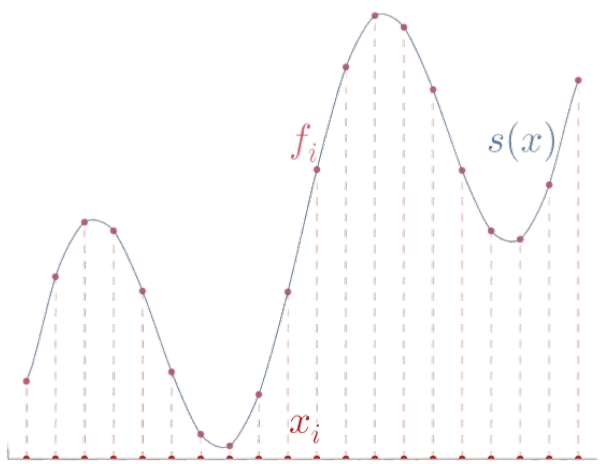
2η Εργασία Στο Μάθημα Της Αριθμητικής Ανάλυσης

Ονοματεπώνυμο : Παπαδόπουλος Αναστάσιος

ΑΕΜ : 3654

Διδάσκων : Τέφας Αναστάσιος

14 Ιανουαρίου 2021



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Περιεχόμενα

1 Εισαγωγή

Η εργασία αυτή αναπτύχθηκε κατά την παρακολούθηση του μάθηματος της Αριθμητικής Ανάλυσης του 3ου εξαμήνου του τμήματος Πληροφορικής ΑΠΘ. Περιλαμβάνει υλοποιήσεις και πειραματισμούς με αλγορίθμους Αριθμητικής Ανάλυσης που αναφέρονται σε θεματικές ενότητες όπως της προσέγγισης συναρτήσεων (**Άσκηση 5 και 7**) για παράδειγμα η μέθοδος των ελαχίστων τετραγώνων και της αριθμητικής ολοκλήρωσης (**Άσκηση 6**) για παράδειγμα η μέθοδος της **Simpson**. Οι αλγόριθμοι αυτοί καθώς και οι γραφικές παραστάσεις που περιέχονται στο παρόν έγγραφο έχουν υλοποιηθεί με την γλώσσα προγραμματισμού **Python** με χρήση των βιβλιοθηκών **NumPy** και **Matplotlib**. Στα παραδοτέα περιλαμβάνεται το παρόν **pdf** αρχείο καθώς και το αντίστοιχο **.tex** και για κάθε μία άσκηση υπάρχει ένας υποφάκελος όπου περίεχει τα αντίστοιχα **.py** αρχεία με τον πηγαίο κώδικα.

2 Άσκηση 5

Στην **5η άσκηση** ζητείται να προγραμματιστεί μία συνάρτηση που να υπολογίζει το ημίτονο μιας οποιασδήποτε γωνίας. Για να δημιουργηθεί αυτή η συνάρτηση θα χρησιμοποιηθούν 10 τιμές του ημιτόνου με πλήρη ακρίβεια που υποστηρίζει το μηχάνημα που τρέχει ο κώδικας για την ανεξάρτητη μεταβλητή και 10 ψηφία ακρίβεια για την εξαρτημένη μεταβλητή. Η συνάρτηση του ημιτόνου θα προσεγγιστεί με τις παρακάτω μεθόδους :

- Με πολυωνυμική προσέγγιση
- Με την μέθοδο Splines
- Με την μέθοδο ελαχίστων τετραγώνων

Τα σημεία που θα χρησιμοποιηθούν είναι τα παρακάτω:

x	$sinx$
0.0	0.0
0.65	0.6051864057
1.3	0.9635581854
1.9500000000000002	0.9289597150
2.6	0.5155013718
3.25	-0.1081951345
3.9000000000000004	-0.6877661591
4.55	-0.9868438585
5.2	-0.8834546557
6.283185307179586	0

Τα σημεία επιλέχθηκαν στο διάστημα $[0, 2\pi]$ καθώς το ημίτονο είναι περιοδική συνάρτηση με περίοδο $T = 2\pi$ και σε διάστημα μήκους ίσο με την περίοδο η συνάρτηση περιέχει όλη την πληροφορία που χρειάζεται. Στην συνέχεια γίνεται σύγκριση των παραπάνω προσεγγίσεων στο διάστημα $[-\pi, \pi]$ όσον αφορά την ακρίβεια προσέγγισης και προβάλλεται σε διάγραμμα το σφάλμα προσέγγισης για 200 σημεία στο διάστημα $[-\pi, \pi]$ και αναφέρεται ο αριθμός των ψηφίων ακρίβειας που πετυχαίνονται.

2.1 Πολυωνυμική Προσέγγιση

Για την πολυωνυμική προσέγγιση χρησιμοποιείται η μέθοδος Newton. Για n σημεία εκπαίδευσης η μέθοδος αυτή παράγει ένα πολυώνυμο βαθμού $n-1$, το οποίο μάλιστα είναι και μοναδικό. Αρχικά υποθέτουμε ότι τα σημεία εκπαίδευσης προέρχονται από μία συνάρτηση $f(x)$, οπότε στόχος μας είναι να προσεγγίσουμε τα σημεία $(x_1, f(x_1)), ..., (x_n, f(x_n))$. Ορίζουμε ως

$$f[x_1 \ ... \ x_n]$$

τον συντελεστή του όρου x^{n-1} στο πολυώνυμο που προσεγγίζει τα παραπάνω σημεία.

