

A dark blue vertical bar is positioned on the left side of the slide. A blue arrow points to the right from this bar, containing the date '19/6/2018'. In the bottom-left corner, there are several thin, curved lines in dark blue and light grey, creating an abstract graphic element.

19/6/2018

# Κατασκευή Γενετικού Αλγορίθμου για προσέγγιση συνάρτησης

ΓΙΩΡΓΟΣ ΜΠΑΛΑΟΥΡΑΣ (8861)

# Τεχνικές Βελτιστοποίησης

Ονοματεπώνυμο: Μπαλαούρας Γεώργιος

AEM: 8861

E-mail: [mpalaourg@ece.auth.gr](mailto:mpalaourg@ece.auth.gr)

Εξάμηνο: 6<sup>ο</sup> Ηλεκτρονικής

## Project : Κατασκευή γενετικού αλγορίθμου για προσέγγιση συνάρτησης

Αντικείμενο αυτής της άσκησης, είναι η σχεδίαση ενός γενετικού αλγορίθμου που ως σκοπό θα έχει την προσέγγιση μιας συνάρτησης  $f(u_1, u_2)$ , της οποίας ο τύπος είναι άγνωστος αλλά γνωρίζουμε ότι είναι συνεχής. Η προσέγγιση θα βασιστεί στον γραμμικό συνδυασμό γκαουσιανών συναρτήσεων, της μορφής  $G(u_1, u_2) = Ae^{-\left[\frac{(u_1-c_1)^2}{2\sigma_1^2} + \frac{(u_2-c_2)^2}{2\sigma_2^2}\right]}$ .

Τι είναι οι γενετικοί αλγόριθμοι?

Αρχικά, οι γενετικοί αλγόριθμοι χρησιμοποιούνται ευρέως σε περιπτώσεις εύρεσης ελαχίστου, όπου η αναλυτική έκφραση της  $f$  είναι άγνωστη, άρα η χρήση των κλασικών τεχνικών αναζήτησης είναι απαγορευτική. Ωστόσο, μπορούν να χρησιμοποιηθούν κι σε άλλα είδους προβλήματα, όπως -στην περίπτωση μας- την εύρεση της αναλυτικής έκφρασης της  $f$ , αρκεί να υπάρχει μαθηματική περιγραφή του προβλήματος. Στην ουσία πρόκειται, για πιθανοκρατικούς αλγορίθμους που δεν εγγυώνται την εύρεση της βέλτιστης λύσης, αλλά είναι ικανοί να επιστρέψουν μια αρκετά καλή λύση σε εύλογο χρονικό διάστημα, αναζητώντας την σ' έναν χώρο υποψήφιων λύσεων σύμφωνα με κάποιο κριτήριο.

Η λειτουργία των γενετικών αλγορίθμων, βασίζεται στην βιολογία και στην θεωρία της εξέλιξης. Πιο συγκεκριμένα, υπάρχει ο πληθυσμός, που αποτελείται από το σύνολο των λύσεων. Κάθε μια τέτοια λύση ονομάζεται χρωμόσωμα. Τα καλύτερα από αυτά τα χρωμοσώματα, θα

«επιβιώσουν» και θα υπάρχουν κι στην επόμενη γενιά, όπως συμβαίνει κι στην φύση. Το κριτήριο για την εύρεση των καλύτερων χρωμοσωμάτων, ονομάζεται *συνάρτηση ικανότητας*, *fitness function*. Βέβαια, στην επόμενη γενιά θα οδηγηθούν και μερικά τυχαία χρωμοσώματα, έτσι ώστε να αποφύγουμε φαινόμενα γρήγορης σύγκλισης σε λύση, που δεν είναι η βέλτιστη δυνατή. Βασικός κανόνας των γενετικών αλγορίθμων είναι, ότι ο αριθμός του πληθυσμού θα παραμείνει ο ίδιος σ' όλες τις γενιές. Έτσι, έχοντας τις καλύτερες υποψήφιες λύσεις για την επόμενη γενιά, επιλέγονται ορισμένες τυχαία και χρησιμοποιούνται ως γονείς για την παραγωγή νέων λύσεων, μέσω της *διασταύρωσης* (*crossover*), αυτό συμβαίνει καθώς επιθυμούμε να υπάρχουν περισσότερες λύσεις με τα χαρακτηριστικά των βέλτιστων. Σποραδικά συμβαίνουν τυχαίες μεταλλάξεις (*mutate*), σε κάποιο από τα επιμέρους χαρακτηριστικά (*γονίδια* – *gene*) που κωδικοποιούν το χρωμόσωμα. Τέλος, το πρόβλημα του αρχικού πληθυσμού, λύνεται δημιουργώντας τυχαία τα απαραίτητα χρωμοσώματα.

Όλες οι παραπάνω διαδικασίες, *crossover*, *mutate*, *κωδικοποίηση γονιδίων* του χρωμοσώματος και επιλογή της *fitness function* μπορούν να γίνουν με διάφορες τεχνικές και διαφέρουν από πρόβλημα σε πρόβλημα.

## Διαδικασίες και παράμετροι του γενετικού αλγορίθμου

### ➤ Κωδικοποίηση Γονιδίων

Η φύση του προβλήματος, δεν επιτρέπει την αναπαράσταση των γονιδίων από bits 0, 1. Σκεπτόμενος ότι τα γονίδια θα αποτελούνται από τα πλάτη, τα κέντρα κι τις τυπικές αποκλίσεις των γκαουσιανών, καταλληλότερη είναι η αναπαράσταση με την χρήση πραγματικών αριθμών. Ωστόσο, ο κάθε «τύπος» γονιδίου έχει τα δικά του όρια τιμών και συγκεκριμένα:

- Πλάτος:  $A \in (f_{min}, f_{max})$ , όπου  $f_{max}$  και  $f_{min}$ , η μέγιστη και η ελάχιστη τιμή της συνάρτησης αντίστοιχα.
- Κέντρο ( $c_1$ ), το εύρος τιμών του εξαρτάται από το πεδίο ορισμού της εκάστοτε συνάρτησης. Συγκεκριμένα για συνάρτηση με πεδίο ορισμού  $(a, b)$  το  $c_1 \in (a - 2, b + 2)$ .
- Τυπική απόκλιση ( $\sigma_1$ ),  $\sigma_1 \in (0.2, 1.3)$ .

