

Τεχνικές Βελτιστοποίησης: Project

Ιωάννης Μιχάλαϊνας AEM:10902

Δεκέμβριος 2025

Αποθετήριο Κώδικα:

github.com/ioannisam/optimization-techniques

Περίληψη

Η παρούσα εργασία πραγματεύεται την προσέγγιση μιας άγνωστης μη γραμμικής συνάρτησης δύο μεταβλητών, $f(u_1, u_2)$, μέσω ενός Γενετικού Αλγορίθμου. Στόχος είναι η εύρεση μιας αναλυτικής έκφρασης που αποτελείται από γραμμικό συνδυασμό Γκαουσιανών συναρτήσεων. Ο αλγόριθμος υλοποιήθηκε εξ ολοκλήρου στο MATLAB, χωρίς τη χρήση έτοιμων εργαλείων βελτιστοποίησης. Η διαδικασία περιλαμβάνει την εκπαίδευση του μοντέλου, την επαλήθευση σε νέα δεδομένα για την αποφυγή υπερπροσαρμογής και τέλος, μια διαδικασία απλοποίησης για τη μείωση της πολυπλοκότητας της τελικής έκφρασης.

Περιεχόμενα

1	Εισαγωγή	2
2	Γενετικός Αλγόριθμος	3
2.1	Αλγόριθμος	4
2.2	Εκπαίδευση	5
2.2.1	Παράμετροι	5
2.2.2	Αρχικοποίηση	7
2.2.3	Κωδικοποίηση	7
2.2.4	Συνάρτηση Αξιολόγησης	7
2.2.5	Γενετικοί Τελεστές	7
2.3	Αξιολόγηση	8
2.3.1	Επαλήθευση	8
2.3.2	Απλοποίηση	8
3	Αποτελέσματα	9
3.1	Γραφικές Παραστάσεις	9
3.2	Ευρήματα	10
3.3	Παρατηρήσεις	10
A'	Εργαλεία	11

1 Εισαγωγή

Στο πλαίσιο της παρούσας εργασίας, εξετάζεται το πρόβλημα της εκτίμησης μιας άγνωστης, μη γραμμικής συνάρτησης δύο μεταβλητών $y = f(u_1, u_2)$, βάσει μετρήσεων εισόδου-εξόδου. Η αναλυτική μορφή της f θεωρείται άγνωστη, ενώ είναι γνωστό ότι πρόκειται για συνεχή συνάρτηση των μεταβλητών u_1 και u_2 .

Για την προσέγγιση της f , χρησιμοποιείται ένα παραμετρικό μοντέλο που βασίζεται σε γραμμικό συνδυασμό Γκαουσιανών συναρτήσεων βάσης. Συγκεκριμένα, η προσεγγιστική συνάρτηση ορίζεται ως:

$$\hat{f}(u_1, u_2) = \sum_{k=1}^K G_k(u_1, u_2), \quad (1)$$

όπου K είναι το πλήθος των Γκαουσιανών συναρτήσεων. Αρχικά τέθηκε το όριο $K \leq 15$, ωστόσο μέσω διαδικασίας ρύθμισης των παραμέτρων, επιλέχθηκε ως βέλτιστη τιμή $K = 6$, όπως αναλύεται αργότερα.

$$G(u_1, u_2) = w \cdot e^{-\left(\frac{(u_1 - c_1)^2}{2\sigma_1^2} + \frac{(u_2 - c_2)^2}{2\sigma_2^2}\right)}. \quad (2)$$

Η παράμετρος w αποτελεί τον συντελεστή στάθμισης του κάθε Γκαουσιανού όρου, ενώ τα c_1, c_2 και σ_1, σ_2 καθορίζουν τη θέση και τη διασπορά του αντίστοιχα. Για κάθε Γκαουσιανή συνάρτηση απαιτείται ο προσδιορισμός πέντε παραμέτρων, με αποτέλεσμα το συνολικό διάνυσμα παραμέτρων να έχει διάσταση $5 \times K$.

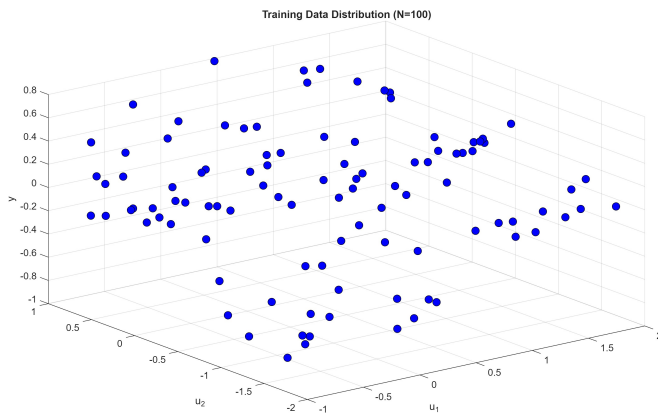
Η διαδικασία εκπαίδευσης του μοντέλου πραγματοποιείται μέσω μετρήσεων εισόδου-εξόδου. Για τις ανάγκες παραγωγής των δεδομένων εκπαίδευσης και επαλήθευσης (διακριτά σύνολα), χρησιμοποιείται η συνάρτηση:

$$f(u_1, u_2) = \sin(u_1 + u_2) \sin(u_2^2), \quad (3)$$

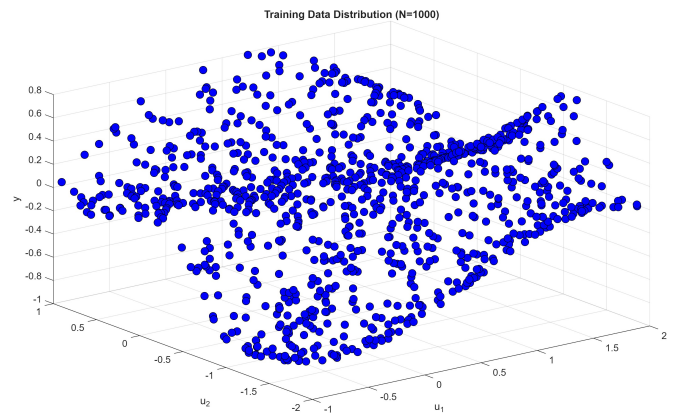
με:

$$u_1 \in [-1, 2], \quad u_2 \in [-2, 1].$$

Στα Σχήματα 1 και 2 οπτικοποιείται η συνάρτηση $f(u_1, u_2)$ για διαφορετικό πλήθος δειγμάτων. Παρατηρείται ότι με την αύξηση του αριθμού των σημείων εισόδου-εξόδου καθίσταται σαφέστερη η μορφολογία της συνάρτησης.