Ορίζοντας λοιπόν τα παραπάνω η μέθοδος Newton ορίζει το ζητούμενο πολυώνυμο ως

$$P(x) = f[x_1] + f[x_1 \ x_2](x - x_1) + f[x_1 \ ... \ x_n](x - x_1)(x - x_{n-1})$$

Όπου τα $f[x_1 \ ... \ x_n]$ υπολογίζονται αναδρομικά ως εξής

$$f[x_k \ x_{k+1}] = \frac{f[x_{k+1}] - f[x_k]}{x_{k+1} - x_k}$$

Στο αρχείο **polynomial.py** υπάρχουν τρεις συναρτήσεις. Η συνάρτηση *polynomial_interpolation* δέχεται ως ορίσματα τα σημεία εκπαίδευσης και επιστρέφει τους συντελεστές του πολυωνύμου που προσεγγίζει αυτά τα σημεία. Οι συντελεστές του πολυωνύμου υπολογίζονται χρησιμοποιώντας το λεγόμενο τρίγωνο Newton, το οποίο μοιάζει όπως το *Σχήμα 1*.

$$
\begin{array}{r|rrrr}
-1 & -5 & & & \\
 & & 4 & & \\
0 & -1 & & -1 & \\
 & & 1 & & 1 \\
2 & 1 & & 3 & \\
 & & 10 & & \\
3 & 11 & & &
\end{array}
$$

Σχήμα 1: Τρίγωνο Newton για 4 σημεία εκπαίδευσης (Αριστερή στήλη η τιμή της ανεξάρτητης μεταβλητής των σημείων εκπαίδευσης)

Από το οποίο παίρνουμε τα στοιχεία της πρώτης διαγωνίου ως συντελεστές του πολυωνύμου με τον σταθερό όρο του πολυωνύμου να είναι ο πρώτος όρος αυτής της διαγωνίου.Στην συνέχεια, η συνάρτηση *calculate_polynomial* υπολογίζει και επιστρέφει την τιμή του πολυωνύμου στην τιμή x που δέχεται ως παράμετρο χρησιμοποιώντας τον βαθμό του πολυωνύμου, τις τιμές της ανεξάρτητης μεταβλητής των σημείων εκπαίδευσης και τους συντελεστές του πολυωνύμου που υπολογίσθηκαν προηγουμένως. Τέλος, η συνάρτηση *custom_sin* ορίζει αρχικά τα σημεία που αναφέρθηκαν στην παράγραφο της εισαγωγής της *Άσκησης 1* και στην συνέχεια χρησιμοποιεί τις άλλες δύο συναρτήσεις για να προσεγγίσει την τιμή του ημιτόνου στο σημείο που δέχεται ως παράμετρο, αφού πρώτα το μεταφέρει στο διάστημα $[0, 2\pi]$.

2.2 Προσέγγιση με Splines

Για την προσέγγιση με την μέθοδο Splines χρησιμοποιούνται φυσικές κυβικές Splines. Αρχικά, στην γενικότερη μορφή η μέθοδος Splines προσεγγίζει με μία καμπύλη (ευθεία, πολυώνυμο 2ου ή 3ου βαθμού) ανά δύο τα σημεία εκπαίδευσης, δηλαδή χρησιμοποιεί διαφορετικούς τύπους για να προσεγγίσει αυτά τα σημεία εκπαίδευσης. Ο όρος κυβικές που χρησιμοποιήθηκε παραπάνω αναφέρεται στο ότι η καμπύλη που θα προσεγγίσει ανά δύο τα σημεία εκπαίδευσης είναι 3ου βαθμού. Επομένως, αφού έχουμε ανάφερει τα παραπάνω καταλήγουμε στον παρακάτω ορισμό για την κυβική Spline για n σημεία εκπαίδευσης

$$
\begin{gathered}
S_1(x) = y_1 + b_1(x - x_1) + c_1(x - x_1)^2 + d_1(x - x_1)^3, \;\; x_1 \leq x \leq x_2 \\
S_2(x) = y_2 + b_2(x - x_2) + c_2(x - x_2)^2 + d_2(x - x_2)^3, \;\; x_2 \leq x \leq x_3 \\
\vdots \\
S_{n-1}(x) = y_{n-1} + b_{n-1}(x - x_{n-1}) + c_{n-1}(x - x_{n-1})^2 + d_{n-1}(x - x_{n-1})^3, \;\; x_{n-1} \leq x \leq x_n
\end{gathered}
$$

Με τις εξής ιδιότητες

- $S_i(x_i) = y_i$ και $S_i(x_{i+1}) = y_{i+1}$ για $i = 1, \ldots, n-1$
- $S'_{i-1}(x_i) = S'_i(x_i)$ για $i = 2, \ldots, n-1$
- $S''_{i-1}(x_i) = S''_i(x_i)$ για $i = 2, \ldots, n-1$

Χρησιμοποιώντας τα παραπάνω λαμβάνουμε $3n - 5$ εξισώσεις ενώ έχουμε $3n - 3$ αγνώστους. Επομένως, υπάρχουν άπειρες κυβικές Splines που προσεγγίζουν τα n σημεία εκπαίδευσης. Εδώ εισέρχεται ο όρος φυσικές κυβικές Splines που εισάγει τις δύο παρακάτω εξισώσεις

$$S_1''(x_1) = 0$$
$$S_{n-1}''(x_n) = 0$$

και επομένως πλέον έχουμε $3n - 3$ εξισώσεις με $3n - 3$ αγνώστους. Για να λύσουμε το σύστημα αρχικά ορίζουμε ως $c_n = S_{n-1}''(x_n)/2$, $\delta_i = x_{i+1} - x_i$ και $\Delta_i = y_{i+1} - y_i$. Τελικά, με τους παραπάνω συμβολισμούς οι συντελεστές υπολογίζονται ως εξής

$$d_i = \frac{c_{i+1} - ci}{3\delta_i}$$
$$b_i = \frac{\Delta_i}{\delta_i} - \frac{\delta_i}{3}(2c_i + c_{i+1})$$

για $i = 1, \ldots, n-1$ και χρησιμοποιώντας τις εξισώσεις των κυβικών Splines έχουμε ότι $2c_1 = 0$ και $2c_n = 0$. Οπότε, καταλήγουμε στο παρακάτω σύστημα προς επίλυση

$$\begin{pmatrix} 1 & 0 & 0 & & & \\ \delta_1 & 2\delta_1 + 2\delta_2 & \delta_2 & \ddots & & \\ 0 & \delta_2 & 2\delta_2 + 2\delta_3 & \delta_3 & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & & \delta_{n-2} & 2\delta_{n-2} + 2\delta_{n-1} & \delta_{n-1} \\ & & & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ \\ \\ \vdots \\ \\ c_n \end{pmatrix}$$

$$
= \begin{pmatrix} 0 \\ 3(\frac{\Delta_2}{\delta_2} - \frac{\Delta_1}{\delta_1}) \\ \\ \vdots \\ \\ 3(\frac{\Delta_{n-1}}{\delta_{n-1}} - \frac{\Delta_{n-2}}{\delta_{n-2}}) \\ 0 \end{pmatrix}
$$