➤ Χρωμοσώματα

Εκτός από την κωδικοποίηση των γονιδίων του κάθε χρωμοσώματος, υπάρχουν κι άλλες παράμετροι που πρέπει να επιλεγούν. Αρχικά, ο αριθμός του πληθυσμού δεν πρέπει να είναι ούτε πολύ μικρός, για να υπάρχει μεγαλύτερο πλήθος λύσεων, αλλά ούτε και πολύ μεγάλος καθώς δημιουργούνται προβλήματα αποθήκευσης και διαχείρισης της μνήμης. Έπειτα, ο αριθμός των καλύτερων χρωμοσωμάτων και ο αριθμός των τυχαίων που θα επιλεγθούν από το υπόλοιπο του πληθυσμού για να «μεταφερθούν» στην επόμενη γενιά. Τέλος, το ποσοστό χρωμοσωμάτων από διασταυρώσεις (*crossover*), επιλέγεται έτσι ώστε ο πληθυσμός να παραμένει σταθερός.

➤ *Fitness Function*

Η προσέγγιση μιας συνάρτησης, αμέσως οδηγεί στην χρήση του σφάλματος, δηλαδή την απόκλιση από την πραγματική τιμή, ως συνάρτηση ικανότητας. Κινούμενος σ' αυτό το μοτίβο, αρχικά χρησιμοποιήθηκε το απόλυτο σφάλμα (*absolute error*),  $e_{abs} = \sum |f(x_i) - \bar{f}(x_i)|$ , με  $f(x_i)$  συμβολίζεται η πραγματική τιμή και με  $\bar{f}(x_i)$  η προσέγγιση της. Ωστόσο, οι τιμές του σφάλματος ήταν υψηλές με αποτέλεσμα την αργή σύγκλιση του αλγορίθμου. Έτσι, στην συνέχεια χρησιμοποιήθηκε το μέσο τετραγωνικό σφάλμα (*mean squared error*),  $e_{mse} = \frac{1}{n} \sum (f(x_i) - \bar{f}(x_i))^2$ , με  $f(x_i)$  και  $\bar{f}(x_i)$  όμοια με πριν και  $n$  ο αριθμός των σημείων που χρησιμοποιήθηκαν για την προσέγγιση.

➤ *Crossover*

Για την παραγωγή των νέων χρωμοσωμάτων, υπάρχουν διάφορες τεχνικές διασταύρωσης που χρησιμοποιούνται, στις εκάστοτε καλύτερες λύσεις κάθε γενιάς. Σαν πρώτη σκέψη, δοκιμάστηκε η εναλλαγή γονιδίων, μεταξύ δύο χρωμοσωμάτων. Δηλαδή, τα παρακάτω χρωμοσώματα

1<sup>ο</sup> χρωμόσωμα

$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_7$	$A_8$	$A_9$
-------	-------	-------	-------	-------	-------	-------	-------	-------

2<sup>ο</sup> χρωμόσωμα

$B_1$	$B_2$	$B_3$	$B_4$	$B_5$	$B_6$	$B_7$	$B_8$	$B_9$
-------	-------	-------	-------	-------	-------	-------	-------	-------

θα δημιουργούσαν τα εξής χρωμοσώματα:

1<sup>ο</sup> παιδί

$B_1$	$B_2$	$B_3$	$A_4$	$A_5$	$A_6$	$B_7$	$B_8$	$B_9$
-------	-------	-------	-------	-------	-------	-------	-------	-------

2<sup>ο</sup> παιδί

$A_1$	$A_2$	$A_3$	$B_4$	$B_5$	$B_6$	$A_7$	$A_8$	$A_9$
-------	-------	-------	-------	-------	-------	-------	-------	-------

Ωστόσο, τα αποτελέσματα στο τέλος της προσέγγισης παρουσίαζαν μεγάλη απόκλιση μεταξύ τους και δεν υπήρχε συνοχή. Έτσι, σαν μέθοδος διασταύρωσης επιλέχθηκε το νέο παιδί να αποτελείται από το ημι-άθροισμα των όρων, των δύο γονέων του. Δηλαδή,

$\frac{A_1 + B_1}{2}$	$\frac{A_2 + B_2}{2}$	$\frac{A_3 + B_3}{2}$	$\frac{A_4 + B_4}{2}$	$\frac{A_5 + B_5}{2}$	$\frac{A_6 + B_6}{2}$	$\frac{A_7 + B_7}{2}$	$\frac{A_8 + B_8}{2}$	$\frac{A_9 + B_9}{2}$
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

#### ➤ *Mutate*

Στην μετάλλαξη, εμπλέκονται όλα τα χρωμοσώματα και είναι τυχαίο φαινόμενο. Έτσι, αφενός πρέπει να ορισθεί η πιθανότητα για την οποία θα υφίσταται και αφετέρου να βρεθεί η μέθοδος της μετάλλαξης. Σαν μέθοδος μετάλλαξης, χρησιμοποιείται η εναλλαγή ενός τυχαίου γονιδίου, με μια τυχαία τιμή, από το εύρος των τιμών του.

