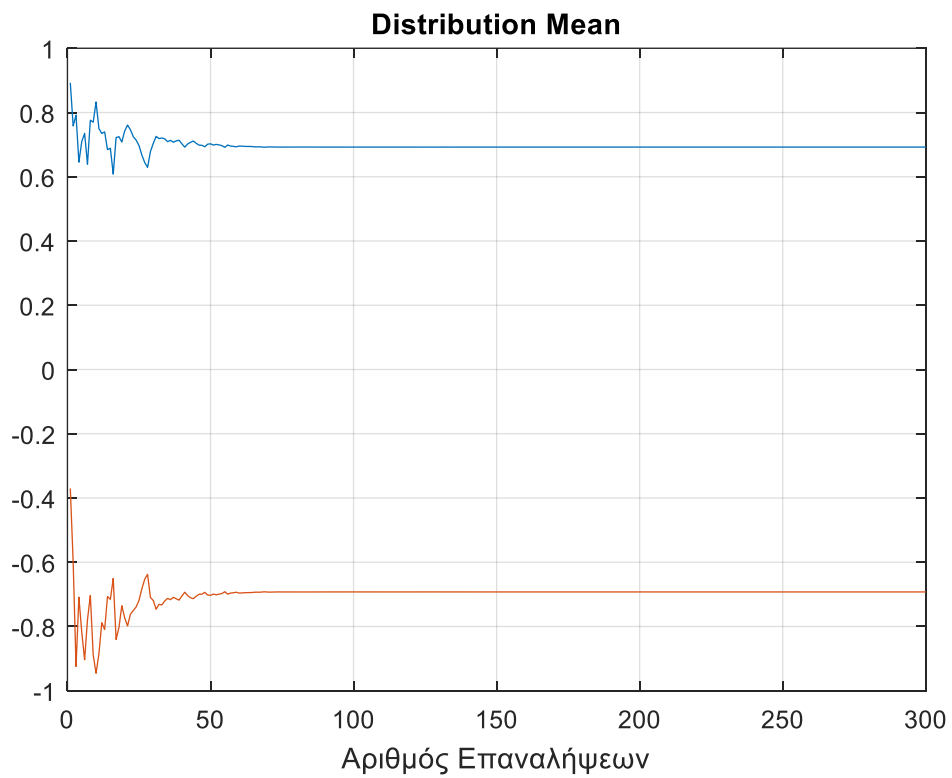
Πίνακας 4.5: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

Σχολιασμός: Με βάση τα διαγράμματά μας και τον πίνακα θα μπορούσαμε εύκολα να πούμε πως κάθε σωματίδιο ακολουθεί σταθερή πορεία κατά την διάρκεια της διαδικασίας μας και εύκολα να υποθέσουμε ότι η ελάχιστη τιμή της είναι το -31 αφού και η απόκλιση στις 5 επαναλήψεις είχε το ίδιο αποτέλεσμα. Συγκεκριμένα, η σταθερότητα της πορείας, πέρα από τον πίνακα φαίνεται και όσο από την πορεία της μέσης τιμής της κατανομής κάθε σωματιδίου που παραμένει σταθερή όσο και από την σταθερότητα της fitness function μετά από ένα σημείο στην διάρκεια της διαδικασίας μας (περίπου 200) μέχρι το οποίο όμως φθίνει με σταθερή κλίση.

Πρόβλημα 6: $\max f(x)$ με $f(x) = \frac{\sin^3(2\pi x_1)\sin(2\pi x_2)}{x_1^3(x_1+x_2)}$



Διάγραμμα 4.11



Διάγραμμα 4.12

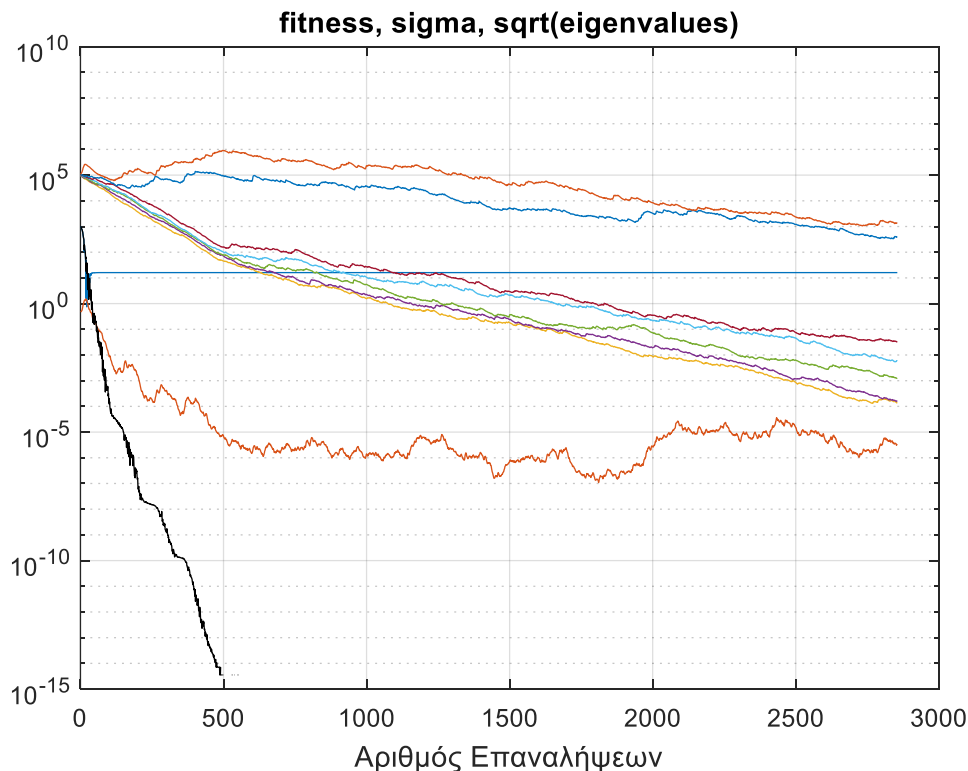
Μέση Τιμή	-2.179×10^{18}
Ελάχιστο	-9.024×10^{18}
Μέγιστο	-0.0175
Τυπική Απόκλιση	3.720×10^{18}