Στο αρχείο **splines.py** υπάρχουν 4 συνάρτησεις. Η συνάρτηση *splines_interpolation* δέχεται ως ορίσματα τα σημεία εκπαίδευσης και επιστρέφει τους συντελεστές της φυσικής κυβικής Splines που προσεγγίζει τα σημεία αυτά, οι οποίοι υπολογίζονται σύμφωνα με αυτά που αναφέρθηκαν παραπάνω. Στην συνέχεια, η συνάρτηση *find_interval_index* επιστρέφει την θέση που θα έπρεπε να μπει το όρισμα search_value στον πίνακα array, ενώ η συνάρτηση *calculate_polynomial* υπολογίζει την τιμή της φυσικής κυβικής Spline στην θέση x που δέχεται σαν παράμετρο. Τέλος, η συνάρτηση *custom_sin* ορίζει αρχικά τα σημεία που αναφέρθηκαν στην παράγραφο της εισαγωγής της *Άσκησης 1* και στην συνέχεια χρησιμοποιεί τις άλλες τρεις συναρτήσεις για να προσεγγίσει την τιμή του ημιτόνου στο σημείο που δέχεται ως παράμετρο, αφού πρώτα το μεταφέρει στο διάστημα $[0, 2\pi]$.

2.3 Μέθοδος Ελαχίστων Τετραγώνων

Για την προσέγγιση με την μέθοδο των ελαχίστων τετραγώνων χρησιμοποιείται πολυώνυμο 10ου βαθμού. Η μέθοδος των ελαχίστων τετραγώνων προσεγγίζει ένα σύνολο σημείων εκπαίδευσης έτσι ώστε το άθροισμα των τετραγώνων του σφάλματος σε κάθε σημείο να είναι ελάχιστο. Για παράδειγμα αν έχουμε n σημεία εκπαίδευσης $(x_i, f_i))$ και θέλουμε να τα προσεγγίσουμε με ένα μοντέλο της μορφής $y = at + b$ πρέπει το

$$\sum_{k=1}^{n}(f_i - (at_i + b))^2$$

να γίνει ελάχιστο ως προς a και b. Το σύστημα που προκύπτει από την παραπάνω απαίτηση δεν έχει πάντα λύση και επομένως υπάρχει ανάγκη για την εισαγωγή του όρου της κοντινότερης λύσης προς την ιδανική (που δεν μπορεί να βρεθεί) λύση. Η κοντινότερη λύση ως προς την Ευκλείδια απόσταση μπορεί να βρέθει λύνοντας το παρακάτω σύστημα αντί το αρχικό $(Ax = b)$ που δεν έχει λύση

$$A^T A\overline{x} = A^T b$$

Οι οποίες ονομάζονται και κανονικές εξισώσεις. Τέλος, για να μετρήσουμε την απόδοση του μοντέλου μας υπολογίζουμε το σφάλμα ως εξής

$$r = b - A\overline{x}$$

Αν το διάνυσμα r είναι το μηδενικό τότε έχουμε λύσει το αρχικό σύστημα $Ax = b$ ακριβώς. Αν όχι, το Ευκλείδιο μέτρο του διανύσματος r είναι μία μετρική για το σφάλμα στην λύση $\overline{x}$. Υπάρχουν τουλάχιστον τρεις μετρικές για να εκφράσουμε το μέγεθος του σφάλματος.

- Το Ευκλείδιο μέτρο

$$\|r\| = \sqrt{r_1^2 + \cdots + r_m^2}$$

- Το τετραγωνικό σφάλμα (**Squared Error - SE**)

$$SE = r_1^2 + \cdots + r_m^2$$

- **Root Mean Squared Error - RMSE**

$$RMSE = \sqrt{SE/m} = \sqrt{(r_1^2 + \cdots + r_m^2)/m}$$

Από τα οποία μπορούμε να συμπεράνουμε ότι αν ελαχιστοποιήσουμε το ένα ελαχιστοποιούνται και τα υπόλοιπα.

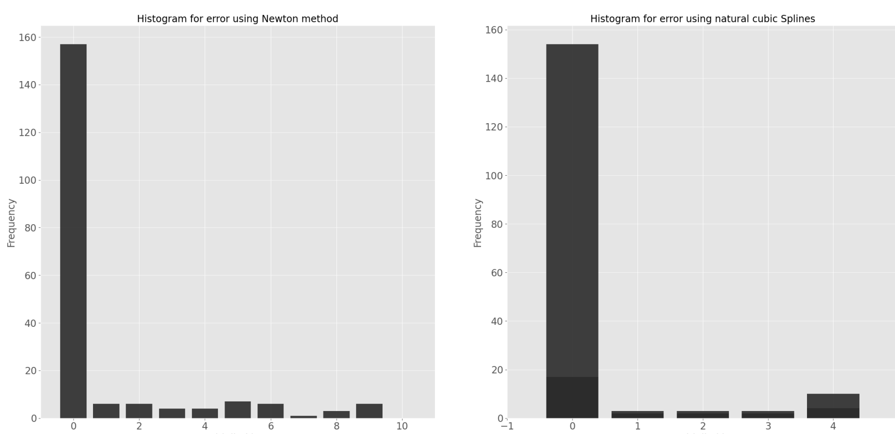
Στο αρχείο **leastSquares.py** υπάρχουν τρεις συναρτήσεις. Η συνάρτηση *least_squares_fit* δέχεται ως ορίσματα τα σημεία εκπαίδευσης και τον βαθμό του πολυωνύμου με το οποίο θα προσεγγιστούν τα σημεία εκπαίδευσης με την διαδικάσια που αναφέρθηκε παραπάνω, δηλαδή λύνοντας το σύστημα $A^T A\overline{x} = A^T b$ και επιστρέφει την λύση $\overline{x}$. Στην συνέχεια, η συνάρτηση *calculate_polynomial* υπολογίζει και επιστρέφει την τιμή του πολυωνύμου στην τιμή x που δέχεται σαν παράμετρο σε συνδυασμό με τον βαθμό του πολυωνύμου που έγινε η εκπαίδευση τον οποίο δέχεται κι αυτόν σαν παράμετρο. Τέλος, η συνάρτηση *custom_sin* ορίζει αρχικά τα σημεία που αναφέρθηκαν στην παράγραφο της εισαγωγής της *Άσκησης 1* και στην συνέχεια χρησιμοποιεί τις άλλες δύο συναρτήσεις για να προσεγγίσει την τιμή του ημιτόνου στο σημείο που δέχεται ως παράμετρο, αφού πρώτα το μεταφέρει στο διάστημα $[0, 2\pi]$. Το μοντέλο επιλέχθηκε να είναι πολυώνυμο 10ου βαθμού καθώς είναι αυτό που ελαχιστοποιεί το **RMSE** από όλα τα μοντέλα που δοκιμάστηκαν (πολυώνυμα βαθμού από 1 έως και 30).