## Προσέγγιση συνάρτησης

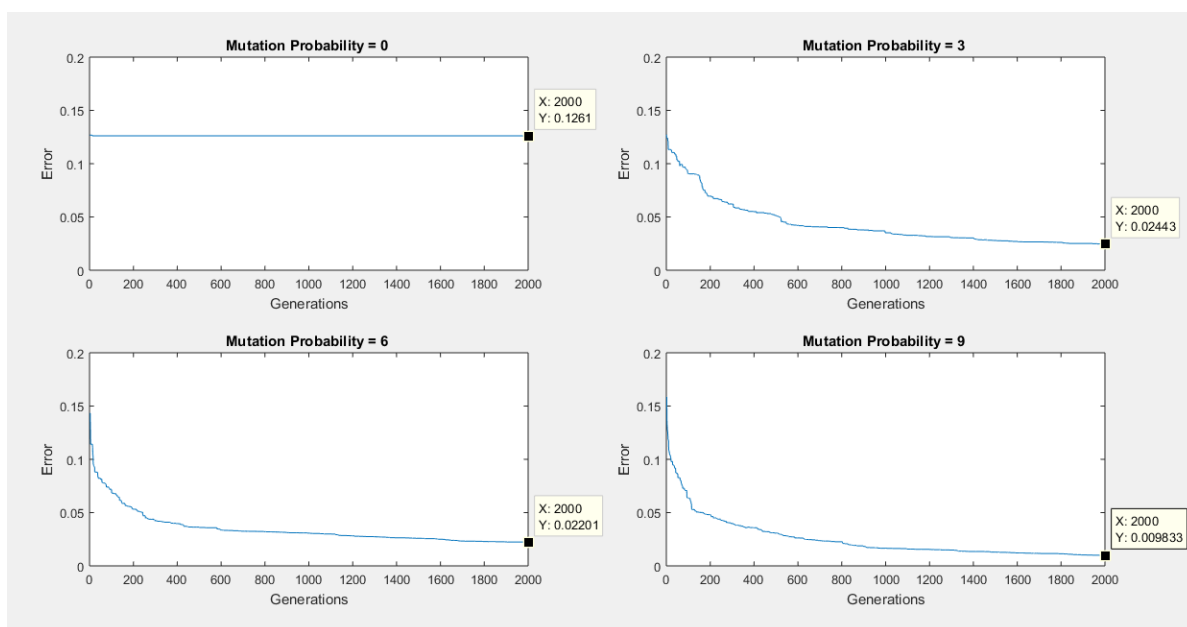
Για την αξιολόγηση του αλγόριθμου, θα χρησιμοποιηθεί η  $f(x, y) = \sin(x + y) * \sin(x^2)$ , με  $x, y \in (-2, 2)$ , έτσι ώστε να παραχθούν τα απαιτούμενα δεδομένα εισόδου-εξόδου.

Αρχικά, με σκοπό την εύρεση των ορίων της  $f$ , υπολογίζεται, μέσω προσέγγισης, η μέγιστη και η ελάχιστη τιμή της,  $f_{max}$  και  $f_{min}$ , αντίστοιχα. Με αυτόν τον τρόπο, ορίζουμε, το εύρος τιμών, των πλατών ( $A$ ) των γκαουσιανών. Πιο συγκεκριμένα,

- Πλάτος:  $A \in (f_{min}, f_{max})$ .
- Κέντρα:  $c_1, c_2 \in (a - 2, b + 2)$ , όπου  $a, b$  το πεδίο ορισμού της συνάρτησης.
- Τυπικές αποκλίσεις:  $\sigma_1, \sigma_2 \in (0.2, 1.3)$ .

Ο πληθυσμός ορίζεται στα 50 χρωμοσώματα, με τον πληθυσμό της επόμενης γενιάς να αποτελείται από το 30% των καλύτερων της προηγούμενης, το 10% τυχαίων λύσεων και το υπόλοιπο 60% δημιουργείται μέσω του crossover.

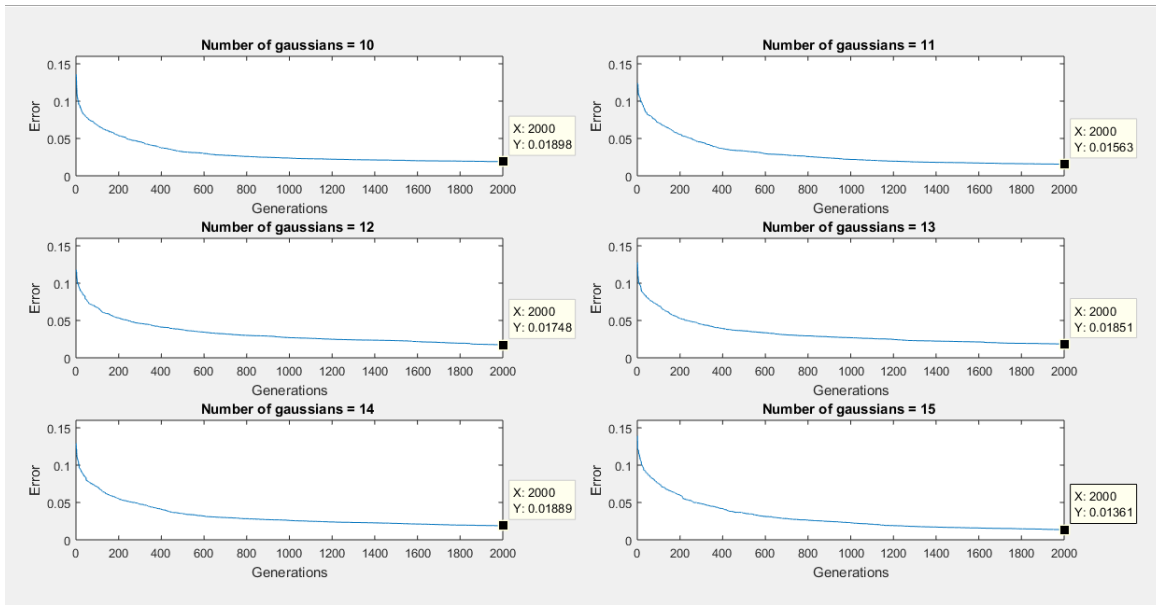
Για την επιλογή της πιθανότητας του *mutation*, αλλά και για να αποδειχθεί η σημαντικότητα των μεταλλάξεων στην σύγκλιση του αλγορίθμου στην βέλτιστη δυνατή λύση, ο αλγόριθμος έτρεξε για σταθερό αριθμό επαναλήψεων, συγκεκριμένα για 2000 επαναλήψεις και παρατηρούμε ότι