Σχήμα 1: Οπτικοποίηση της $f(u_1, u_2)$
(100 δείγματα)



Σχήμα 2: Οπτικοποίηση της $f(u_1, u_2)$
(1000 δείγματα)

Λόγω της έντονα μη γραμμικής φύσης του προβλήματος και του μεγάλου αριθμού προς εκτίμηση παραμέτρων, η εύρεση του βέλτιστου διανύσματος παραμέτρων ($w, c_1, \sigma_1, c_2, \sigma_2$) πραγματοποιείται μέσω ενός Γενετικού Αλγορίθμου, ο οποίος παρουσιάζεται και αναλύεται στην επόμενη ενότητα.

2 Γενετικός Αλγόριθμος

Οι Γενετικοί Αλγόριθμοι αποτελούν τεχνικές ολικής βελτιστοποίησης. Είναι ευριστικές μέθοδοι βελτιστοποίησης που εμπνέονται από τη διαδικασία της φυσικής επιλογής, η οποία έχει επιδείξει την επιτυχημένη επίλυση πολύπλοκων προβλημάτων βελτιστοποίησης, όπως η δημιουργία και ανάπτυξη νέων οργανισμών.

Οι Γενετικοί Αλγόριθμοι λειτουργούν πάνω σε έναν πληθυσμό υποψήφιων λύσεων (άτομα), οι οποίες εξελίσσονται μέσω γενετικών τελεστών (επιλογή, διασταύρωση, μετάλλαξη) με στόχο την εύρεση της βέλτιστης λύσης σε πολύπλοκα προβλήματα όπου οι αναλυτικές μέθοδοι είναι δύσκολο να εφαρμοστούν.

Όλοι οι εξελικτικοί αλγόριθμοι θεωρούν έναν πληθυσμό από άτομα (χρωμοσώματα), καθένα από τα οποία αντιπροσωπεύει μία λύση του προβλήματος βελτιστοποίησης. Ο πληθυσμός εξελίσσεται σε γενεές, ως αποτέλεσμα της εφαρμογής διαφόρων εξελικτικών τελεστών. Με αυτό τον τρόπο δημιουργούνται νέα άτομα που παριστάνουν διαφορετικά σημεία στον χώρο αναζήτησης, καλύπτοντας ολοένα και περισσότερες περιοχές αυτού.

Σε κάθε άτομο του πληθυσμού αντιστοιχεί μια τιμή της αντικειμενικής συνάρτησης, η οποία στην ορολογία των εξελικτικών αλγορίθμων καλείται *συνάρτηση ικανότητας* (fitness function). Η διαδικασία επιλογής διασφαλίζει ότι τα άτομα με την καλύτερη απόδοση, η οποία αξιολογείται με βάση τη συνάρτηση ικανότητας, θα έχουν μεγαλύτερη πιθανότητα να επιβιώσουν και να αναπαραχθούν. Κατά συνέπεια, ο πληθυσμός τείνει να εξελιχθεί προς λύσεις με βέλτιστη απόδοση.

Η δημιουργία νέων λύσεων (απογόνων) επιτυγχάνεται κυρίως μέσω δύο μηχανισμών:

1. Διασταύρωση: Αποτελεί τον κύριο μηχανισμό ανταλλαγής πληροφορίας. Στην παρούσα εργασία, εφαρμόζεται η *Αριθμητική Διασταύρωση*. Δοθέντων δύο γονέων \mathbf{p}_1 και \mathbf{p}_2 , παράγονται δύο απόγονοι $\mathbf{c}_1, \mathbf{c}_2$ μέσω γραμμικού συνδυασμού:

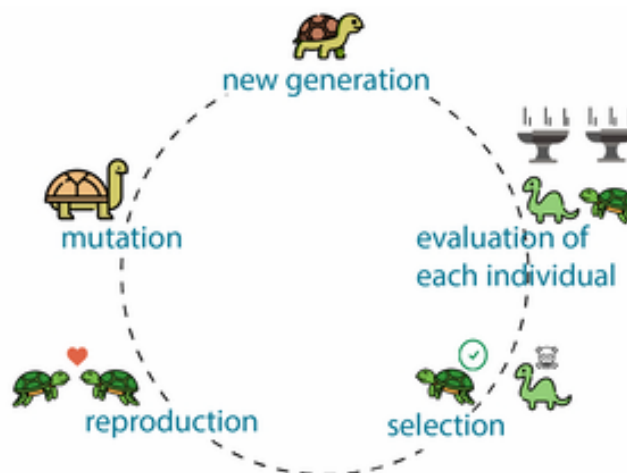
$$\begin{aligned}\mathbf{c}_1 &= \alpha \cdot \mathbf{p}_1 + (1 - \alpha) \cdot \mathbf{p}_2 \\ \mathbf{c}_2 &= (1 - \alpha) \cdot \mathbf{p}_1 + \alpha \cdot \mathbf{p}_2\end{aligned}\tag{4}$$

όπου α είναι ένας τυχαίος αριθμός στο διάστημα $(0, 1)$. Η μέθοδος αυτή εξασφαλίζει ότι οι απόγονοι θα βρίσκονται εντός του υπερ-ορθογωνίου που ορίζουν οι γονείς.

2. Μετάλλαξη: Η μετάλλαξη εισάγει τυχαίες μικρές διαταραχές στα γονίδια των απογόνων, αποτρέποντας την πρόωρη σύγκλιση σε τοπικά ακρότατα. Στην παρούσα υλοποίηση χρησιμοποιείται προσθετικός Γκαουσιανός θόρυβος:

$$x'_i = x_i + \delta, \quad \text{όπου } \delta \sim \mathcal{N}(0, \sigma^2)\tag{5}$$

όπου x_i η τιμή του γονιδίου και δ ο θόρυβος που προκύπτει από την κανονική κατανομή με διασπορά που καθορίζεται από την παράμετρο θορύβου (mutation_noise).