2.4 Σύγκριση των προσεγγίσεων και ψηφία ακρίβειας

Για την σύγκριση των παραπάνω προσεγγίσεων υπολογίζουμε το **RMSE** και για τις τρεις μεθόδους για 200 σημεία στο $[-\pi, \pi]$ και καταλήγουμε στον παρακάτω πίνακα

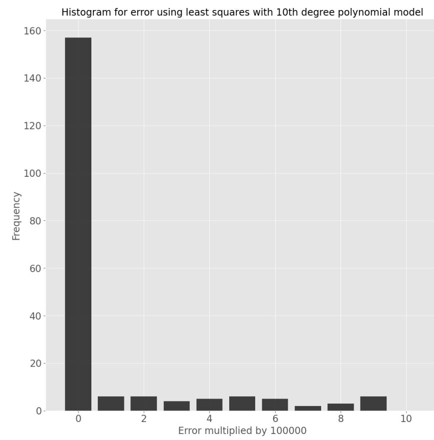
Μέθοδος Προσέγγισης	RMSE
Πολυωνυμική (Newton)	$2.644882474077062e-05$
Φυσικές Κυβικές Splines	0.0012736124616251868
Μέθοδος Ελαχίστων Τετραγώνων	$2.6267765538364208e-05$

Απ΄ όπου παρατηρούμε ότι το μικρότερο **RMSE** το πετυχαίνει η μέθοδος των ελαχίστων τετραγώνων με μοντέλο ένα πολυώνυμο 10ου βαθμού. Στην συνέχεια, την αμέσως καλύτερη επίδοση έχει η πολυωνιμκή προσέγγιση με την μέθοδο του Newton ενώ το μεγαλύτερο **RMSE** το πετυχαίνουν οι φυσικές κυβικές Splines.



Σχήμα 2: Ιστογράμματα συχνοτήτων του σφάλματος προσέγγισης για 200 σημεία στο [-π, π] για τις μεθόδους **Newton** και **Splines**.

Από τα ιστογράμματα των *Σχημάτων 2* και *3* παρατηρούμε ότι οι μέθοδοι Newton και ελαχίστων τετραγώνων προσεγγίζουν το ημίτονο με το σφάλμα να κυμαίνεται



Σχήμα 3: Ιστογράμματα συχνοτήτων του σφάλματος προσέγγισης για 200 σημεία στο [-π, π] για την μέθοδο των ελαχίστων τετραγώνων με πολυώνυμο 10ου βαθμού.

από -0.00009 εώς και 0.00009, ενώ οι φυσικές κυβικές Splines προσεγγίζουν το ημίτονο με το σφάλμα να κυμαίνεται από -0.004 έως 0.004. Το μέγιστο σφάλμα που κάνει η μέθοδος Newton είναι $9.727780525789487e-05$ ενώ το ελάχιστο $1.6041064032634722e-09$, επομένως η μέθοδος Newton έχει ακρίβεια από 3 έως 8 ψηφία. Αντίστοιχα, οι φυσικές κυβικές Splines έχουν μέγιστο σφάλμα ίσο με 0.004398121629302976 και ελάχιστο $2.547804256458619e-08$, επομένως η μέθοδος αυτή έχει ακρίβεια από 2 εώς 7 ψηφία. Τέλος, η μέθοδος ελαχίστων τετραγώνων έχει μέγιστο σφάλμα ίσο με $9.656790921613867e-05$ και ελάχιστο $5.535089053765319e-10$, επομένως η μέθοδος των ελαχίστων τετραγώνων έχει ακρίβεια από 3 έως 8 ψηφία. Τα παραπάνω αποτελέσματα φαίνονται συγκεντρωτικά και στον παρακάτω πίνακα

Μέθοδος Προσέγγισης	Ψηφία Ακρίβειας
Πολυωνυμική (Newton)	3 έως 8
Φυσικές Κυβικές Splines	2 έως 7
Μέθοδος Ελαχίστων Τετραγώνων	3 έως 8

3 Άσκηση 6

Στην **6η άσκηση** ζητείται να υπολογιστεί το ολοκλήρωμα της συνάρτησης του ημιτόνου στο διάστημα $[0, \pi/2]$ χρησιμοποιώντας 11 σημεία με τις μεθόδους **Simpson** και **τραπεζίου**. Ζητείται δηλαδή η τιμή του

$$\int_0^{\frac{\pi}{2}} sin(x)\, dx,$$

καθώς και να αναφερθεί το θεωρητικό και αριθμητικό σφάλμα για κάθε μέθοδο αντίστοιχα. Τα 11 σημεία που επιλέχθηκαν είναι τα παρακάτω

0.0
0.15707963
0.31415927
0.4712389
0.62831853
0.78539816
0.9424778
1.09955743
1.25663706
1.41371669
1.57079633

Αν υπολογίσουμε το ολοκλήρωμα χρησιμοποιώντας γνώσεις του ολοκληρωτικού λογισμού έχουμε διαδοχικά

$$\int_0^{\frac{\pi}{2}} sin(x)\, dx = -cosx\Big|_0^{\frac{\pi}{2}} = 1$$

3.1 Μέθοδος Simpson

Η μέθοδος Simpson προσεγγίζει την τιμή ενός ολοκληρώματος $\int_a^b f(x)\,dx$ μιας συνεχούς συνάρτησης σε ένα κλειστό και φραγμένο διάστημα [a, b] με χρήση εμβαδών παραβολών, οι οποίες προκύπτουν από την προσέγγιση της συνάρτησης f σε στοιχειώδη υποδιαστήματα του [a, b] από πολυώνυμα 2ου βαθμού, δηλαδή από παραβολές. Η μέθοδος περιγράφεται παρακάτω