Πίνακας 4.6: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

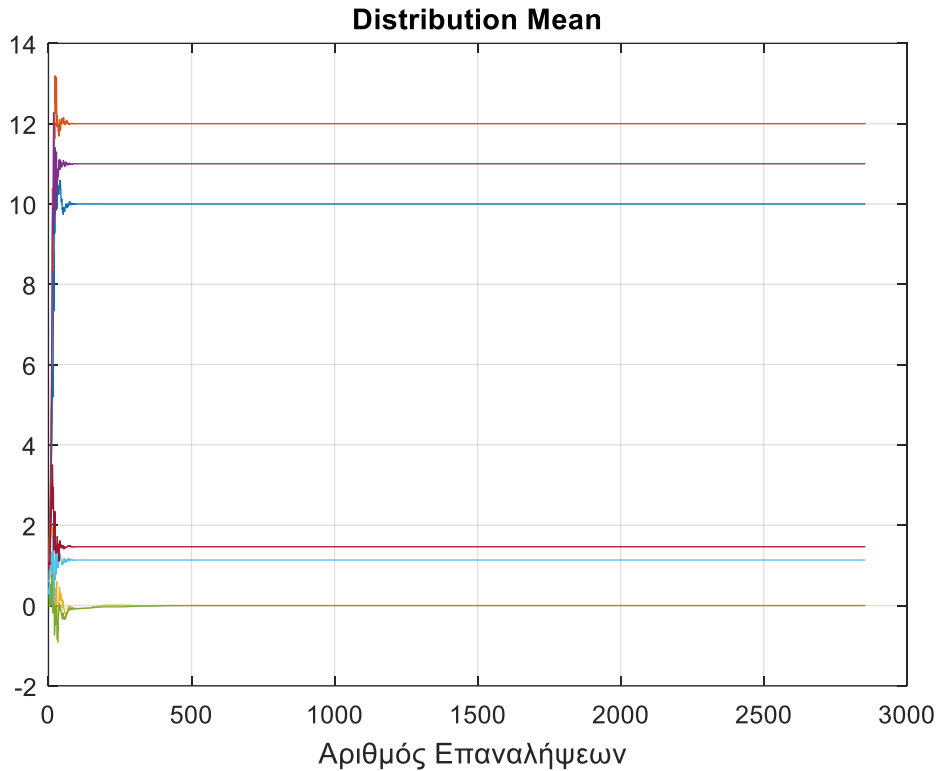
Σχολιασμός: Όπως είναι φανερό, η μέση τιμή της κατανομής για κάθε σωματίδιο ακολουθεί μία πανομοιότυπη κίνηση μέχρι να σταθεροποιηθεί στην 50^η επανάληψη. Επίσης παρατηρούμε πως μειώνεται σταδιακά η fitness function στο διάγραμμα (κίτρινη γραμμή) ενώ η τυπική απόκλιση του βήματος και το τετραγώνον των ιδιοτιμών (μπλε και σκούρα μπλε γραμμή) κινούνται άτακτα καθ'όλη την διάρκεια της διαδικασίας μας.

Γενικά, είναι δύσκολο να εξάγουμε συμπέρασμα περί αριθμού σύγκλισης αφού και η τυπική απόκλιση στον πίνακα μας είναι πολύ μεγάλη πράγμα που σημαίνει ότι σε κάθε επανάληψη, τα αποτελέσματά μας παρουσιάζανε μεγάλες διαφορές.

Πρόβλημα 7: $\min f(x)$ με $f(x) = (x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 + 10x_5^6 + 7x_6^2 + x_7^4 - 4x_6x_7 - 10x_6 - 8x_7$



Διάγραμμα 4.13



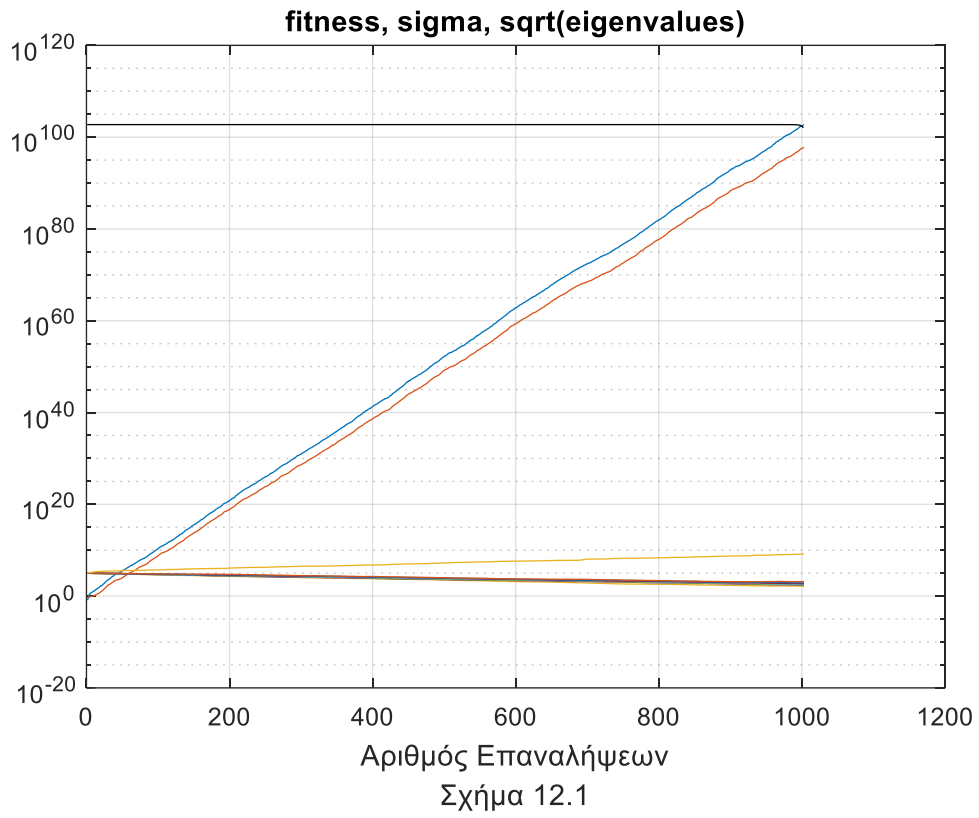
Διάγραμμα 4.14

Μέση Τιμή	inf
Ελάχιστο	-16.097
Μέγιστο	-16.097
Τυπική Απόκλιση	8.558×10^{-14}

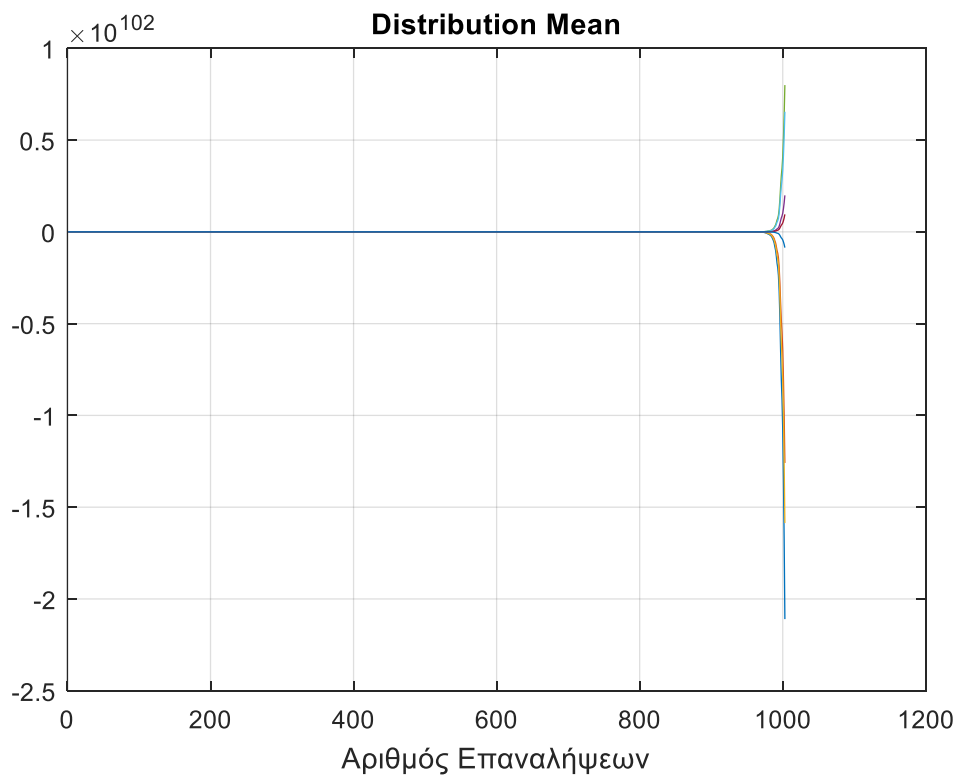
Πίνακας 4.7: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

Σχολιασμός: Η μέση τιμή της κατανομής κάθε σωματιδίου παρατηρούμε ότι παραμένει σταθερή μετά από ελάχιστες επαναλήψεις της CMA. Παρόλα αυτά, η fitness function βλέπουμε πως συνεχίζει να ελαττώνεται καθ'όλη την διάρκεια του αλγορίθμου ενώ επίσης παρατηρούμε ότι το ελάχιστο και μέγιστο στοιχείο για την fitness function από όλες τις επαναλήψεις μας είναι το ίδιο άρα το ελάχιστο είναι το -16.097.