Σχήμα 3: Εικονογράφηση των βημάτων ενός Γενετικού Αλγορίθμου⁵

2.1 Αλγόριθμος

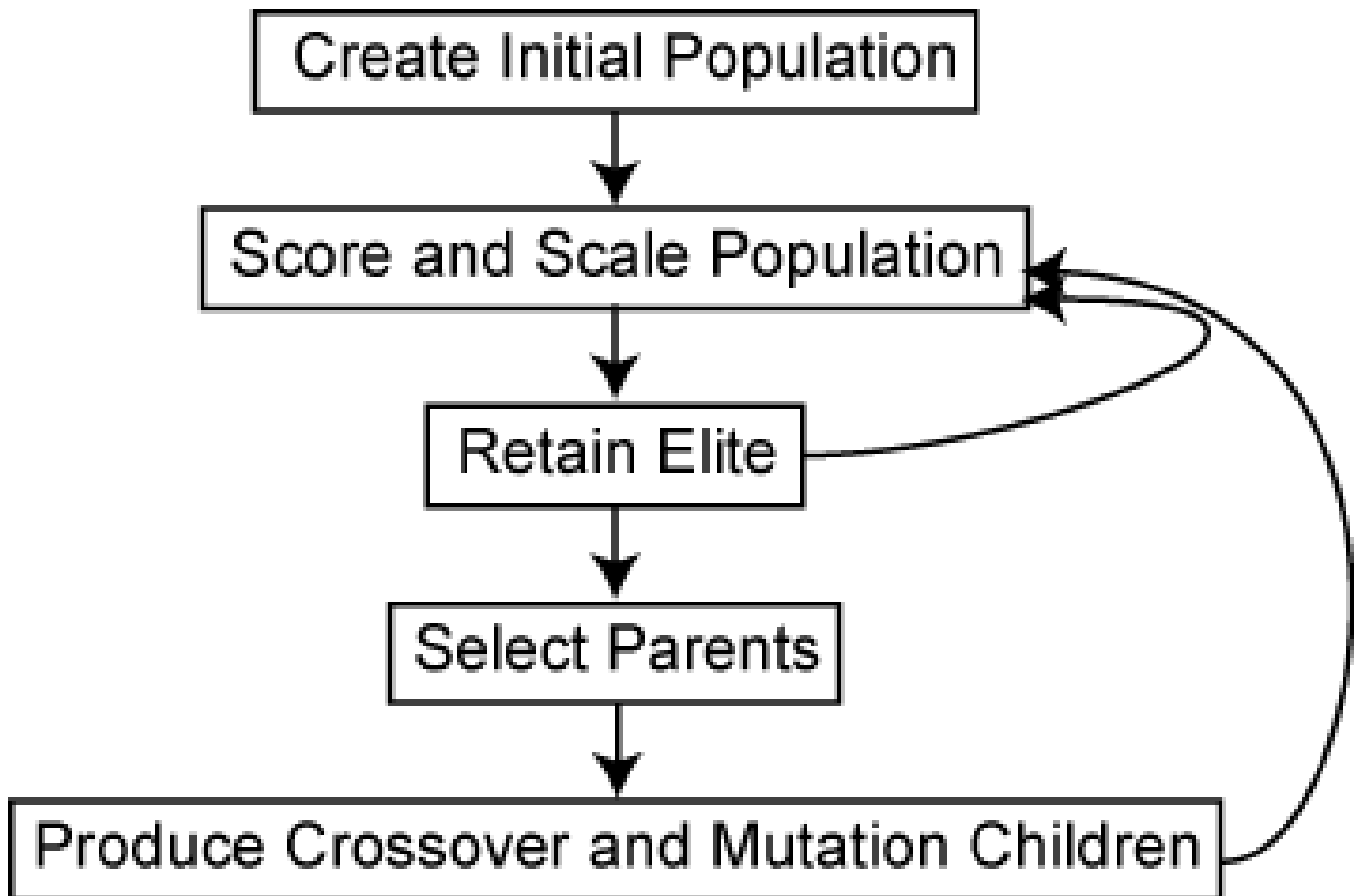
Εδώ παρουσιάζεται η παραπάνω διαδικασία του αλγοριθμικά σε ψευδοκώδικα (Αλγόριθμος 1) και διαγραμματικά (Σχήμα 4).

Algorithm 1 Γενετικός Αλγόριθμος Βελτιστοποίησης Παραμέτρων

Require: Μέγεθος πληθυσμού N , αριθμός Γκαουσιανών K , μέγιστος αριθμός γενεών $MaxGen$, πιθανότητα διασταύρωσης p_c , πιθανότητα μετάλλαξης p_m

Ensure: Βέλτιστο διάνυσμα παραμέτρων θ_{best}

- 1: **Αρχικοποίηση:** Δημιουργία τυχαίου πληθυσμού P_0 μεγέθους N εντός των επιτρεπτών ορίων.
 - 2: **Αξιολόγηση:** Υπολογισμός MSE και Fitness $f(x) = \frac{1}{1+MSE(x)}$ για κάθε άτομο $x \in P_0$.
 - 3: **for** $gen = 1$ **to** $MaxGen$ **do**
 - 4: **Ελιτισμός:** Εντοπισμός και αποθήκευση του καλύτερου ατόμου x_{best} της τρέχουσας γενιάς.
 - 5: **Επιλογή:** Επιλογή N γονέων από τον P_{gen-1} με τη μέθοδο του Τροχού της Τύχης.
 - 6: **Διασταύρωση:** Εφαρμογή Διασταύρωσης σε ζεύγη γονέων με πιθανότητα p_c .
 - 7: **Μετάλλαξη:** Εφαρμογή θορύβου στα γονίδια των απογόνων με πιθανότητα p_m .
 - 8: **Επαναφορά Ορίων:** Διόρθωση τιμών που υπερβαίνουν τα όρια αναζήτησης.
 - 9: $P_{gen} \leftarrow$ Απόγονοι.
 - 10: **Εισαγωγή Elite:** Αντικατάσταση του πρώτου ατόμου του P_{gen} με το x_{best} .
 - 11: **Αξιολόγηση:** Υπολογισμός Fitness για τον νέο πληθυσμό P_{gen} .
 - 12: **end for**
 - 13: **return** Το άτομο με το μέγιστο Fitness από την τελευταία γενιά, θ_{best} .
-



Σχήμα 4: Διαγραμματική παρουσίαση του αλγορίθμου³

2.2 Εκπαίδευση

Στην παρούσα ενότητα περιγράφεται αναλυτικά η διαδικασία που ακολουθήθηκε για την εκπαίδευση του Γενετικού Αλγορίθμου, συμπεριλαμβανομένου του καθορισμού των παραμέτρων και υπερπαραμέτρων, της αρχικοποίησης του πληθυσμού, της κωδικοποίησης των λύσεων, καθώς και της επιλογής και εφαρμογής των γενετικών τελεστών.