Σχήμα 1: Σφάλμα προσέγγισης συναρτήσεως της πιθανότητας μετάλλαξης

όταν δεν συμβαίνουν μεταλλάξεις, ο αλγόριθμος συγκλίνει σε μια τοπικά βέλτιστη λύση, που παρουσιάζει, ωστόσο, μεγάλο σφάλμα χωρίς να έχει δυνατότητα να «εξερευνήσει» γειτονικές περιοχές για την εύρεση καλύτερης λύσης. Επιπλέον, από το διάγραμμα φαίνεται ότι για πιθανότητα μετάλλαξης ίση με 9, ο αλγόριθμος παρουσιάζει συνεχής μείωση στο σφάλμα προσέγγισης.

Τα αποτελέσματα του παραπάνω διαγράμματος, είναι με την χρήση 15 γκαουσιανών. Με σταθερή την πιθανότητα μετάλλαξης, για τον ίδιο αριθμό επαναλήψεων και μεταβάλλοντας τον αριθμό των γκαουσιανών παρατηρούμε

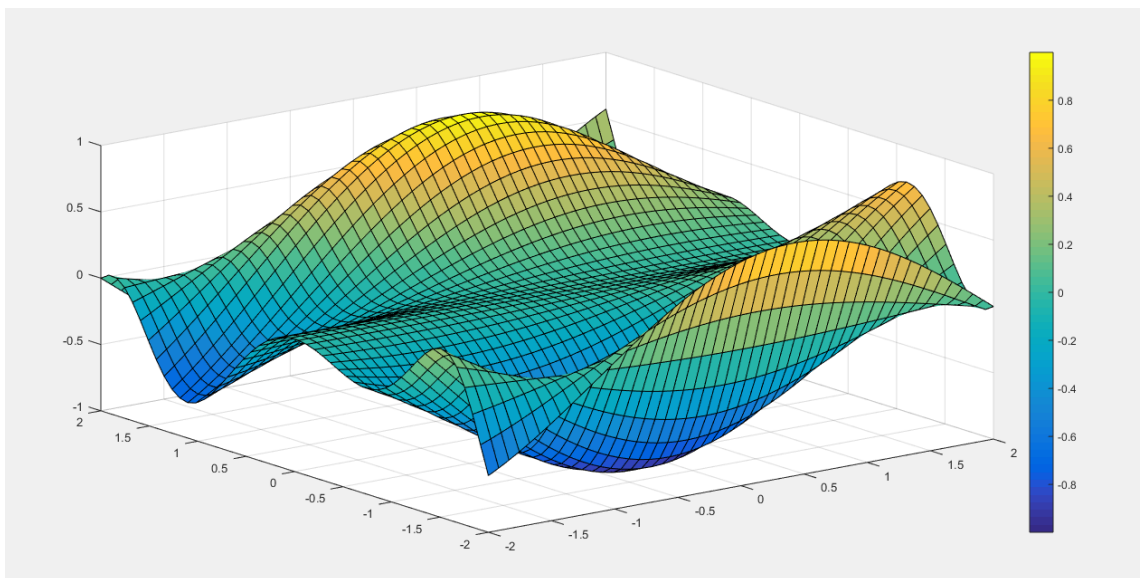


Σχήμα 2: Σφάλμα για διαφορετικό αριθμό γκαουσιανών

ότι το ελάχιστο σφάλμα, εμφανίζεται με την χρήση των 15 γκαουσιανών.

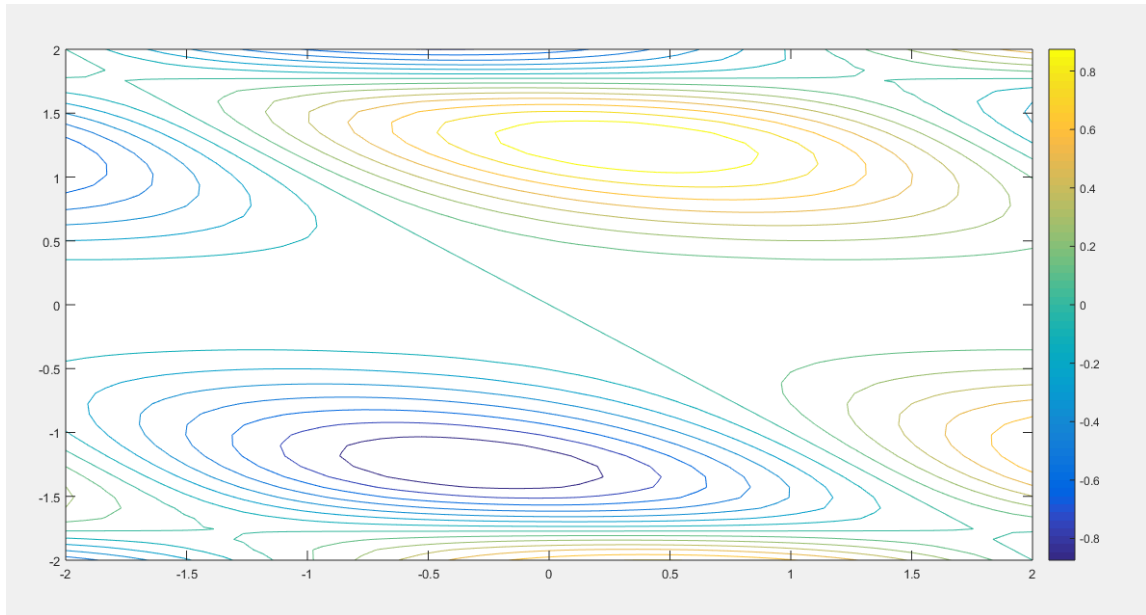
**Σημείωση:** Για το παραπάνω διάγραμμα, ο αλγόριθμος εκτελέστηκε από 10 φορές για κάθε περίπτωση, έτσι ώστε να εξαλειφθεί η τυχαιότητα και ο πιθανοκρατικός χαρακτήρας των αποτελεσμάτων. Σε κάθε περίπτωση εμφανίζεται ο μέσος όρος των σφαλμάτων.