Πρόβλημα 8: $\min f(x)$ με $f(x) = x_1 + x_2 + x_3$



Διάγραμμα 15



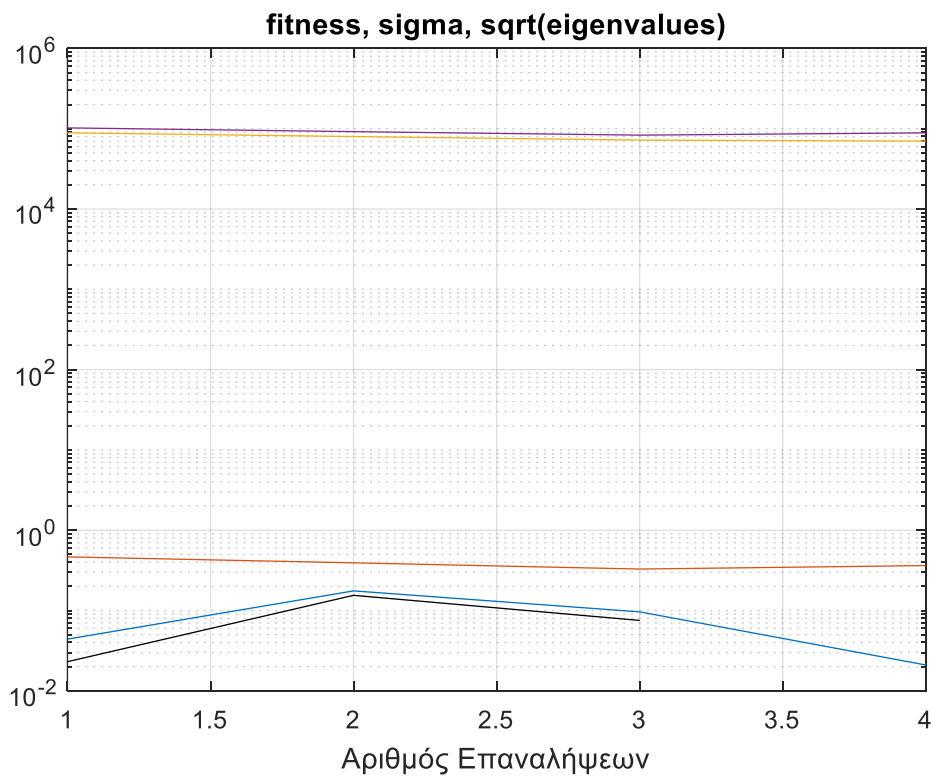
Διάγραμμα 4.16

Μέση Τιμή	-2.755×10^{108}
Ελάχιστο	-1.653×10^{109}
Μέγιστο	-3.166×10^{101}
Τυπική Απόκλιση	6.749×10^{108}

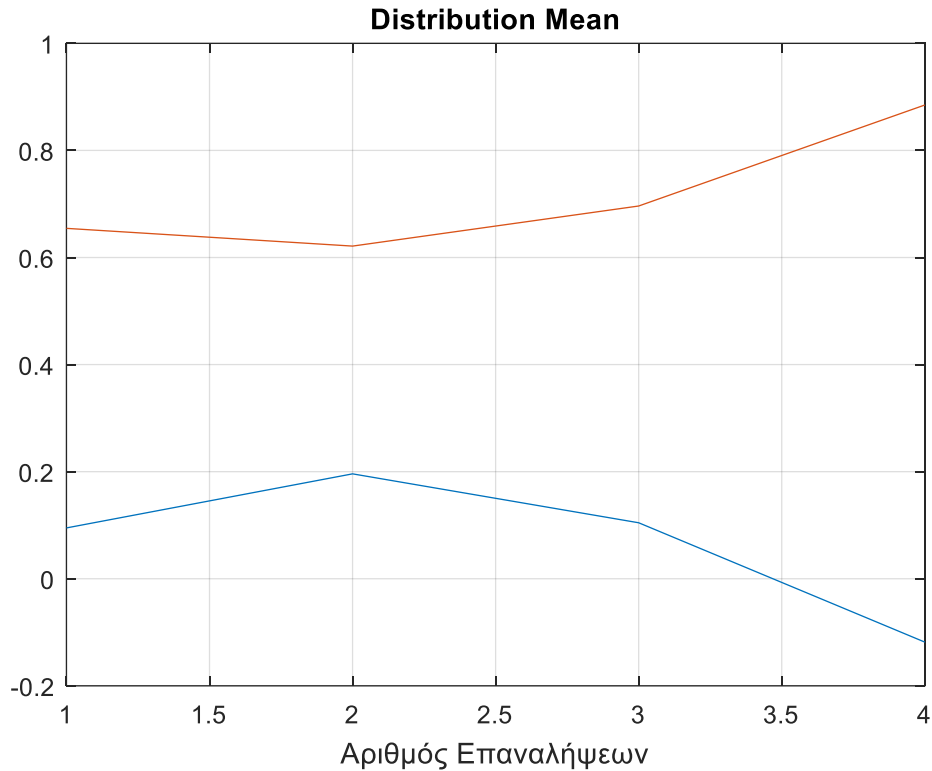
Πίνακας 4.8: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

Σχολιασμός: Το πρόβλημα μας έχει παρόμοια αποτελέσματα με αυτά του προβλήματος 2. Οι διαφορές είναι στην επανάληψη που τελειώνει η διαδικασία, που στην περίπτωσή μας είναι η 1000^η ενώ σημαντικό είναι το γεγονός ότι η τυπική απόκλιση δεν είναι στο άπειρο αφού αυτό σημαίνει ότι τα αποτελέσματά μας είναι καλύτερα.