2.2.1 Παράμετροι

Κατά την εκπόνηση της εργασίας ορίστηκαν ορισμένες παράμετροι και υπερπαραμέτροι, η επιλογή των οποίων είναι ιδιαίτερης σημασίας, καθώς επηρεάζουν καθοριστικά τη συμπεριφορά και την απόδοση του αλγορίθμου. Η κατάλληλη ρύθμισή τους επιτρέπει την επίτευξη ικανοποιητικής ισορροπίας μεταξύ ακρίβειας και υπολογιστικής αποδοτικότητας. Οι παράμετροι είναι κεντροποιημένες στο αρχείο `parameters.m` για την εύκολη διαχείριση και ρύθμισή τους.

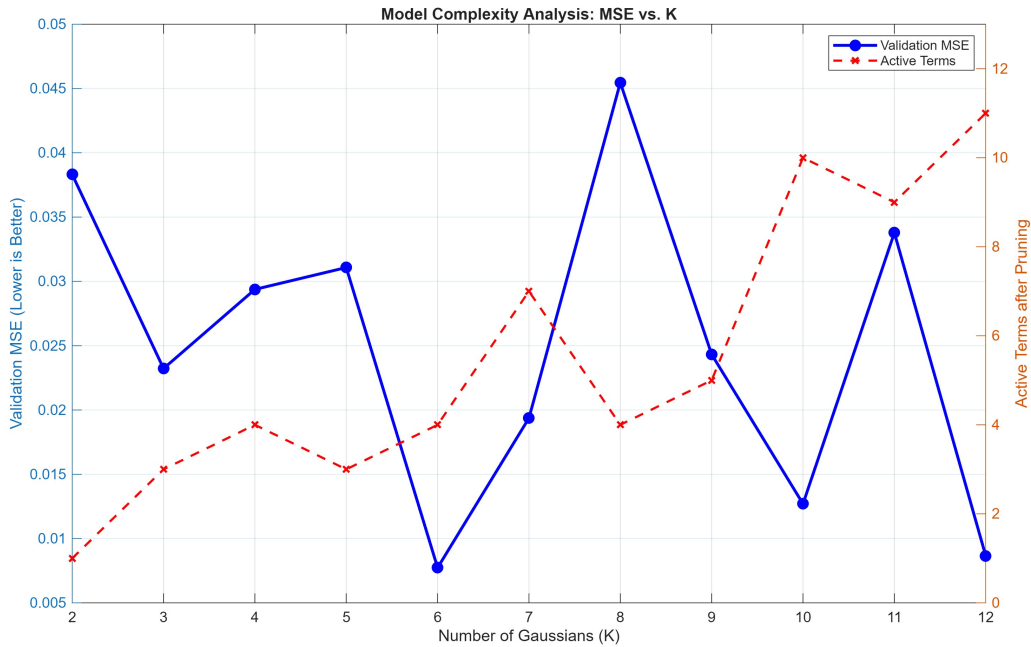
Πριν την τελική εκπαίδευση, πραγματοποιήθηκε συστηματική αναζήτηση για τον βέλτιστο καθορισμό των υπερπαραμέτρων. Στον Πίνακα 1 που ακολουθεί συνοψίζονται οι επιλεγμένες τιμές των παραμέτρων, καθώς και το σκεπτικό που οδήγησε στην επιλογή τους.

Παράμετρος	Τιμή	Περιγραφή / Αιτιολόγηση
num_samples	1000	Πλήθος δειγμάτων εκπαίδευσης για επαρκή κάλυψη του πεδίου ορισμού.
c1_range	$[-1, 2]$	Όρια αναζήτησης για τα κέντρα c_1 , σύμφωνα με το πεδίο ορισμού του u_1 .
c2_range	$[-2, 1]$	Όρια αναζήτησης για τα κέντρα c_2 , σύμφωνα με το πεδίο ορισμού του u_2 .
w_range	$[-10, 10]$	Περιορισμός του εύρους των βαρών για αποφυγή ακραίων τιμών.
sigma_range	$[0.2, 2.0]$	Όρια για το πλάτος των Γκαουσιανών, ώστε να αποφεύγονται τόσο υπερβολικά αιχμηρές όσο και υπερβολικά επίπεδες συναρτήσεις.
pop_size	100	Μέγεθος πληθυσμού που προσφέρει ισορροπία μεταξύ υπολογιστικού κόστους και ποικιλομορφίας.
num_gaussians	6	Επιλεγμένος αριθμός Γκαουσιανών όρων βάσει ελαχιστοποίησης του σφάλματος επαλήθευσης.
max_generations	3000	Μέγιστος αριθμός γενεών για την εξασφάλιση σύγκλισης του αλγορίθμου.
crossover_rate	0.6	Πιθανότητα διασταύρωσης που επιτρέπει αποτελεσματική εξερεύνηση του χώρου λύσεων χωρίς αστάθεια.
mutation_rate	0.1	Πιθανότητα μετάλλαξης για διατήρηση της ποικιλομορφίας και αποφυγή πρόωρης σύγκλισης.
mutation_noise	0.05	Ένταση του θορύβου που προστίθεται κατά τη μετάλλαξη για ελεγχόμενες τοπικές μεταβολές.
pruning_threshold	0.6	Κατώφλι για την αφαίρεση Γκαουσιανών όρων με μικρή συνεισφορά στο τελικό μοντέλο.