Η πραγματική μορφή τη  $f$ , παρουσιάζεται παρακάτω



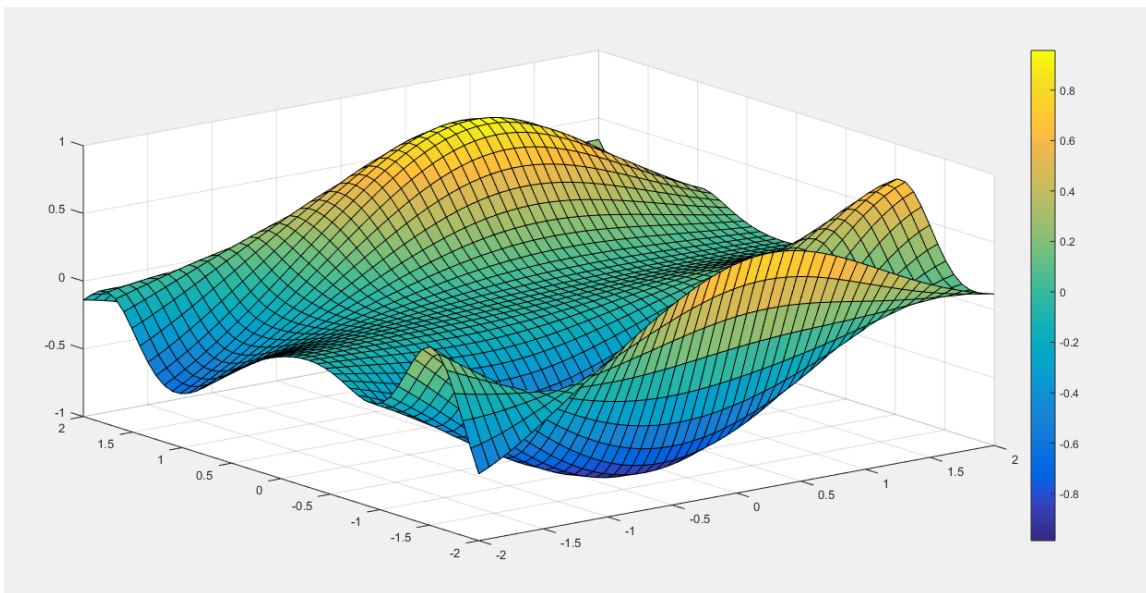
Σχήμα 3: 3D απεικόνιση της  $f$

και αντίστοιχα οι ισοβαρείς της



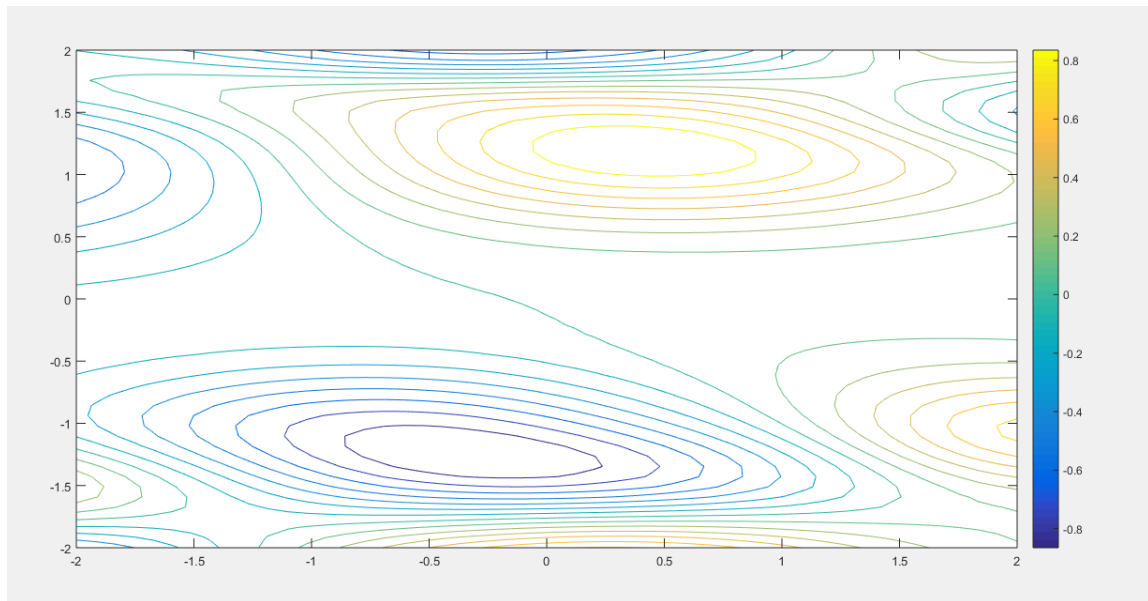
Σχήμα 4: Ισοβαρείς καμπύλες της  $f$

Εκτελώντας τον αλγόριθμο και μετά την διέλευση 10000 επαναλήψεων, έχουμε το εξής αποτέλεσμα