Πρόβλημα 9: $\min f(x)$ με $f(x) = x_1^2 + (x_2 - 1)^2$



Διάγραμμα 17



Διάγραμμα 18

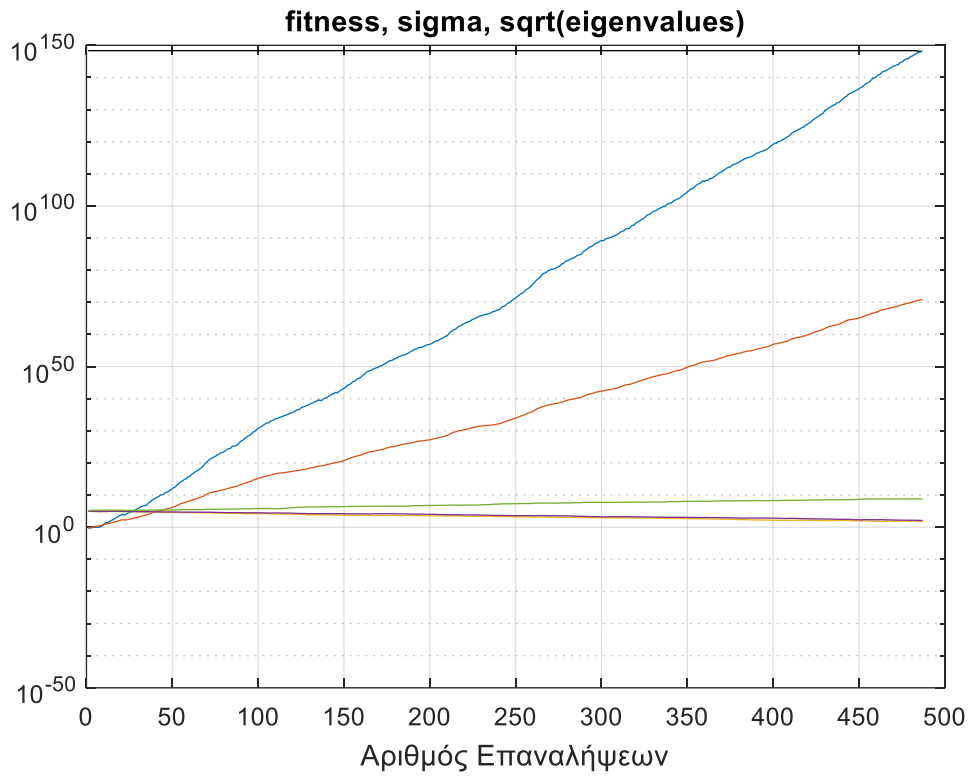
Μέση Τιμή	0.029
Ελάχιστο	0.011
Μέγιστο	0.053
Τυπική Απόκλιση	0.016

Πίνακας 9: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

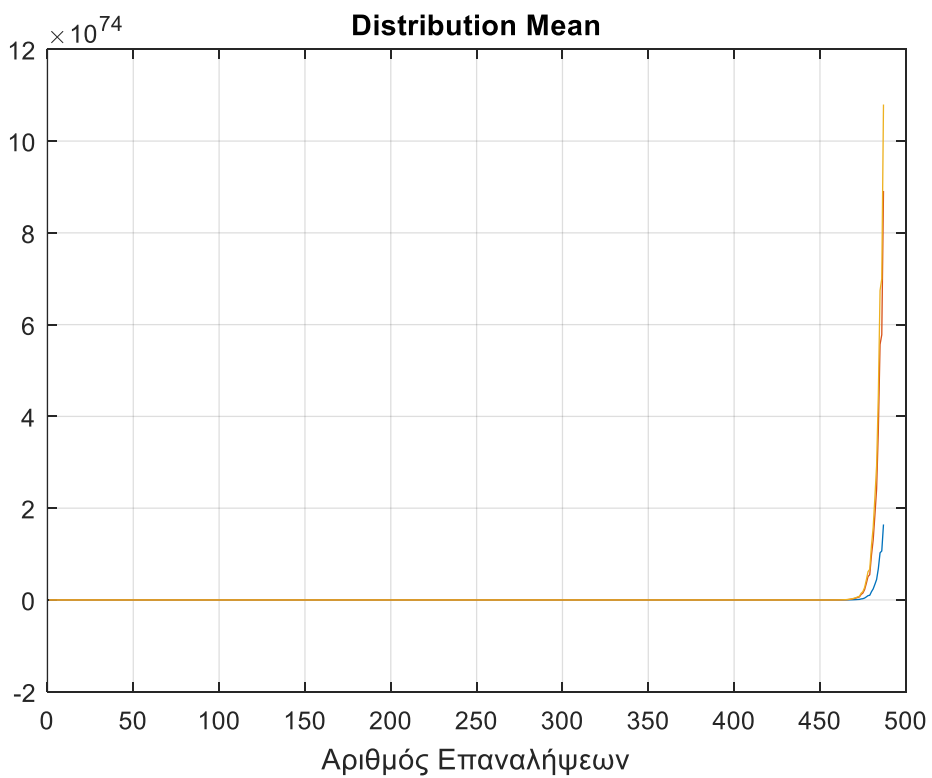
Σχολιασμός: Σε αυτή την περίπτωση θα μπορούσαμε να πούμε ότι η μέση τιμή της κατανομής των 2 σωματιδίων ακολουθούν συμμετρική (ως προς την ευθεία $y=0.2$ στο διάγραμμα) διαδρομή και φαίνεται ότι μετά την 3^η επανάληψη ακολουθούν ίδια πορεία.

Βέβαια, από τον πίνακα μας φαίνεται ότι η τυπική απόκλιση δεν είναι μικρή σχετικά με τις τιμές του προβλήματός μας οπότε δεν μπορούμε να αποφανθούμε σίγουρα για κάποιο αποτέλεσμα.

Πρόβλημα 10: $\min f(x)$ με $f(x) = (100 - (x_1 - 5)^2 - (x_2 - 5)^2 - (x_3 - 5)^2) / 100$



Διάγραμμα 4.19



Διάγραμμα 4.20

Μέση Τιμή	-3.781×10^{147}
-----------	--------------------------

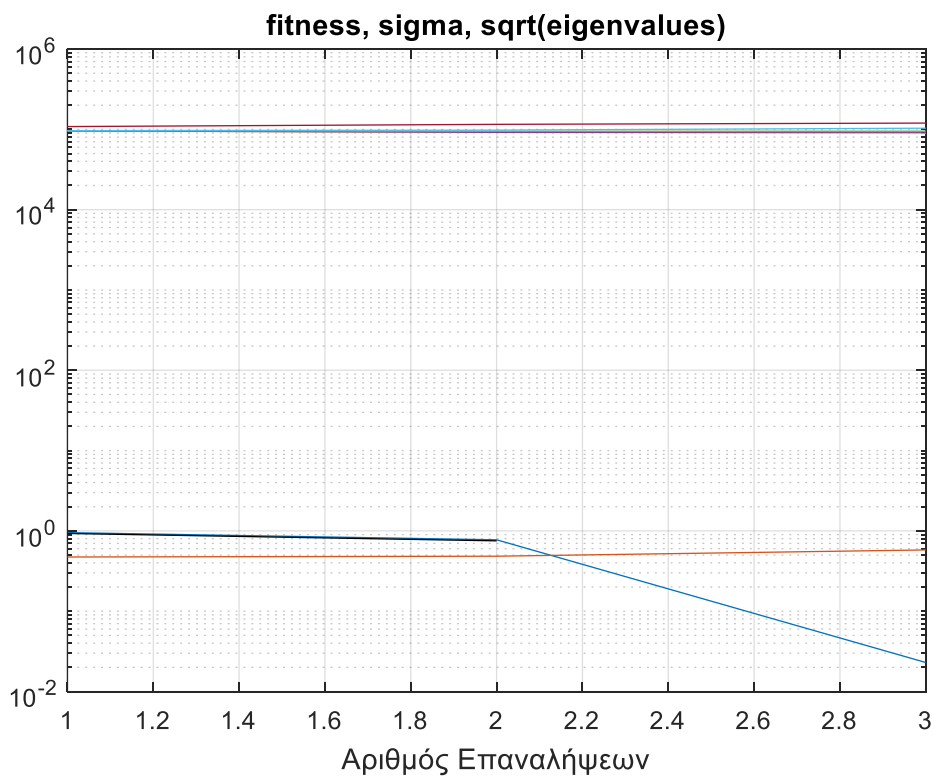
Ελάχιστο	-2.268×10^{148}
Μέγιστο	-1
Τυπική Απόκλιση	9.261×10^{147}