- Έστω $\{x_0 = a, \ldots, x_N = b\}$, με $x_0 < x_1 < \cdots < x_N$ ένας ομοιόμορφος διαμερισμός του [a, b], δηλαδή χωρίζουμε το [a, b] σε N ισομήκη υποδιαστήματα.
- Τότε $x_i = x_0 + \varkappa\frac{b-a}{N}$, $\varkappa = 0, \ldots, N$
- Υπολογίζουμε τις τιμές $f(x_i), i = 0, \ldots, N$
- Σχηματίζουμε τις διαδοχικές παραβολές που διέρχονται από τα σημεία $f(x_i)$, $f(x_{i+1}), f(x_{i+2})$, $i = 0, \ldots, N$, οπότε πρέπει N **ζυγός**.
- Υπολογίζουμε τα εμβαδά των N/2 παραβολών που σχηματίζονται και μετά από πράξεις έχουμε τελικά

$$\int_a^b f(x)\,dx \approx \frac{b-a}{3N}(f(x_0) + f(x_N) + 2\sum_{i=1}^{\frac{N}{2}-1} f(x_{2i}) + 4\sum_{i=1}^{\frac{N}{2}} f(x_{2i-1}))$$

ενώ το άνω όριο του σφάλματος είναι

$$|e| \leq \frac{(b-a)^5}{180N^4}M,$$

όπου

$$M = max\{|f^{(4)}(x)| : x \in [\text{a, b}]\}$$

Στο αρχείο **simpson.py** έχουν αναπτυχθεί δύο συναρτήσεις, μία για τον υπολογισμό ενός ολοκληρώματος από το a στο b για μία συνεχή συνάρτηση f

με την μέθοδο Simpson, καθώς και μία για τον υπολογισμό του άνω ορίου του σφάλματος για τις προαναφερόμενες παραμέτρους. Αν καλέσουμε την συνάρτηση ***simpson_ integrate*** με ορίσματα την συνάρτηση του ημιτόνου, το 0, το $\frac{\pi}{2}$, το 10 και τα σημεία που αναφέρθηκαν στην εισαγωγή αυτής της άσκησης, καθώς και την συνάρτηση **simpson_ error_ bound** με τα ίδια ορίσματα εκτός της λίστας των σημείων λαμβάνουμε τα παρακάτω

```
Simpson integration result = 1.0000033922209004
Simpson integration error = 3.3922209004000337e-06
Simpson theoretical error bound = 5.312841749744469e-06
```

Σχήμα 4: Αποτελέσματα κλήσης της συνάρτησης ***simpson_ integrate*** και ***simpson_ error_ bound*** στο διάστημα $[0, \frac{\pi}{2}]$.

Απ΄ όπου παρατηρούμε ότι το σφάλμα προσέγγισης θεωρητικά είναι $5.312841749744469e-06$ ενώ αριθμητικά (δηλαδή αφαιρώντας από την πραγματική τιμή του ολοκληρώματος, που είναι 1, και εφαρμόζοντας απόλυτη τιμή) είναι $3.3922209004000337e-06$ και τέλος, η προσεγγιστική τιμή του ολοκληρώματος με την μέθοδο **Simpson** είναι 1.0000033922209004.

3.2 Μέθοδος τραπεζίου

Η μέθοδος τραπεζίου προσεγγίζει την τιμή ενός ολοκληρώματος $\int_a^b f(x)\,dx$ μιας συνεχούς συνάρτησης σε ένα κλειστό και φραγμένο διάστημα [a, b] με χρήση εμβαδών τραπεζίων που προκύπτουν από την προσέγγιση της συνάρτησης f από μία τεθλασμένη γραμμή. Η μέθοδος περιγράφεται παρακάτω

- Έστω $\{x_0 = a, \ldots, x_N = b\}$, με $x_0 < x_1 < \cdots < x_N$ ένας ομοιόμορφος διαμερισμός του [a, b], δηλαδή χωρίζουμε το [a, b] σε N ισομήκη υποδιαστήματα.
- Τότε $x_i = x_0 + \kappa\frac{b-a}{N}$, $\kappa = 0, \ldots, N$

- Υπολογίζουμε τις τιμές $f(x_i), i = 0, \ldots, N$
- Σχηματίζουμε τα διαδοχικά ευθύγραμμα τμήματα με άκρα τα $f(x_0), \ldots, f(x_N)$ οπότε σχηματίζεται μία τεθλασμένη γραμμή.
- Υπολογίζουμε τα εμβαδά των N τραπεζίων που σχηματίζονται και μετά από πράξεις έχουμε τελικά

$$\int_a^b f(x)\,dx \approx \frac{b-a}{2N}(f(x_0) + f(x_N) + 2\sum_{i=1}^{N-1} f(x_i))$$

ενώ το άνω όριο του σφάλματος είναι

$$|e| \leq \frac{(b-a)^3}{12N^2}M,$$

όπου

$$M = max\{|f''(x)| : x \in [a, b]\}$$

Στο αρχείο **trapezoid.py** έχουν αναπτυχθεί δύο συναρτήσεις, μία για τον υπολογισμό ενός ολοκληρώματος από το a στο b για μία συνεχή συνάρτηση f με την μέθοδο τραπεζίου, καθώς και μία για τον υπολογισμό του άνω ορίου του σφάλματος για τις προαναφερόμενες παραμέτρους. Αν καλέσουμε την συνάρτηση ***trapezoid_ integrate*** με ορίσματα την συνάρτηση του ημιτόνου, το 0, το $\frac{\pi}{2}$, το 10 και τα σημεία που αναφέρθηκαν στην εισαγωγή αυτής της άσκησης, καθώς και την συνάρτηση **trapezoid_ error_ bound** με τα ίδια ορίσματα εκτός της λίστας των σημείων λαμβάνουμε τα παρακάτω

```
Trapezoid integration result = 0.9979429863543572
Trapezoid integration error = 0.0020570136456428134
Trapezoid theoretical error bound = 0.0032298204875312307
```

Σχήμα 5: Αποτελέσματα κλήσης της συνάρτησης ***trapezoid_ integrate*** και ***trapezoid_ error_ bound*** στο διάστημα $[0, \frac{\pi}{2}]$.