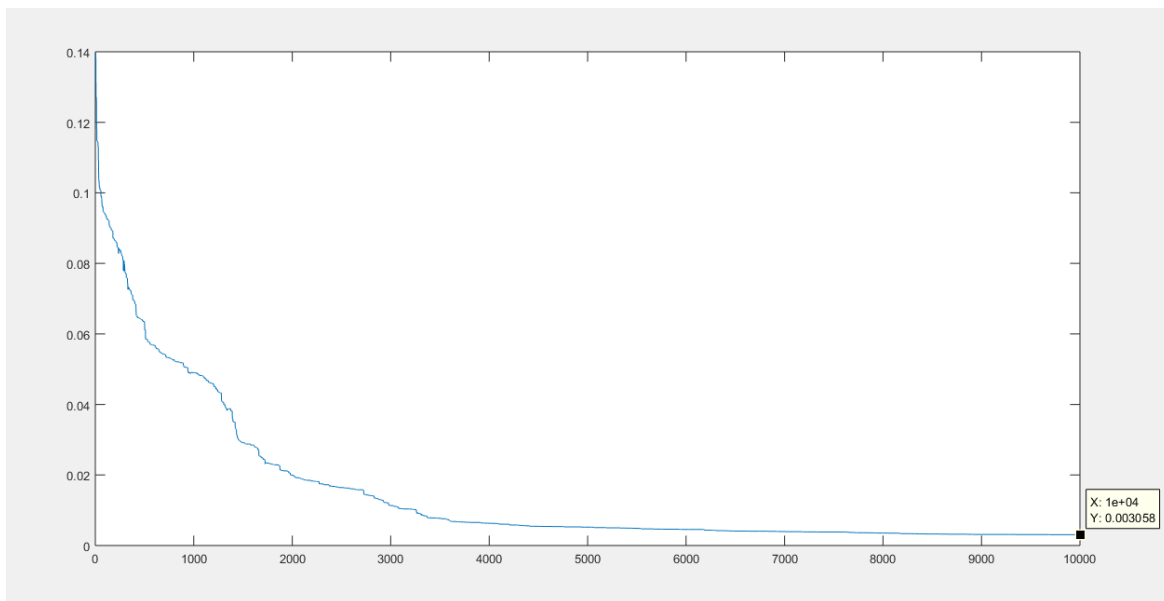
Σχήμα 5: 3D απεικόνιση της προσέγγισης



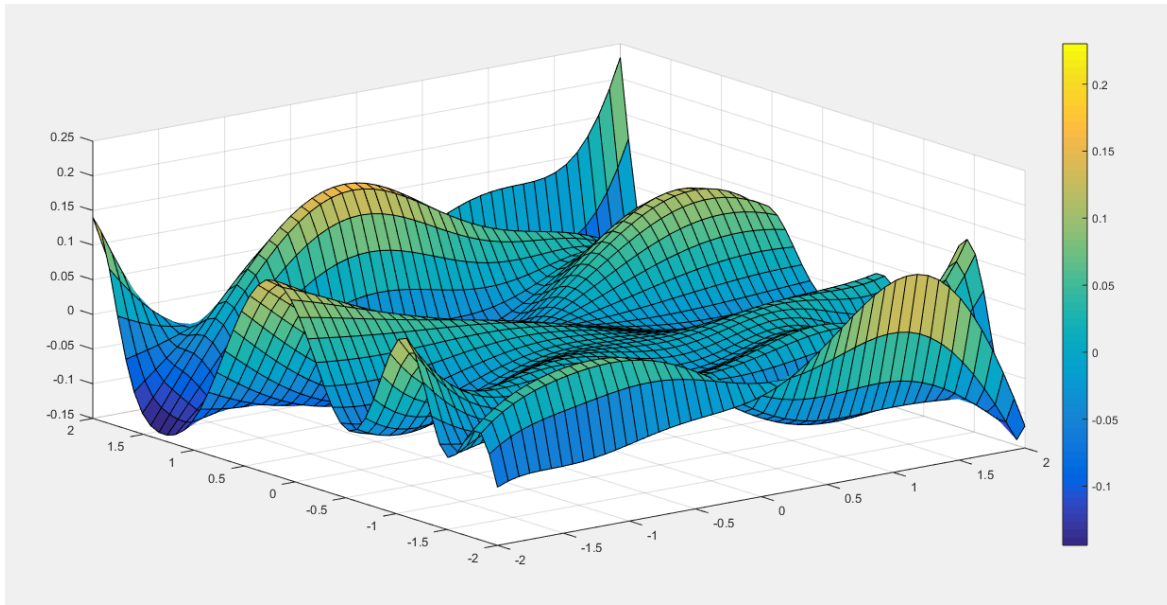


Σχήμα 6: Ισοβαρείς καμπύλες της προσέγγισης

Με το σφάλμα προσέγγισης,



Σχήμα 7: Σφάλμα προσέγγισης ανά γενεά



Σχήμα 8: 3D απεικόνιση του σφάλματος

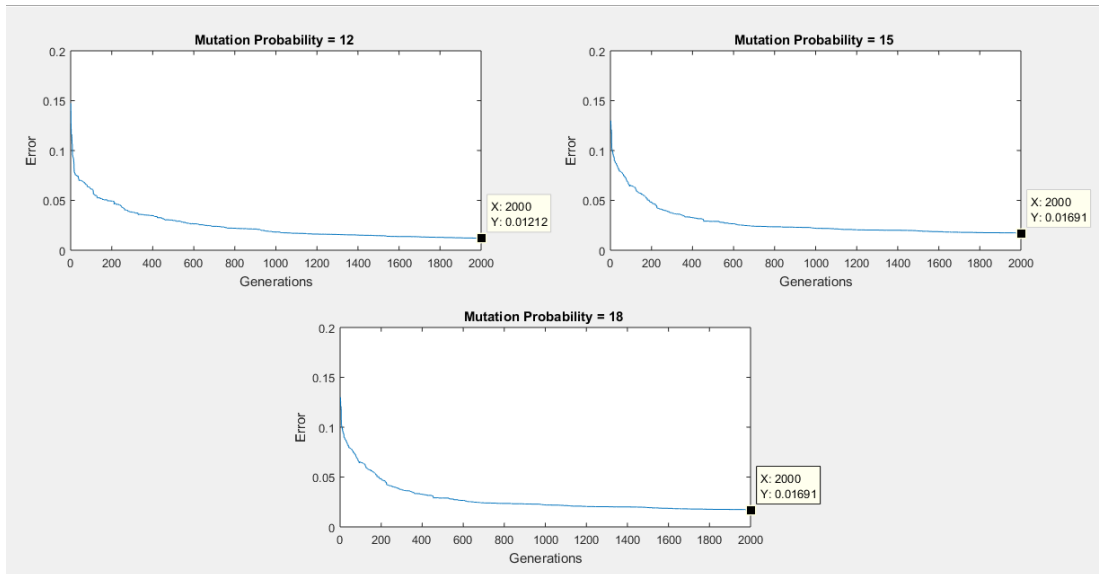
Όπου παρατηρείται, η αδυναμία του αλγορίθμου να προσεγγίσει την συνάρτηση στα άκρα του διαστήματος, όπου και εμφανίζεται και η μεγαλύτερη απόκλιση (0.25) από την πραγματική τιμή.

Για την παραγωγή των δεδομένων εισόδου-εξόδου, χρησιμοποιήθηκαν 25 σημεία για κάθε ένα από το  $x, y$  στο διάστημα  $(-2, 2)$ . Για την δημιουργία των γραφημάτων χρησιμοποιήθηκαν 50 σημεία για κάθε ένα από το  $x, y$  στο διάστημα  $(-1.95, 1.95)$ , έτσι ώστε να χρησιμοποιηθούν διαφορετικά σημεία, για τις δύο αυτές φάσεις.