Πίνακας 4.10: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

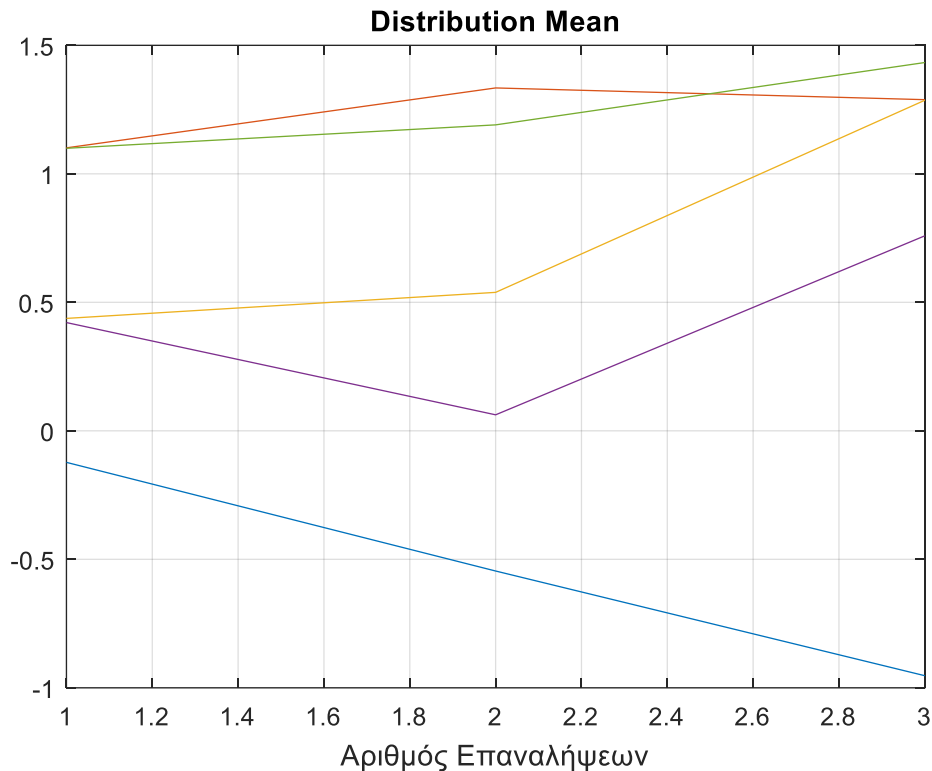
Σχολιασμός: Αρχικά, παρατηρούμε ότι η μέση τιμή της κατανομής για κάθε σωματίδιο ακολουθεί σταθερή πορεία μέχρι περίπου την 470^η επανάληψη, ενώ μετά ακολουθεί εκθετική. Όσον αφορά την fitness function είναι διαρκώς αύξουσα αλλά δεν απειρίζεται και γενικά έχουμε, με την βοήθεια του πίνακα μας, ισχυρές ενδείξεις περί εύρεσης ελαχίστου.

Βέβαια με βάση τα δεδομένα μας, δεν μπορούμε να αποφανθούμε υπέρ μια λύσης λόγω της τεράστιας τιμής της τυπικής απόκλισης μεταξύ των δεδομένων μας.

Πρόβλημα 11: $\min f(x)$ με $f(x) = e^{x_1 x_2 x_3 x_4}$



Διάγραμμα 4.21



Διάγραμμα 4.22

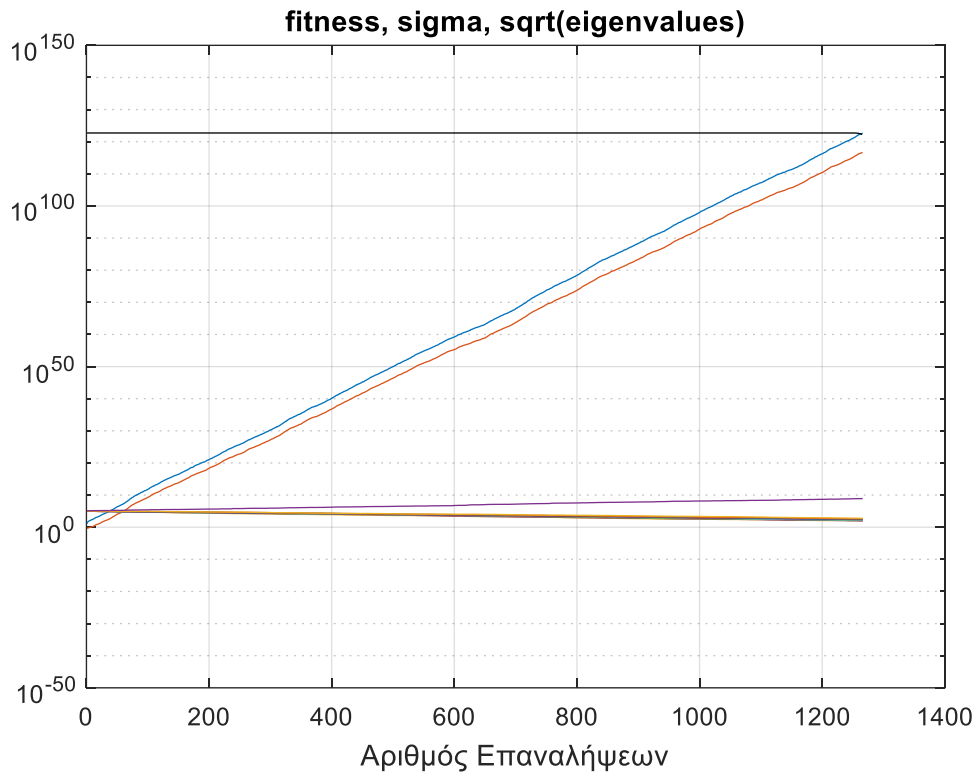
Μέση Τιμή	0.078
Ελάχιστο	0.011
Μέγιστο	0.319
Τυπική Απόκλιση	0.119