Πίνακας 1: Παράμετροι και υπερπαραμέτροι του Γενετικού Αλγορίθμου

Ρύθμιση Παραμέτρων

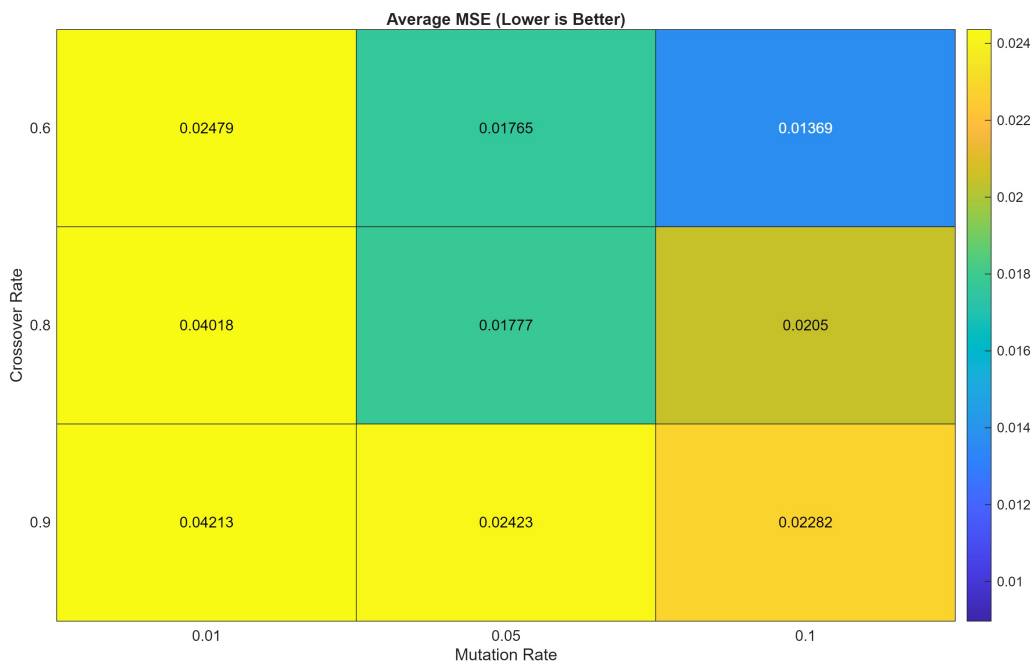
Εξετάστηκε η επίδραση της πολυπλοκότητας του μοντέλου στην ακρίβεια πρόβλεψης. Όπως φαίνεται στο Σχήμα 5, δοκιμάστηκαν τιμές $K \in [2, 12]$. Παρατηρήθηκε ότι για $K = 6$, το σφάλμα επαλήθευσης (Validation MSE) ελαχιστοποιείται. Αξίζει να σημειωθεί ότι η αύξηση του K πέραν του 6 δεν οδήγησε σε βελτίωση, αλλά σε αστάθεια, καθώς ο χώρος αναζήτησης γίνεται υπερβολικά πολύπλοκος λόγω της Κατάρας της Διαστατικότητας (Curse of Dimensionality).



Σχήμα 5: MSE ανά αριθμό Γκαουσιανών

Ρύθμιση Υπερπαραμέτρων

Εφαρμόστηκε αναζήτηση πλέγματος (Grid Search) για τον προσδιορισμό των πιθανοτήτων διασταύρωσης (p_c) και μετάλλαξης (p_m). Τα αποτελέσματα (Σχήμα 6) έδειξαν ότι ο συνδυασμός $p_c = 0.6$ και $p_m = 0.1$ αποδίδει το χαμηλότερο σφάλμα, επιτρέποντας ικανή εξερεύνηση του χώρου λύσεων χωρίς να διαταράσσεται η σύγκλιση.



Σχήμα 6: Heatmap σφάλματος MSE για διάφορους συνδυασμούς p_c και p_m .

2.2.2 Αρχικοποίηση

Η διαδικασία ξεκινά με τη δημιουργία ενός αρχικού πληθυσμού N ατόμων (τυπικά 100 άτομα είναι ικανοποιητικά).¹ Στη συνάρτηση `initialize_population` του κώδικα, κάθε γονίδιο του χρωμοσώματος λαμβάνει μια τυχαία τιμή από μια ομοιόμορφη κατανομή εντός των προκαθορισμένων ορίων (Πίνακας 1). Αυτό εξασφαλίζει ότι ο αλγόριθμος ξεκινά με μια ευρεία κάλυψη του χώρου αναζήτησης πριν αρχίσει η διαδικασία της εξέλιξης.

2.2.3 Κωδικοποίηση

Σε αντίθεση με τους κλασικούς γενετικούς αλγόριθμους που χρησιμοποιούν δυαδική κωδικοποίηση,² στην παρούσα εργασία επιλέχθηκε η κωδικοποίηση πραγματικών αριθμών (real-valued encoding). Κάθε άτομο (χρωμόσωμα) αναπαρίσταται ως ένα διάνυσμα πραγματικών αριθμών που περιέχει τις παραμέτρους των $K = 6$ Γκαουσιανών συναρτήσεων.

Συγκεκριμένα, για κάθε Γκαουσιανή k , απαιτούνται 5 παράμετροι: το βάρος w_k , τα κέντρα $c_{1,k}, c_{2,k}$ και οι τυπικές αποκλίσεις $\sigma_{1,k}, \sigma_{2,k}$. Επομένως, το μήκος του χρωμοσώματος είναι $L = 5 \times 6 = 30$ γονίδια. Η δομή του χρωμοσώματος έχει τη μορφή:

$$\mathbf{x} = [w_1, c_{1,1}, \sigma_{1,1}, c_{2,1}, \sigma_{2,1}, \dots, w_6, c_{1,6}, \sigma_{1,6}, c_{2,6}, \sigma_{2,6}] \quad (6)$$

2.2.4 Συνάρτηση Αξιολόγησης

Η αξιολόγηση της ποιότητας κάθε λύσης γίνεται μέσω του Μέσου Τετραγωνικού Σφάλματος (MSE) μεταξύ της πραγματικής συνάρτησης $f(u_1, u_2)$ και της εκτίμησης $G(u_1, u_2)$ στα δεδομένα εκπαίδευσης.

Επειδή οι γενετικοί αλγόριθμοι συνήθως μεγιστοποιούν μια συνάρτηση ικανότητας (Fitness Function), ενώ εμείς επιθυμούμε την ελαχιστοποίηση του σφάλματος, χρησιμοποιήθηκε ο ακόλουθος μετασχηματισμός στη συνάρτηση `fitness`:

$$Fitness(\mathbf{x}) = \frac{1}{1 + MSE(\mathbf{x})} \quad (7)$$

Με αυτόν τον τρόπο, όσο μικρότερο είναι το σφάλμα ($MSE \rightarrow 0$), τόσο η τιμή `Fitness` πλησιάζει τη μονάδα (μέγιστη τιμή).