Απ΄ όπου παρατηρούμε ότι το σφάλμα προσέγγισης θεωρητικά είναι 0.0032298204875312307 ενώ αριθμητικά (δηλαδή αφαιρώντας από την πραγματική τιμή του ολοκληρώματος, που είναι 1, και εφαρμόζοντας απόλυτη τιμή) είναι 0.0020570136456428134 και τέλος, η προσεγγιστική τιμή του ολοκληρώματος με την μέθοδο τραπεζίου είναι 0.9979429863543572.

4 Άσκηση 7

Στην **7η άσκηση** ζητείται να εκτιμηθεί η τιμή κλεισίματος ημέρας δύο διαφορετικών μετοχών εταιρειών του Χρηματιστηρίου Αθηνών για την ημέρα που είναι κοντινότερη στις **17 Μαρτίου 2020**. Οι εταιρείες που επιλέχθηκαν είναι η **AUTOHELLAS A.E. (ΟΤΟΕΛ)** και τα **Ελληνικά Πετρέλαια (ΕΛΠΕ)**. Η εκτίμηση θα γίνει για την τιμή κλεισίματος αυτών των μετοχών στις 18 και 24 Μαρτίου 2020 χρησιμοποιοώντας 10 τιμές κλείσιματος από τις συνεδριάσεις που έγιναν από τις 4 έως και τις 17 Μαρτίου 2020, ενώ τα μοντέλα που θα χρησιμοποιηθούν είναι πολυώνυμα 2ου,3ου και 4ου βαθμού χρησιμοποιώντας την μέθοδο των ελαχίστων τετραγώνων. Για την αξιολόγηση των μοντέλων θα χρησιμοποιηθεί η μετρική **Root Mean Squared Error - RMSE**, ενώ εκτός από τις 10 τιμές για την εκπαίδευση των μοντέλων θα χρησιμοποιηθούν και επιπλέον 5 τιμές για την αξιολόγηση των μοντέλων. Έτσι, συνολικά έχουμε 15 τιμές κλεισίματος από τις οποίες οι 10 ($\approx$ **70%** του συνολικού συνόλου δεδομένων) θα χρησιμοποιηθούν για την εκπαίδευση **(train set)**, ενώ οι 5 ($\approx$ **30%** του συνολικού συνόλου δεδομένων) θα χρησιμοποιηθούν για την αξιολόγηση των μοντέλων **(test set)**. Τέλος, τονίζεται ότι στα διαγράμματα αυτής της άσκησης η αριθμήση ξεκινάει στους άξονες από τις 4 Μαρτίου 2020, δηλαδή το $x = 0$ αναφέρεται στις 4 Μαρτίου 2020 ενώ στην συνέχεια κάθε αύξηση του x κατά 1 αναφέρεται στην επόμενη μέρα που είναι ανοιχτό το χρηματιστήριο.

4.1 Εκτίμηση της τιμής κλεισίματος της μετοχής της εταιρείας AUTOHELLAS

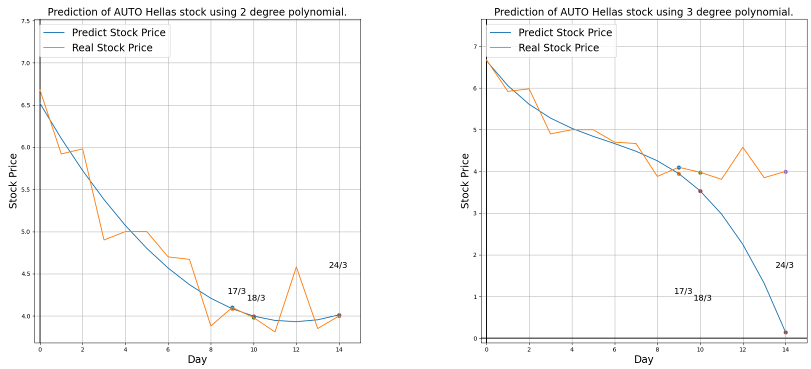
Στον πίνακα του *Σχήματος 4* παρουσιάζονται οι τιμές κλείσιματος της μετοχής της **AUTOHELLAS A.E.** για την περίοδο από τις 4 έως τις 24 Μαρτίου 2020.

Ημερομηνία	Τιμή Κλείσιματος Μετοχής
4 Μαρτίου 2020	6.68
5 Μαρτίου 2020	5.92
6 Μαρτίου 2020	5.98
9 Μαρτίου 2020	4.9
10 Μαρτίου 2020	5.0
11 Μαρτίου 2020	5.0
12 Μαρτίου 2020	4.7
13 Μαρτίου 2020	4.67
16 Μαρτίου 2020	3.88
17 Μαρτίου 2020	4.1
18 Μαρτίου 2020	3.98
19 Μαρτίου 2020	3.81
20 Μαρτίου 2020	4.58
23 Μαρτίου 2020	3.85
24 Μαρτίου 2020	4.0

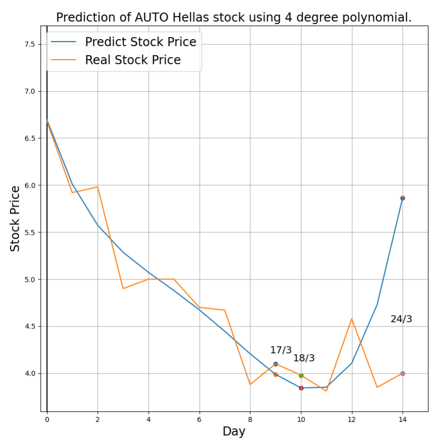
Σχήμα 6: Τιμές κλεισίματος της μετοχής της εταιρείας **AUTOHELLAS A.E.**

Στο αρχείο **auto_hellas_stock.py** χρησιμοποιώντας τις συναρτήσεις που αναπτύχθηκαν στην **Άσκηση 5** στην παράγραφο με την μέθοδο των ελαχίστων τετραγώνων και τις τιμές κλεισίματος από τις 4 έως και τις 17 Μάρτιου 2020 καταλήγουμε στα διαγράμματα των *σχημάτων 5 και 6.*