Πίνακας 4.11: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

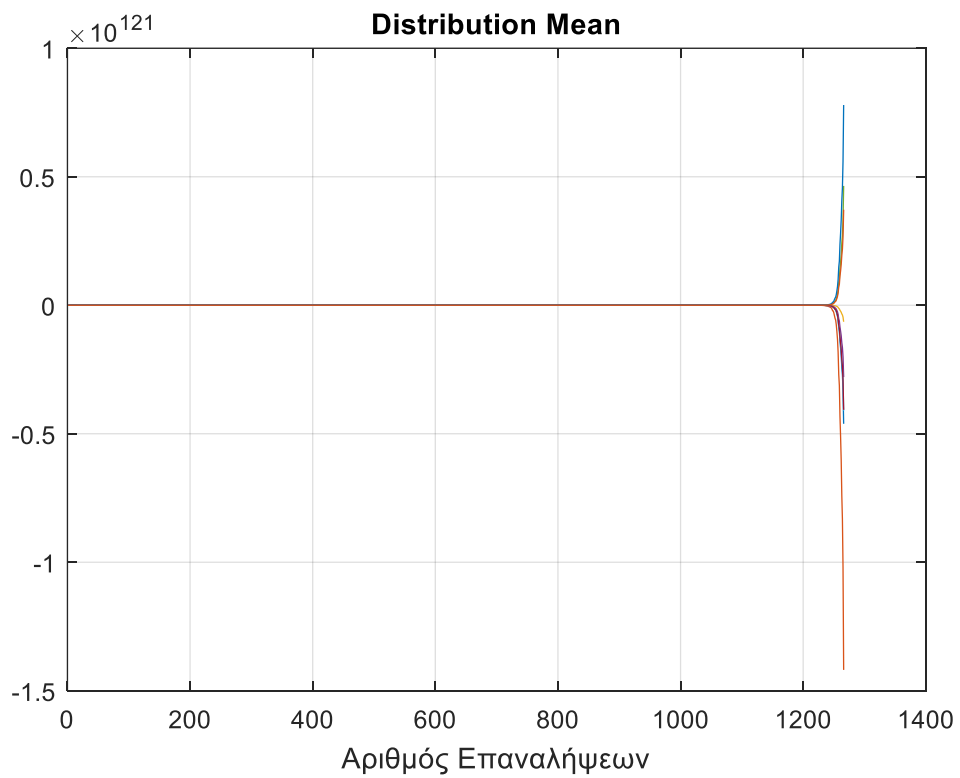
Σχολιασμός: Κοιτώντας το διάγραμμα θα μπορούσαμε να πούμε πως η διαδικασία της σύγκλισης είναι σχεδόν άμεση αφού η fitness function είναι σταθερή καθ'όλη την διάρκεια εκτέλεσης της CMA που διαρκεί μάλιστα μόλις 3 επαναλήψεις.

Βέβαια δεν θα μπορούσαμε με βεβαιότητα να αποφανθούμε για το κατά πόσο θα μπορούσαμε να εντοπίσουμε την ελάχιστη τιμή της αφού η τυπική απόκλιση είναι σχετικά μεγάλη οπότε έχουμε διαφορές στα αποτελέσματά μας.

Πρόβλημα 12: $\min f(x)$ με $f(x) = -9x_5 - 15x_8 + 6x_1 + 16x_2 + 10x_6 + 10x_7$



Διάγραμμα 4.23



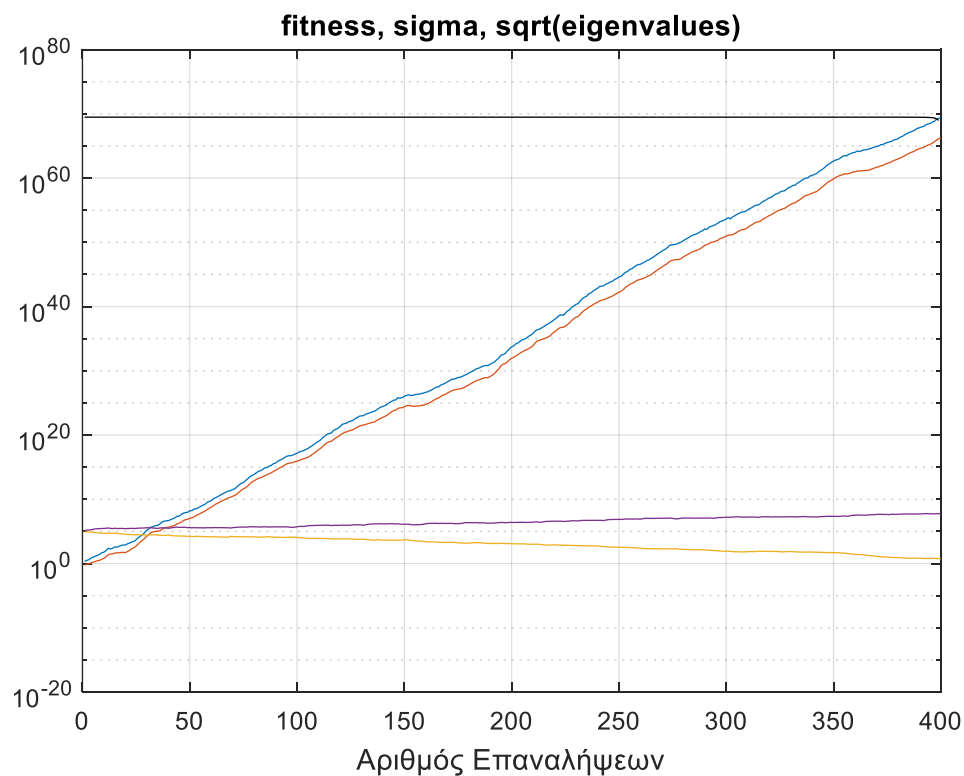
Διάγραμμα 4.24

Μέση Τιμή	-2.664×10^{123}
Ελάχιστο	-1.545×10^{124}
Μέγιστο	-2.845×10^{104}
Τυπική Απόκλιση	-6.269×10^{123}

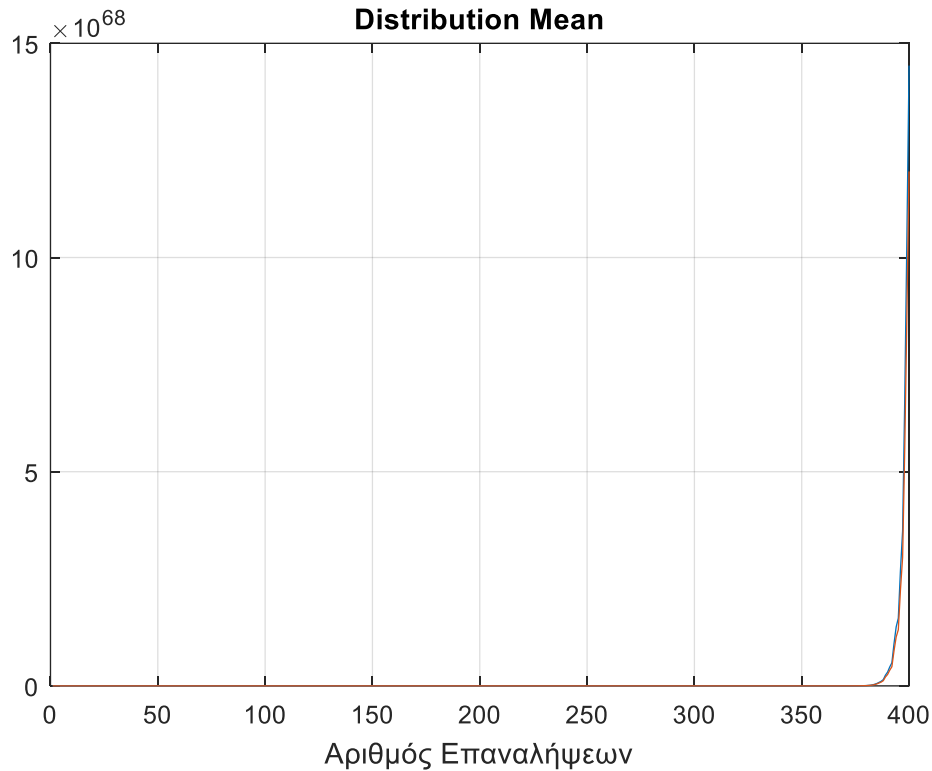
Πίνακας 4.12: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

Σχολιασμός: Στο συγκεκριμένο πρόβλημα θα μπορούσαμε να πούμε ότι η πορεία της fitness function είναι ίδια με του προβλήματος 2 μόνο που η διαδικασία άρχισε να εξελίσσεται κάπου στην 1300^η επανάληψη ενώ πάλι δεν μπορούμαστε να ισχυριστούμε με βεβαιότητα ότι βρήκαμε το σημείο ελαχιστοποίησης αφού η τυπική απόκλιση είναι μεγάλη.