2.2.5 Γενετικοί Τελεστές

- **Επιλογή (Selection):** Χρησιμοποιήθηκε η μέθοδος του Τροχού της Τύχης (Roulette Wheel Selection). Σύμφωνα με τη μέθοδο αυτή, κάθε άτομο έχει πιθανότητα επιλογής ανάλογη της ικανότητάς του. Η πιθανότητα P_i για το άτομο i υπολογίζεται ως $P_i = \frac{f_i}{\sum_{j=1}^N f_j}$, όπου f_i η τιμή της συνάρτησης ικανότητας του ατόμου.

Ωστόσο, καθώς προχωρούν οι γενεές και ο γενετικός αλγόριθμος πλησιάζει στη σύγκλιση, η ικανότητα των ατόμων τείνει να έχει την ίδια περίπου τιμή. Κατά συνέπεια, όλα τα άτομα αποκτούν σχεδόν την ίδια πιθανότητα επιβίωσης και ο γενετικός αλγόριθμος κινδυνεύει να εκφυλιστεί σε μέθοδο τυχαίας αναζήτησης.

Για την επίλυση αυτού του προβλήματος, εφαρμόστηκε *κλιμάκωση* της συνάρτησης ικανότητας, ώστε η υψηλότερη τιμή του τρέχοντος πληθυσμού να απεικονίζεται στο 1 και η χαμηλότερη στο 0 (όπως υλοποιείται στη συνάρτηση `selection`).

- **Διασταύρωση (Crossover):** Εφαρμόστηκε αριθμητική διασταύρωση με πιθανότητα 0.6. Δύο γονείς p_1, p_2 παράγουν απογόνους μέσω γραμμικού συνδυασμού: $c_1 = \alpha \cdot p_1 + (1 - \alpha) \cdot p_2$ και $c_2 = (1 - \alpha) \cdot p_1 + \alpha \cdot p_2$, όπου α τυχαίος αριθμός στο $(0, 1)$.
- **Μετάλλαξη (Mutation):** Εφαρμόστηκε μετάλλαξη με προσθήκη Γκαουσιανού θορύβου με πιθανότητα 0.1 ανά γονίδιο. Στις νέες τιμές επιβάλλονται ξανά τα όρια, ώστε να παραμένουν εντός των επιτρεπτών ορίων του προβλήματος.

2.3 Αξιολόγηση

Μετά την ολοκλήρωση του καθορισμένου αριθμού γενεών (`max_generations`), επιλέγεται από τον τελικό πληθυσμό το άτομο με την υψηλότερη τιμή ικανότητας (Elite individual). Αυτό το άτομο αποτελεί τη βέλτιστη λύση που εντόπισε ο αλγόριθμος. Η αξιολόγηση της ποιότητας της λύσης πραγματοποιείται σε τρία επίπεδα:

1. **Υπολογισμός Σφάλματος:** Υπολογίζεται το τελικό Μέσο Τετραγωνικό Σφάλμα (MSE) στα δεδομένα εκπαίδευσης, το οποίο αποτελεί δείκτη της ακρίβειας προσαρμογής του μοντέλου στα γνωστά δεδομένα.
2. **Έλεγχος Σύγκλισης:** Εξετάζεται το διάγραμμα ιστορικού του σφάλματος (MSE History) ανά γενιά, για να επιβεβαιωθεί ότι ο αλγόριθμος έχει συγκλίνει ομαλά και έχει σταθεροποιηθεί σε μια ελάχιστη τιμή.
3. **Οπτική Επιθεώρηση:** Δημιουργούνται τρισδιάστατες γραφικές παραστάσεις για την οπτική σύγκριση της επιφάνειας που παράγει η προσεγγιστική συνάρτηση $\hat{f}(u_1, u_2)$ έναντι της πραγματικής συνάρτησης $f(u_1, u_2)$.

2.3.1 Επαλήθευση

Για να διασφαλιστεί ότι το μοντέλο δεν έχει υποστεί υπερπροσαρμογή (overfitting) στα δεδομένα εκπαίδευσης, εφαρμόζεται διαδικασία επαλήθευσης. Συγκεκριμένα:

- Παράγεται ένα **νέο σύνολο δεδομένων** $N = 1000$ δειγμάτων, χρησιμοποιώντας την ίδια γεννήτρια δεδομένων, αλλά με διαφορετική τυχαία δειγματοληψία εντός του πεδίου ορισμού.
- Το βέλτιστο μοντέλο αξιολογείται σε αυτό το νέο, άγνωστο σύνολο δεδομένων και υπολογίζεται το Σφάλμα Επαλήθευσης (MSE_{val}).

Κριτήριο επιτυχίας αποτελεί η σχέση μεταξύ του σφάλματος εκπαίδευσης (MSE_{train}) και του σφάλματος επαλήθευσης. Εάν $MSE_{val} \approx MSE_{train}$, τότε το μοντέλο γενικεύει σωστά. Αντιθέτως, αν το MSE_{val} είναι σημαντικά μεγαλύτερο, υπάρχει ένδειξη υπερπροσαρμογής.

2.3.2 Απλοποίηση

Στόχος της εργασίας είναι η εύρεση μιας αναλυτικής έκφρασης *χαμηλής πολυπλοκότητας*. Για τον σκοπό αυτό, μετά την εκπαίδευση εφαρμόζεται μια διαδικασία «κλαδέματος» των Γκαουσιανών όρων. Η διαδικασία βασίζεται στο πλάτος (βάρος) w_k κάθε Γκαουσιανής:

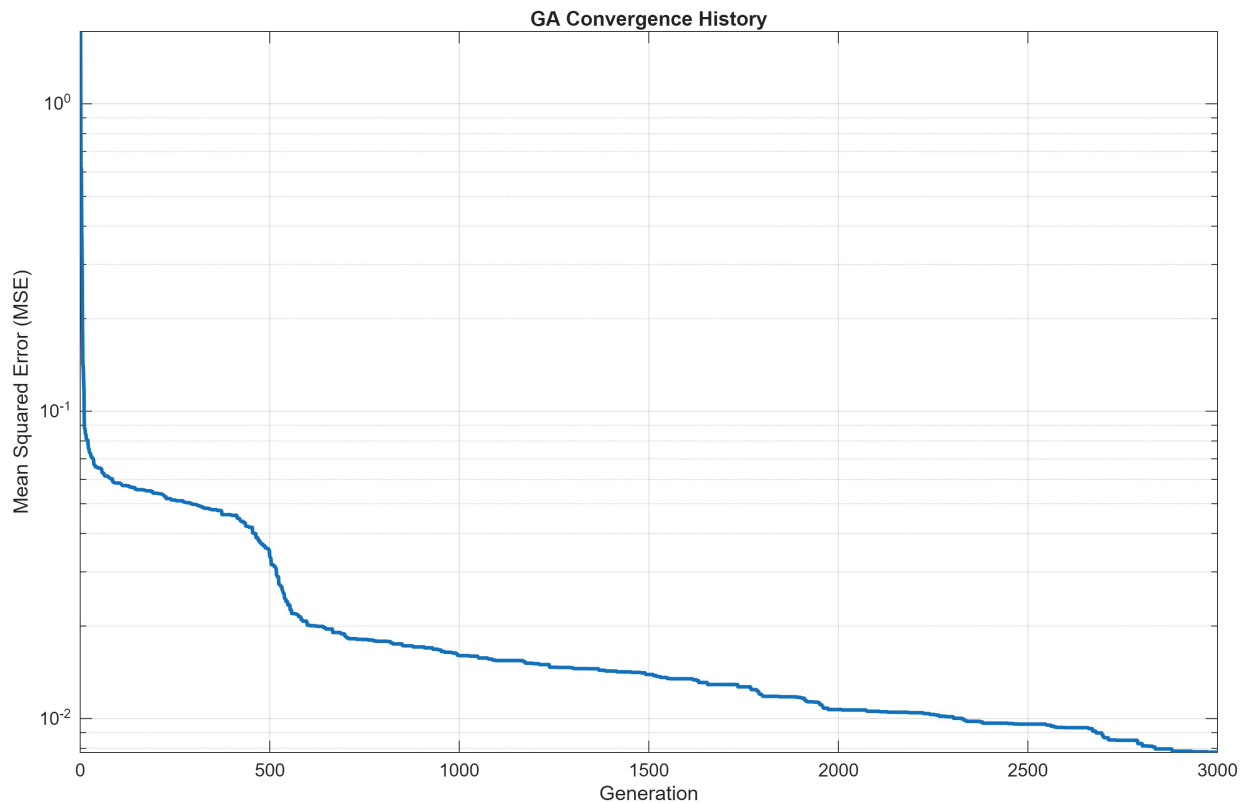
$$\text{Αν } |w_k| < \text{pruning_threshold, τότε } w_k = 0 \quad (8)$$

όπου `pruning_threshold` είναι το κατώφλι απλοποίησης που ορίστηκε στις παραμέτρους. Οι όροι με αμελητέα συνεισφορά αφαιρούνται από την τελική έκφραση. Στη συνέχεια, το απλοποιημένο μοντέλο επαναξιολογείται για να διαπιστωθεί εάν η αφαίρεση των όρων προκάλεσε σημαντική αύξηση του σφάλματος. Εάν η αύξηση είναι μικρή, η απλοποιημένη μορφή υιοθετείται ως η τελική λύση.

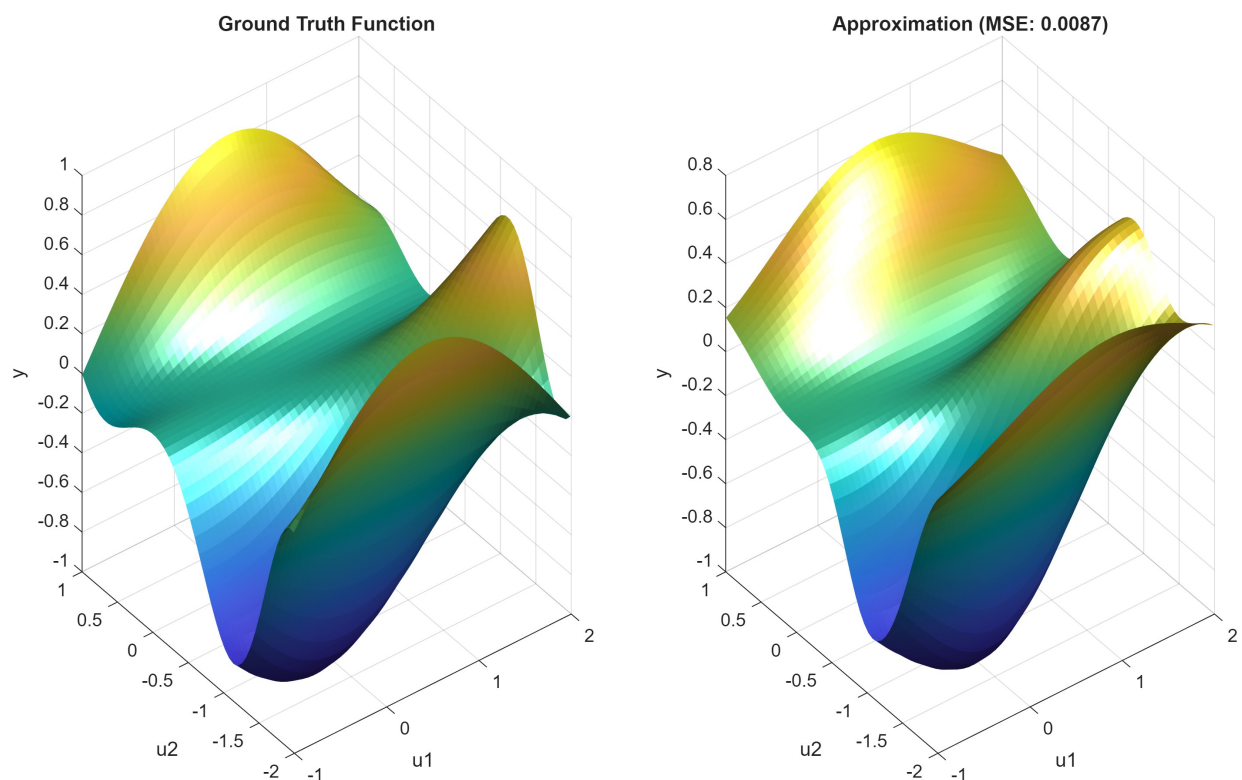
3 Αποτελέσματα

Τα παρακάτω αποτελέσματα προέκυψαν τρέχοντας τον αλγόριθμο για `max_generation` γενιές.