## Συμπεράσματα

Έγινε προσπάθεια, ο αλγόριθμος να είναι γενικός και να μην εξαρτάται από την εκάστοτε συνάρτηση και τις ιδιότητες της, ωστόσο κάθε συνάρτηση έχει τα δικά της χαρακτηριστικά και μπορεί να χρειαστούν μικρές αλλαγές, πχ στα όρια των τιμών των γονιδίων.

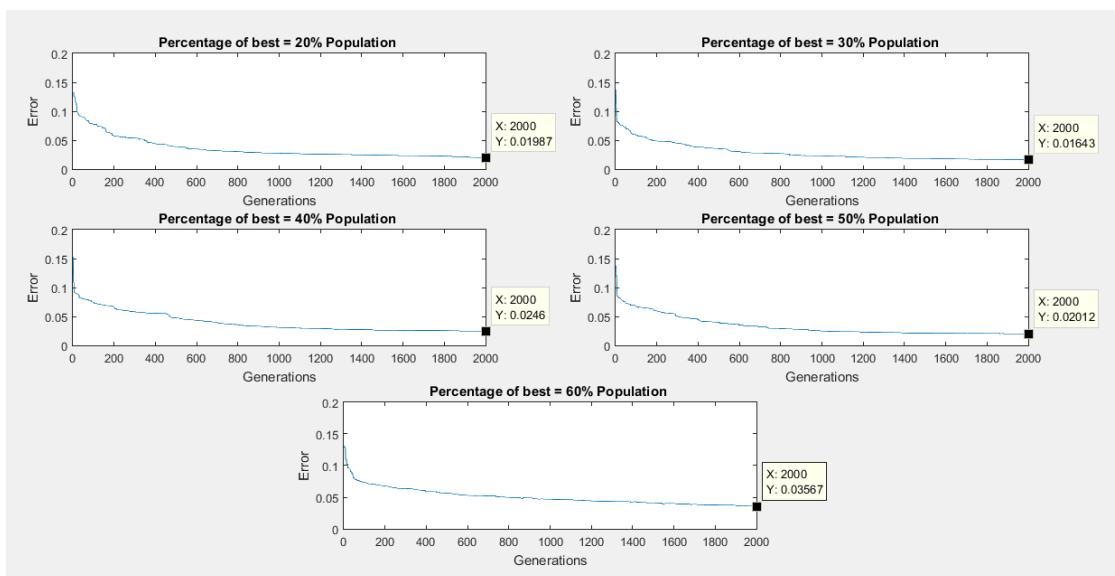
Με το σχήμα 1, μπορεί να δημιουργηθεί η άποψη ότι μεγαλώνοντας το *mutation rate*, βελτιώνονται τα αποτελέσματα του αλγορίθμου. Ωστόσο, κάτι τέτοιο δεν ισχύει αφού οι συνεχείς μεταλλάξεις, αποτρέπουν την σύγκλιση του αλγορίθμου στην λύση.



Σχήμα 9: Αύξηση πιθανότητας μετάλλαξης

Κατασκευάζοντας τον αλγόριθμο, έπρεπε να επιλέγει το ποσοστό του πληθυσμού -της επόμενης γενιάς- που θα αναλογεί στις καλύτερες λύσεις, δηλαδή στα χρωμοσώματα με το χαμηλότερο σφάλμα.