Πρόβλημα 13: $\min f(x)$ με $f(x) = -x_1 - x_2$



Διάγραμμα 25



Διάγραμμα 4.26

Μέση Τιμή	-1.600×10^{77}
Ελάχιστο	-9.470×10^{77}
Μέγιστο	-2.219×10^{57}
Τυπική Απόκλιση	3.856×10^{77}

Πίνακας 4.13: Στατιστικά της fitness function έπειτα από 5 φορές εκτέλεσης του αλγορίθμου

Σχολιασμός: Από το διάγραμμα γίνεται φανερό ότι μετά από περίπου από την 380^η επανάληψη για κάθε σωματίδιο η μέση τιμή της κατανομής του ενώ ακολουθούσε σταθερή μηδενική πορεία, μεγαλώνει εκθετικά. Ταυτόχρονα, παρατηρούμε ότι σε όλη την πορεία του διαγράμματος η fitness function μειώνεται και έτσι μπορούμε να υποθέσουμε ότι η διαδικασία συνεχίζεται και μετά την 400^η επανάληψη.

Από τον πίνακα δεν μπορούμε να είμαστε βέβαιοι ότι βρήκαμε το ελάχιστο αφού παρατηρούμε ότι η τυπική απόκλιση του προβλήματος είναι μεγάλη μεταξύ των διάφορων επαναλήψεων.

4.2 Σύνοψη Αποτελεσμάτων με την μορφή πίνακα:

Συνοψίζοντας, θα παραθέσουμε τα αποτελέσματα μας με την μορφή πίνακα όπου θα περιλαμβάνονται όλα τα ευρήματά μας ούτως ώστε να μπορεί να γίνει μια πιο λεπτομερής μελέτη:

Πρόβλημα	Μέση Τιμή	Ελάχιστο	Μέγιστο	Τυπική Απόκλιση
1. $\min f(x)$ με $f(x) = \sqrt{n} \prod_{i=1}^n x_i$ ($n = 10$)	inf	-1.670×10^{308}	-7.311×10^{307}	inf
2. $\min f(x)$ με $f(x) =$ $5.3578547x_3^2 + 0.8356891x_1x_5 +$ $37.293239x_1 - 40792.141$	$-5,355 \times 10^{157}$	$-3,194 \times 10^{158}$	$-3,220 \times 10^{123}$	inf
3. $\min f(x)$ με $f(x) = 3x_1 +$ $0,000001x_1^3 + (0,000002/3)x_2^3$	inf	-1.814×10^{219}	-1.978×10^{189}	inf
4. $\min f(x)$ με $f(x) = (x_1 - 10)^3 +$ $(x_2 - 20)^3$	inf	-1.670×10^{308}	-7.311×10^{208}	inf
5. $\min f(x)$ με $f(x) = x_1^2 + x_2^2 +$ $x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 +$ $4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 -$ $1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 -$ $10)^2 + (x_{10} - 7)^2 + 45$	-31	-31	-31	0
6. $\max f(x)$ με $f(x) = \frac{\sin^3(2\pi x_1)\sin(2\pi x_2)}{x_1^3(x_1+x_2)}$	-2.179×10^{18}	-9.024×10^{18}	-0.0175	3.720×10^{18}
7. $\min f(x)$ με $f(x) = (x_1 - 10)^2 +$ $5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 +$ $10x_5^6 + 7x_6^2 + x_7^4 - 4x_6x_7 - 10x_6 -$ $8x_7$	inf	-16.097	-16.097	8.558×10^{-14}
8. $\min f(x)$ με $f(x) = x_1 + x_2 + x_3$	-2.755×10^{108}	-1.653×10^{109}	-3.166×10^{101}	6.749×10^{108}
9. $\min f(x)$ με $f(x) = x_1^2 + (x_2 - 1)^2$	0,029	0,011	0,053	0,016
10. $\min f(x)$ με $f(x) = (100 -$ $(x_1 - 5)^2 - (x_2 - 5)^2 - (x_3 - 5)^2)/$ 100	$-3,781 \times 10^{147}$	$-2,268 \times 10^{148}$	-1	$9,261 \times 10^{147}$
11. $\min f(x)$ με $f(x) = e^{x_1x_2x_3x_4}$	0,078	0,011	0,319	0,119
12. $\min f(x)$ με $f(x) = -9x_5 -$ $15_8 + 6x_1 + 16x_2 + 10x_6 + 10x_7$	$-2,664 \times 10^{123}$	$-1,545 \times 10^{124}$	$-2,845 \times 10^{104}$	$-6,269 \times 10^{123}$
13. $\min f(x)$ με $f(x) = -x_1 - x_2$	$-1,600 \times 10^{77}$	$-9,470 \times 10^{77}$	$-2,219 \times 10^{57}$	$3,856 \times 10^{77}$

Πίνακας 4.14: Σύνοψη των αποτελεσμάτων

Συμπεράσματα

Μελετώντας την CMA αλλά και την φύση των γενετικών αλγορίθμων είναι εύκολα διακριτό το πόσο ισχυρό εργαλείο αποτελεί η CMA στον τομέα της βελτιστοποίησης και ως εκ τούτου και της επιχειρησιακής έρευνας. Είναι μία ελκυστική επιλογή για προβλήματα τα οποία παρουσιάζουν προβλήματα όταν εφαρμόζονται σε αυτές οι κλασικές μέθοδοι βελτιστοποίησης (π.χ. μέθοδος συζυγών κλίσεων) λόγω του ότι μπορεί να τα επιλύσει ακόμα και αν οι συναρτήσεις παρουσιάζουν ασυνέχεια, μη γραμμικότητα κτλ ενώ, ακόμα και αν δεν μελετήθηκε, η CMA διαχειρίζεται πολύ καλά τους περιορισμούς που υπάρχουν σε πολλά προβλήματα.

