Δεύτερη Εργασία Αριθμητικής Ανάλυσης

Ευάγγελος Χαλκιάς

23/12/2020

Προσέγγιση συνάρτησης με διάφορες μεθόδους.

Για αυτή την άσκηση ζητείται η προσέγγιση της συνάρτησης του ημιτόνου χρησιμοποιώντας 10 τιμές, που στην συγκεκριμένη υλοποίηση ανήκουν στο διάστημα $[-\pi, \pi]$ και έχουν ακρίβεια 5 δεκαδικών ψηφίων. Με χρήση διαφόρων μεθόδων, θα μπορούμε να υπολογίζουμε μια σχετικά ακριβής προσεγγιστική τιμή για κάθε γωνία στο διάστημα αυτό.

Προσεγγιστική μέθοδος πολυωνύμου Lagrange.

Η μέθοδος πολυωνύμου Lagrange χρησιμοποιείται για την προσέγγιση συναρτήσεων με βάση κάποια δεδομένα σημεία (x, y). Έστω για $n+1$ σημεία, με τάξη πολυωνύμου n επιτυγχάνουμε η τελική συνάρτηση να περνάει από κάθε σημείο x και έτσι να έχουμε μία πολύ καλή προσέγγιση. Το πολυώνυμο Lagrange έχει την μορφή:

$$L_i(x) = \frac{(x-x_0)(x-x_1)...(x-x_{i-1})(x-x_{i+1})...(x-x_n)}{(x_i-x_0)(x_i-x_1)...(x_i-x_{i-1})(x_i-x_{i+1})...(x_i-x_n)}, i = 0, ..., n$$

Αθροίζοντας τα αποτελέσματα παίρνουμε:

$$P_n(x) = \sum_{i=0}^{n} y_i L_i(x)$$

Το κάθε πολυώνυμο είναι βαθμού n και έτσι το αθροισμά τους είναι πάλι βαθμού n.

Η συνάρτηση Lagrange Polynomial

Η συνάρτηση δέχεται ως ορίσματα έναν πίνακα από tuples όπου το πρώτο στοιχείο είναι το x και το δεύτερο το y. Το δεύτερο όρισμα είναι το x για το οποίο θέλουμε να προσεγγίσουμε το y. Μέσα στην συνάρτηση, πρώτα από όλα οριστικοποιείται το y και στην συνέχεια εκτελείται διπλός βρόγχος για τον υπολογισμό των $L_i(x), i = 0, ..., n$. Χρησιμοποιούνται οι βοηθητικές μεταβλητές sum1 και sum2 για τον υπολογισμό του αριθμητή και παρανομαστή του κάθε συντελεστή Lagrange αντίστοιχα. Στο τέλος της κάθε επανάληψης διαιρείται το sum1 και το sum2 και πολλαπλασιάζεται με το αντίστοιχο y_i. Τέλος, με έναν βρόγχο, αθροίζονται τα πολυώνυμα και αποθηκεύεται το αποτέλεσμα στην μεταβλητή y η οποία και επιστρέφεται.

Η μέθοδος Natural Cubic Splines

Η μέθοδος Natural Cubic Splines είναι ένας τρόπος προσέγγισης μιας συνάρτησης στο διάστημα $[a, b]$ με $a = x_0 < x_1 < ... < x_n = b$ χωρίζοντας την σε τμηματικά πολυώνυμα τρίτου βαθμού. Κάθε πολυώνυμο ορίζεται σε ένα διάστημα $[x_i, x_{i+1}]$ όπου είναι δύο φορές παραγωγίσιμο και έχει μορφή:

$$s^{(i)}(x) = a^{(i)} + b^{(i)}x + c^{(i)}x^2 + d^{(i)}x^3, i = 0, ..., n$$

Άρα, συνολικά πρέπει να υπολογισθούν $4n$ σταθερές. Έχοντας $n + 1$ σχέσεις απο τις συνθήκες παρεμβολής, $n - 1$ σχέσεις απο τις συνθήκες συνέχειας και $2n - 2$ σχέσεις από τις συνθήκες παραγωγισιμότητας, καταλήγουμε με $4n - 2$ συνθήκες που δεν είναι αρκετές για τον υπολογισμό του συστήματος. Έτσι, ορίζουμε εμείς $s''(x_0) = 0, s''(x_n) = 0$ για να έχει λύση το σύστημα. Από εκεί προέρχεται και το Natural στο Natural Cubic Splines.