Πιο συγκεκριμένα, το πολυώνυμο 2ου βαθμού για την πρόβλεψη της τιμής κλεισίματος στις **18 Μαρτίου 2020** επιστρέφει την τιμή 3.9971666666666508 ενώ η πραγματική τιμή κλεισίματος για εκείνη την ημέρα ήταν 3.98. Για την πρόβλεψη 5 ημέρες μετά, δηλαδή στις **24 Μαρτίου 2020**, το μοντέλο επιστρέφει την τιμή 4.012469696969636 ενώ η πραγματική τιμή κλεισίματος ήταν 4.0. Για το **test set** το πολυώνυμο 2ου βαθμού πέτυχε **rmse = 0.3000386931735489**, ενώ και



Σχήμα 7: Διάγραμμα της πρόβλεψης καθώς και της πραγματικής τιμής κλεισίματος της μετόχης της εταιρείας **AUTOHELLAS A.E.** για τα μοντέλα με βαθμό 2 και 3 για την περίοδο 4 με 24 Μαρτίου 2020.



Σχήμα 8: Διάγραμμα της πρόβλεψης καθώς και της πραγματικής τιμής κλεισίματος της μετόχης της εταιρείας **AUTOHELLAS A.E.** για το μοντέλο με βαθμό 4 για την περίοδο 4 με 24 Μαρτίου 2020.

για τις δύο προβλέψεις παρατηρούμε ότι είναι αρκετά κοντά στις πραγματικές τιμές.

Στην συνέχεια, το πολυώνυμο 3ου βαθμού για την πρόβλεψη της τιμής κλεισιματος στις **18 Μαρτίου 2020** επιστρέφει την τιμή 3.5346666666667828 ενώ η πραγματική τιμή κλεισίματος για εκείνη την ημέρα ήταν 3.98. Για την πρόβλεψη 5 ημέρες μετά, δηλαδή στις **24 Μαρτίου 2020**, το μοντέλο επιστρέφει την τιμή 0.14105361305466246 ενώ η πραγματική τιμή κλεισίματος ήταν 4.0. Για το **test set** το πολυώνυμο 3ου βαθμού πέτυχε **rmse = 2.351096182011764**, ενώ για την πρώτη πρόβλεψη παρατηρούμε μία μικρή απόκλιση, η οποία γίνεται αρκετά μεγαλύτερη στις 24 Μαρτίου.

Τέλος, το πολυώνυμο 4ου βαθμού για την πρόβλεψη της τιμής κλεισιματος στις **18 Μαρτίου 2020** επιστρέφει την τιμή 3.8441666666660677 ενώ η πραγματική τιμή κλεισίματος για εκείνη την ημέρα ήταν 3.98. Για την πρόβλεψη 5 ημέρες μετά, δηλαδή στις **24 Μαρτίου 2020**, το μοντέλο επιστρέφει την τιμή 5.861392773880528 ενώ η πραγματική τιμή κλεισίματος ήταν 4.0. Για το **test set** το πολυώνυμο 4ου βαθμού πέτυχε **rmse = 0.9468997720978967**, ενώ για τις προβλέψεις παρατηρούμε ότι η πρώτη είναι σχετικά κοντά στην πραγματική τιμή, με την δεύτερη όμως να αποκλίνει όπως και στο πολυώνυμο 3ου βαθμού.

Συμπερασματικά, τα 3 μοντέλα καταντάσσονται με βάση το **RMSE** στον παρακάτω πίνακα

Βαθμός Πολυωνύμου	RMSE
2	0.3000386931735489
4	0.9468997720978967
3	2.351096182011764

ενώ σχετικά με τα διαγράμματα των *σχήματων 5 και 6* παρατηρούμε ότι όσο ανεβαίνει ο βαθμός του πολυωνύμου έχουμε καλύτερη προσέγγιση στο κέντρο του διαστήματος των σημείων εκπαίδευσης αλλά μεγαλύτερο σφάλμα όσο προχωράμε προς τα άκρα αυτού του διαστήματος, στο οποίο ανήκουν και οι ημερομηνίες για τις οποίες γίνεται η πρόβλεψη της τιμής κλεισίματος με αποτελέσμα να υπάρχει μεγαλύτερο σφάλμα από το 2ου βάθμου λόγω της κυρτότητας των πολυωνύμων 3ου και 4ου βαθμού.

4.2 Εκτίμηση της τιμής κλεισίματος της μετοχής της εταιρείας ΕΛΛΗΝΙΚΑ ΠΕΤΡΕΛΑΙΑ

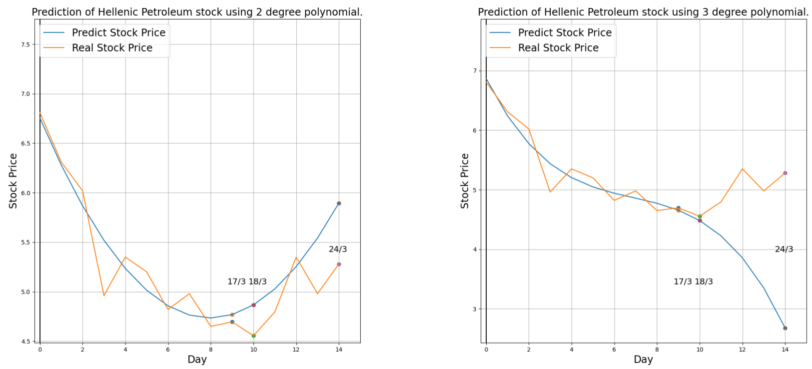
Στον πίνακα του *Σχήματος 7* παρουσιάζονται οι τιμές κλείσιματος της μετοχής της **ΕΛΛΗΝΙΚΑ ΠΕΤΡΕΛΑΙΑ** για την περίοδο από τις 4 έως τις 24 Μαρτίου 2020.