Επίσης, η CMA παρόλο που ανήκει στην οικογένεια των εξελικτικών αλγορίθμων μπορεί και ξεπερνάει ορισμένα προβλήματα που παρουσιάζουν ορισμένοι εξελικτικοί αλγόριθμοι όπως είναι:

1. Η κακή απόδοση που έχουν σε κακώς κλιμακωτές και / ή εξαιρετικά μη διαχωρίσιμες αντικειμενικές συναρτήσεις. Συγκεκριμένα, αυτό φαίνεται από την εξίσωση (30) όπου προσαρμόζεται η κατανομή αναζήτησης σε προβλήματα που παρουσιάζουν ανεπαρκώς κλιμακούμενα και μη διαχωρίσιμα αποτελέσματα.
2. Η τάση χρήσης μεγάλων αριθμών πληθυσμού. Ένας από τους πιο συχνούς και ταυτόχρονα πιο δύσκολα ανιχνεύσιμους λόγους για τον οποίο πολλές αποτυγχάνουν οι αλγόριθμοι είναι εκφυλισμός του πληθυσμού σε έναν υπόχωρο. Αυτό συνήθως εμποδίζεται είτε από τα μη προσαρμοστικά συστατικά στον αλγόριθμο είτε από ένα μεγάλο μέγεθος πληθυσμού (σημαντικά μεγαλύτερο από την διάσταση του προβλήματος) είτε και από τα δύο. Στη CMA, το μέγεθος του πληθυσμού μπορεί να επιλεγεί ελεύθερα, επειδή το οι ρυθμοί εκμάθησης c_1 και c_μ στην (30) αποτρέπουν τον εκφυλισμό ακόμη και για τα μικρά μεγέθη του πληθυσμού. Τα μικρά μεγέθη του πληθυσμού συνήθως οδηγούν σε ταχύτερη σύγκλιση, σε μεγάλο πληθυσμό ενώ τα μεγάλα μεγέθη βοηθούν στην αποφυγή τοπικών optima.

3. Η πρόωρη σύγκλιση του πληθυσμού. Ο έλεγχος μεγέθους βήματος στην (37) εμποδίζει τον πληθυσμό να συγκλίνουν πρόωρα. Δεν εμποδίζει την αναζήτηση να καταλήξει σε τοπικό βέλτιστο.

Κατά την διάρκεια, λοιπόν, που τρέξαμε τον αλγόριθμό μας παρατηρήσαμε ότι ανάλογα την φύση και την πολυπλοκότητα του προβλήματος, εκτελούντουσαν περισσότερες ή λιγότερες επαναλήψεις γεγονός που έκανε το πρόγραμμα του Matlab να δυσκολεύεται περισσότερο. Γενικά, το πρόγραμμα θα μπορούσε να χαρακτηριστεί «βαρύ» για τον υπολογιστή αφού εκτελούνται πάρα πολλές πράξεις κατά την διάρκεια λειτουργίας του. Βέβαια, τις περισσότερες φορές θα λέγαμε ότι έφτανε σε αποτέλεσμα, ενώ κιόλας πρέπει να λάβουμε υπ' όψιν μας ότι το Matlab δεν επιτρέπει τιμές από ένα όριο και κάτω να εμφανιστούν και τις εμφανίζει ως άπειρο (βλ. Πρόβλημα 1). Για αυτό πολύ πιθανό είναι ότι θα μπορούσαμε να βρούμε και άλλες λύσεις (ίσως και όλες) άμα δεν υπήρχε αυτό το πρόβλημα. Παρόλα αυτά, οι τυπικές αποκλίσεις μεταξύ των διάφορων εκτελέσεων της CMA για το ίδιο πρόβλημα παρουσίαζαν μεγάλη τυπική απόκλιση γύρω από την μέση τιμή οπότε ίσως θα έπρεπε για κάποιες να γίνει εκτέλεση περισσότερες φορές ούτως ώστε να ελαχιστοποιηθεί. Άλλωστε, ο αλγόριθμος είναι στοχαστικός οπότε όσες περισσότερες φορές τον εκτελέσουμε, τόσο πιο ασφάλεια θα μας παρέχει όσον αφορά τα αποτελέσματά μας.

Εν κατακλείδι, θα λέγαμε ότι η CMA θα είναι προτιμότερο να χρησιμοποιείται μετά από ανεπιτυχείς προσπάθειες εκτέλεσης πιο απλών αλγορίθμων βελτιστοποίησης για τον λόγο ότι η χρήση της δυσκολεύει λίγο τον υπολογιστή από το πλήθος των πράξεων που έχει. Επίσης, θέλει μεγάλη προσοχή στον ορισμό των συναρτήσεων και την σύνδεση τους με τον αλγόριθμο καθώς επίσης και το ότι ίσως θα χρειαστεί μια παραπάνω γνώση μαθηματικών και προγραμματισμού προκειμένου να μπορέσουμε να καταφέρουμε να επιλύσουμε προβλήματα στα οποία υπάρχουν περιορισμοί. Παρόλα αυτά, η ενασχόληση είναι πολύ ουσιαστική καθώς φαίνεται ότι μέσω αυτής της πολύ σύγχρονης διαδικασίας είναι πολύ πιθανό να δοθεί επίλυση σε κάθε πρόβλημα που μπορεί να προκύψει.

Κατάλογος βιβλιογραφικών αναφορών

- [1] Abdi, H. (2010). Principal Component Analysis
- [2] Arnold. (n.d.). Weighted multirecombination evolution strategies. Theoretical computer. 2006.
- [3] Beyer, H.-G. (2007). Evolution strategies.
- [4] Cassioli, A. (2013). A Tutorial on Black-Box Optimization. Paris.
- [5] Fogel, D. (1991). System Identification through Simulated Evolution: A Machine Learning Approach to Modeling. Ginn Press, Needham Heights, MA
- [6] Goldberg, D., Holland, J. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Massachusetts.
- [7] Hansen, A. A. (2013). Tutorial CMA-ES — Evolution Strategies and Covariance Matrix Adaptation. Amsterdam.
- [8] Hansen, N. (2006). Towards a new evolutionary computation. Advances. Springer.
- [9] Hansen, N. (2016). Nikolaus Hansen. Paris.
- [10] Holland, J. (1975). Adaptation in Natural and Artificial Systems
- [11] http://www.scholarpedia.org/article/Evolution_strategies. (n.d.).
- [12] Jones T. (1995), Evolutionary Algorithms, Fitness Landscapes and Search, New Mexico
- [13] Koza, J. (1992). Genetic Programming: On the Programming of Computers by Means of Natural Selection
- [14] Rechenberg, I. (1994) Evolutionsstrategie '94, volume 1 of Werkstatt Bionik und Evolutionstechnik. Frommann-Holzboog Verlag, Stuttgart, Germany.
- [15] Salimans, T., J. H. (n.d.). Evolution Strategies as a Scalable Alternative to Reinforcement Learning.
- [16] T. R. CONSTRAINED EVOLUTIONARY OPTIMIZATION-the penalty function approach.

- [17] Αννίνου, Ν. (2009). Γενετικοί και Μέτα-Γενετικοί Αλγόριθμοι και η Εφαρμογή τους στην Εκτίμηση ARMA Μοντέλων. Πάτρα.
- [18] Γκούντη, Α. (2013). ΕΞΕΛΙΚΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ: Ο ΑΛΓΟΡΙΘΜΟΣ ΤΗΣ ΔΙΑΦΟΡΙΚΗΣ ΕΞΕΛΙΞΗΣ. Θεσσαλονίκη.
- [19] Καλκούνου, Δ. (2013). Πολυμεταβλητή Στατιστική Ανάλυση, Πάτρα
- [20] Κολέτσος, Ι., Δ. Σ. (2012). Εισαγωγή στην Επιχειρησιακή Έρευνα. Αθήνα.
- [21] Λυκοθανάσης, Σ. (2001). Γενετικοί αλγόριθμοι και εφαρμογές. Πάτρα.