3.1 Γραφικές Παραστάσεις



Σχήμα 7: Σύγκλιση του αλγορίθμου ανά τις γενιές



Σχήμα 8: Οπτική σύγκριση της πραγματικής συνάρτησης με την εκτίμηση

3.2 Ευρήματα

Ο Γενετικός Αλγόριθμος, μετά τη βελτιστοποίηση των παραμέτρων, κατέληξε σε ένα συμπαγές μοντέλο 6 Γκαουσιανών συναρτήσεων. Τα αποτελέσματα είναι θεαματικά βελτιωμένα σε σχέση με την αρχική προσέγγιση των 15 συναρτήσεων

- **Μείωση Πολυπλοκότητας:** Το μοντέλο χρησιμοποιεί μόλις 30 παραμέτρους (αντί για 75).
- **Βελτίωση Ακρίβειας:** Το MSE επαλήθευσης μειώθηκε στο **0.00869** (από ≈ 0.017 με 15 Γκαουσιανές).
- **Βελτίωση Ταχύτητας:** Ο συνολικός χρόνος εκπαίδευσης ανήλθε σε 15.46 δευτερόλεπτα (προηγουμένως 34+ δευτερόλεπτα).

Στον Πίνακα 2 παρουσιάζονται οι παράμετροι των 6 όρων του τελικού μοντέλου. Για μέγιστη ακρίβεια ($MSE=0.008$) απαιτείται η χρήση και του 6ου όρου που αφαιρέθηκε κατά την απλοποίηση.

k	Ενεργός	w_k	$c_{1,k}$	$\sigma_{1,k}$	$c_{2,k}$	$\sigma_{2,k}$
1	Ναι	-0.944	-0.51	0.99	-0.93	0.34
2	Ναι	0.765	0.92	1.69	-0.80	0.61
3	Όχι	-0.482	-0.05	1.34	-0.20	0.96
4	Ναι	1.189	0.53	1.10	-0.33	1.97
5	Ναι	-1.646	0.35	0.98	-1.42	0.33
6	Ναι	-1.253	0.73	1.04	-0.26	0.60

Πίνακας 2: Οι παράμετροι των 6 Γκαουσιανών όρων του μοντέλου.

3.3 Παρατηρήσεις

Από την εκπαίδευση, επαλήθευση και ανάλυση των αποτελεσμάτων του Γενετικού Αλγορίθμου προκύπτουν τα ακόλουθα βασικά συμπεράσματα:

1. **Επίδραση Πολυπλοκότητας Μοντέλου:** Η χρήση περιορισμένου αριθμού συναρτήσεων βάσης ($K = 6$ αντί για $K = 15$) οδήγησε σε σαφώς καλύτερη απόδοση. Η μείωση της διάστασης του χώρου αναζήτησης (από 75 σε 30 παραμέτρους) επέτρεψε στον Γενετικό Αλγόριθμο να συγκλίνει ταχύτερα και πιο αξιόπιστα, αποφεύγοντας τοπικά ελάχιστα και φαινόμενα αστάθειας που σχετίζονται με την Κατάρτα της Διαστατικότητας.
2. **Σύγκλιση Αλγορίθμου:** Ο αλγόριθμος παρουσίασε ταχεία μείωση του σφάλματος κατά τις πρώτες γενεές, ενώ στη συνέχεια ακολούθησε φάση σταδιακής βελτίωσης έως τη γενιά 3000. Το τελικό σφάλμα εκπαίδευσης ήταν $MSE_{\text{train}} = 0.00776$, γεγονός που υποδεικνύει επιτυχή προσαρμογή του μοντέλου στα δεδομένα εκπαίδευσης.
3. **Ικανότητα Γενίκευσης:** Η αξιολόγηση σε ανεξάρτητο σύνολο δεδομένων επαλήθευσης απέδωσε $MSE_{\text{val}} = 0.00869$, τιμή πολύ κοντινή στο σφάλμα εκπαίδευσης. Η μικρή αυτή απόκλιση καταδεικνύει ότι το μοντέλο δεν έχει υποστεί υπερπροσαρμογή και γενικεύει ικανοποιητικά σε άγνωστα δεδομένα.
4. **Απλοποίηση Μοντέλου:** Εφαρμόστηκε διαδικασία απλοποίησης με κατώφλι βάρους $|w| < 0.6$. Καθώς αφαιρέθηκε ένας από τους έξι Γκαουσιανούς όρους (Active Gaussians: 5), παρατηρήθηκε σημαντική αύξηση του σφάλματος επαλήθευσης (από 0.00869 σε 0.08606). Το αποτέλεσμα αυτό υποδεικνύει ότι και οι έξι όροι είναι ουσιώδεις για την ακριβή περιγραφή της συνάρτησης και ότι το μοντέλο των $K = 6$ όρων επιτυγχάνει βέλτιστη ισορροπία μεταξύ ακρίβειας και απλότητας.
5. **Χρόνος Εκτέλεσης:** Η μείωση του μήκους του χρωμοσώματος και η κατάλληλη ρύθμιση των υπερπαραμέτρων οδήγησαν σε σημαντική μείωση του χρόνου εκτέλεσης, καθιστώντας τη μέθοδο αποδοτική και κατάλληλη ακόμη και για εφαρμογές με αυξημένες χρονικές απαιτήσεις.

Α' Εργαλεία

Τα εργαλεία που χρησιμοποιήθηκαν κατά την εκπόνηση της εργασίας είναι τα εξής:

- MATLAB
- LaTeX
- Git

Αναφορές

- [1] Γ. Ροβιθάκης, *Τεχνικές Βελτιστοποίησης*, Εκδόσεις Τζιόλα, 2007.
- [2] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley, 1989.
- [3] MathWorks, What Is the Genetic Algorithm?, *Math Works Documentation*.
- [4] S. Sanskar, How to Generate a Genetic Algorithm with MATLAB, *Medium*, Dec. 2024.
- [5] Generative Design, Genetic Algorithms, *Generative Design Library*.