Οι δοκιμές που έγιναν, για την επιλογή του παρουσιάζονται στο παρακάτω διάγραμμα

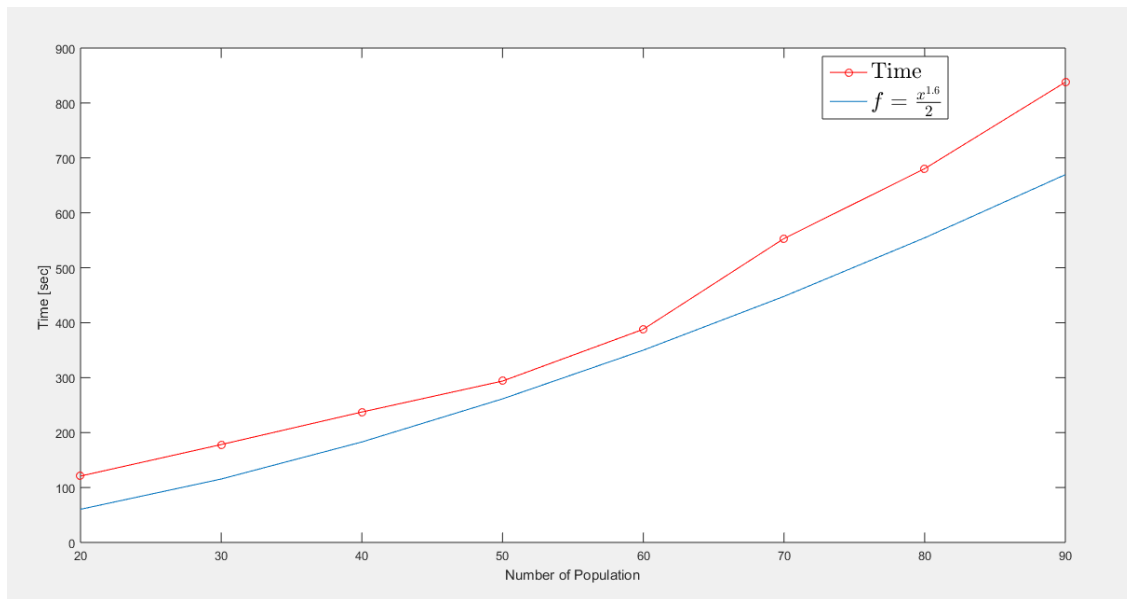


Σχήμα 10: Ποσοστό καλύτερων και σφάλμα προσέγγισης

όπου παρατηρείται ότι η βέλτιστη επιλογή δεν είναι ούτε ένα πολύ μικρό ποσοστό της τάξεως του 20%, καθώς απορρίπτονται λύσεις που μπορεί να βοηθούσαν στην σύγκλιση του αλγορίθμου, αλλά ούτε και ένα μεγάλο ποσοστό της τάξεως του 50% – 60%, καθώς έτσι οδηγείτε ο αλγόριθμος σε μια τοπικά βέλτιστη λύση.

Έτσι, επιλέχθηκε μια ενδιάμεση λύση, εκείνη του 30%. Θέλοντας, τα χρωμοσώματα που δημιουργούνται μέσω της διασταύρωσης (*crossover*) να αποτελούν το 60% του πληθυσμού και για να παραμένει σταθερός ο πληθυσμός, το υπόλοιπο 10% αποτελείται από τυχαίες λύσεις.

Θέλοντας να εξεταστεί η πολυπλοκότητα, εκτελώ τον αλγόριθμο, για σταθερό αριθμό επαναλήψεων = 1000 και μεταβάλλοντας τον αριθμό του πληθυσμού. Προκύπτει το εξής διάγραμμα



Σχήμα 11: Χρόνος εκτέλεσης και Αριθμός Πληθυσμού

**Σημείωση:** Ο αλγόριθμος εκτελέστηκε 10 φορές για κάθε διαφορετική περίπτωση και στο διάγραμμα παρουσιάζεται ο μέσος χρόνος εκτέλεσης.

Παρατηρείται, ότι ο πειραματικός χρόνος εκτέλεσης μπορεί να προσεγγιστεί από την  $f = \frac{x^{1.6}}{2}$  και έτσι συμπεραίνουμε, ο αλγόριθμος είναι  $\theta(n^{1.6})$ , δηλαδή  $T(n) \in \theta(n^{1.6})$ , όπου  $T(n)$  ο χρόνος εκτέλεσης του αλγορίθμου.

Η τελική λύση του αλγορίθμου φαίνεται παρακάτω

Αριθμός Γκαουσιανής	Παράμετρος				
	Πλάτος (A)	Κέντρο ( $c_1$ )	Κέντρο ( $c_2$ )	Τυπική Απόκλιση ( $\sigma_1$ )	Τυπική Απόκλιση ( $\sigma_2$ )
1 <sup>η</sup>	0.8071	-1.1086	2.1135	0.3305	0.6854
2 <sup>η</sup>	-0.5305	-1.0018	-0.8904	0.3230	0.8717
3 <sup>η</sup>	-0.8455	-1.4577	0.1640	0.2919	1.0360
4 <sup>η</sup>	-0.9883	1.4453	-2.5365	0.6909	1.0252
5 <sup>η</sup>	-0.9893	1.5038	3.1356	0.2433	1.2732
6 <sup>η</sup>	-0.9933	2.0408	-0.3329	0.2180	1.0254
7 <sup>η</sup>	0.9559	-2.9086	0.2213	0.8227	0.8848
8 <sup>η</sup>	0.1625	-0.5838	0.6082	0.3520	1.1561
9 <sup>η</sup>	0.9919	1.2047	0.4660	0.3963	0.9732
10 <sup>η</sup>	0.9968	-1.5176	-3.2227	0.2593	1.1650
11 <sup>η</sup>	0.8718	1.8056	-2.1983	0.3517	1.2932
12 <sup>η</sup>	0.9905	3.0050	1.9290	0.8120	0.5761
13 <sup>η</sup>	0.9925	-2.9645	0.2188	0.7285	0.9170
14 <sup>η</sup>	-0.3088	-1.0378	-0.1932	0.5347	0.8861
15 <sup>η</sup>	-0.9943	-2.2967	-2.3792	0.3636	0.7453