Ημερομηνία	Τιμή Κλείσιματος Μετοχής
4 Μαρτίου 2020	6.81
5 Μαρτίου 2020	6.31
6 Μαρτίου 2020	6.02
9 Μαρτίου 2020	4.96
10 Μαρτίου 2020	5.35
11 Μαρτίου 2020	5.2
12 Μαρτίου 2020	4.82
13 Μαρτίου 2020	4.98
16 Μαρτίου 2020	4.65
17 Μαρτίου 2020	4.695
18 Μαρτίου 2020	4.555
19 Μαρτίου 2020	4.795
20 Μαρτίου 2020	5.35
23 Μαρτίου 2020	4.98
24 Μαρτίου 2020	5.28

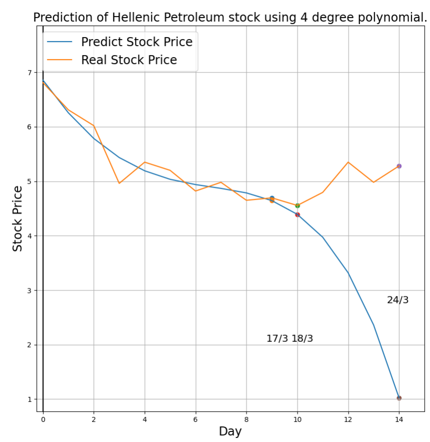
Σχήμα 9: Τιμές κλεισίματος της μετοχής της εταιρείας **ΕΛΛΗΝΙΚΑ ΠΕΤΡΕΛΑΙΑ**

Στο αρχείο **hellenic_petroleum_stock.py** χρησιμοποιώντας τις συναρτήσεις που αναπτύχθηκαν στην **Άσκηση 5** στην παράγραφο με την μέθοδο των ελαχίστων τετραγώνων και τις τιμές κλεισίματος από τις 4 έως και τις 17 Μάρτιου 2020 καταλήγουμε στα διαγράμματα των *σχημάτων 8 και 9.*

Πιο συγκεκριμένα, το πολυώνυμο 2ου βαθμού για την πρόβλεψη της τιμής κλεισίματος στις **18 Μαρτίου 2020** επιστρέφει την τιμή 4.866166666666659 ενώ η πραγματική τιμή κλεισίματος για εκείνη την ημέρα ήταν 4.555. Για την πρόβλεψη 5 ημέρες μετά, δηλαδή στις **24 Μαρτίου 2020**, το μοντέλο επιστρέφει



Σχήμα 10: Διάγραμμα της πρόβλεψης καθώς και της πραγματικής τιμής κλεισίματος της μετόχης της εταιρείας **ΕΛΛΗΝΙΚΑ ΠΕΤΡΕΛΑΙΑ** για τα μοντέλα με βαθμό 2 και 3 για την περίοδο 4 με 24 Μαρτίου 2020.



Σχήμα 11: Διάγραμμα της πρόβλεψης καθώς και της πραγματικής τιμής κλεισίματος της μετόχης της εταιρείας **ΕΛΛΗΝΙΚΑ ΠΕΤΡΕΛΑΙΑ** για το μοντέλο με βαθμό 4 για την περίοδο 4 με 24 Μαρτίου 2020.

την τιμή 5.894499999999967 ενώ η πραγματική τιμή κλεισίματος ήταν 5.28. Για το **test set** το πολυώνυμο 2ου βαθμού πέτυχε **rmse = 0.41320119494294816**, ενώ και για τις δύο προβλέψεις παρατηρούμε σχετικά καλές εκτιμήσεις με την δεύτερη να έχει μεγαλύτερη απόκλιση από την πρώτη.

Πιο συγκεκριμένα, το πολυώνυμο 3ου βαθμού για την πρόβλεψη της τιμής κλεισίματος στις **18 Μαρτίου 2020** επιστρέφει την τιμή 4.482000000000168 ενώ η πραγματική τιμή κλεισίματος για εκείνη την ημέρα ήταν 4.555. Για την πρόβλεψη 5 ημέρες μετά, δηλαδή στις **24 Μαρτίου 2020**, το μοντέλο επιστρέφει την τιμή 2.6787832167846926 ενώ η πραγματική τιμή κλεισίματος ήταν 5.28. Για το **test set** το πολυώνυμο 3ου βαθμού πέτυχε **rmse = 1.5483780363486506**, ενώ παρατηρούμε για την πρώτη πρόβλεψη ότι είναι καλύτερη σε σχέση με την πρόβλεψη του πολυωνύμου 2ου βαθμού αλλά με την πρόβλεψη στις 24 Μαρτίου να έχει αρκετά μεγαλύτερη απόκλιση από την πραγματική τιμή κλεισίματος.

Τέλος, το πολυώνυμο 4ου βαθμού για την πρόβλεψη της τιμής κλεισιματος στις **18 Μαρτίου 2020** επιστρέφει την τιμή 4.392499999999325 ενώ η πραγματική τιμή κλεισίματος για εκείνη την ημέρα ήταν 4.555. Για την πρόβλεψη 5 ημέρες μετά, δηλαδή στις **24 Μαρτίου 2020**, το μοντέλο επιστρέφει την τιμή 1.0245979020847766 ενώ η πραγματική τιμή κλεισίματος ήταν 5.28. Για το **test set** το πολυώνυμο 4ου βαθμού πέτυχε **rmse = 2.4409310013590737**, ενώ για τις προβλέψεις παρατηρούμε ότι η πρώτη εκτίμηση είναι σχετικά κοντά στην πραγματική τιμή, με την δεύτερη όμως να αποκλίνει με μεγαλύτερο σφάλμα απ΄ ότι στο πολυώνυμο 3ου βαθμού.

Συμπερασματικά, τα 3 μοντέλα κατατάσσονται με βάση το **RMSE** στον παρακάτω πίνακα

Βαθμός Πολυωνύμου	RMSE
2	0.41320119494294816
3	1.5483780363486506
4	2.4409310013590737

ενώ σχετικά με τα διαγράμματα των *σχήματων 8 και 9* παρατηρούμε ότι όσο ανεβαίνει ο βαθμός του πολυωνύμου έχουμε σχετικά καλύτερη προσέγγιση στο κέντρο του διαστήματος των σημείων εκπαίδευσης αλλά μεγαλύτερο σφάλμα όσο προχωράμε προς τα άκρα αυτού του διαστήματος, στο οποίο ανήκουν και οι ημερομηνίες για τις οποίες γίνεται η πρόβλεψη της τιμής κλεισίματος με αποτελέσμα να υπάρχει μεγαλύτερο σφάλμα σε σχέση με το πολυώνυμο 2ου βάθμου και πάλι λόγω της κυρτότητας των πολυωνύμων 3ου και 4ου βαθμού.