Η συνάρτηση Natural Cubic Splines Η συνάρτηση δέχεται ως ορίσματα έναν πίνακα απο tuples με τις συντεταγμένες x και y και ένα όρισμα x του οποίου το y θέλουμε να υπολογίσουμε. Πρώτα, μέσα στην συνάρτηση αρχικοποιούνται 2 πίνακες w και h όπου κάθε στοιχείο του w είναι ίσο με $w[i] = x_{i+1} - x_i$ και του h με $h[i] = \frac{y_{i+1} - y_i}{w[i]}, i \in [0, n-1]$. Οι υπολογισμοί αυτοί εκτελούνται στον βρόγχο που ακολουθεί. Στην συνέχεια, δημιουργείται μία νέα λίστα για την αποθήκευση των τιμών της δευτέρας παραγώγου της συνάρτησης με όνομα fder2 . Ο τύπος υπολογισμού των σημείων είναι ο εξής:

$$fder2[i + 1] = 3\frac{h_{i+1} - h_i}{w_{i+1} + w_i}, i \in [0, n - 2]$$

Διαλέγοντας τα i κατάλληλα αποφεύγουμε την αλλαγη των δευτέρων παραγώγων που εμείς ορίσαμε ίσον με το 0. Ετοιμάζοντας τις βοηθητικές λίστες, μπορούμε να υπολογίσουμε τους πίνακες A,B,C και D . Οι λίστες αυτες έχουν n μέγεθος και αντιπροσωπεύουν τις $4n$ σταθερές που θέλουμε να υπολογίσουμε. Οι υπολογιστικοί τύποι για κάθε λίστα είναι οι εξής:

Για $i \in [0, n-1]$

$$A[i] = \frac{fder2[i+1] - fder2[i]}{6w[i]}$$

$$B[i] = \frac{fder2[i]}{2}$$

$$C[i] = h[i] - w[i]\frac{fder2[i+1] + 2fder2[i]}{6}$$

$D[i] = y[i]$ όπου y οι τιμές που δώσαμε ως όρισμα.

Στην συνέχεια, με βάση τα σημεία x που έχουν διαλεχτεί, γίνεται έλεγχος με συνθήκες if για το ποιο πολυώνυμο πρέπει να χρησιμοποιηθεί. Η συνάρτηση επιστρέφει το αποτέλεσμα της προσέγγισης.

Η μέθοδος Least Squares Η μέθοδος ελαχίστων τετραγώνων είναι άλλο ένα εργαλείο για την προσέγγιση συναρτήσεων με πολυώνυμα διαφόρων δυνάμεων. Η δουλειά της μεθόδου είναι να βρει την καλύτερη δυνατή καμπύλη για την προσέγγιση κάποιων δεδομένων σημείων (x_i, y_i). Στην συγκεκριμένη περίπτωση το πολυώνυμο που ψάχνει είναι βαθμού 5 και έτσι έχει να υπολογίσει:

$$E(a_0, a_1, a_2, a_3, a_4, a_5) = \sum_{i=1}^{n}[y[i] - (a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5)]^2 = min$$

Αυτό επιτυγχάνεται υπολογίζοντας τις μερικές παραγώγους κάθε a και βάζοντας τες ίσες με το 0. Έτσι, δημιουργείται ένα σύστημα $6x6$ που υπολογίζεται ως εξής:

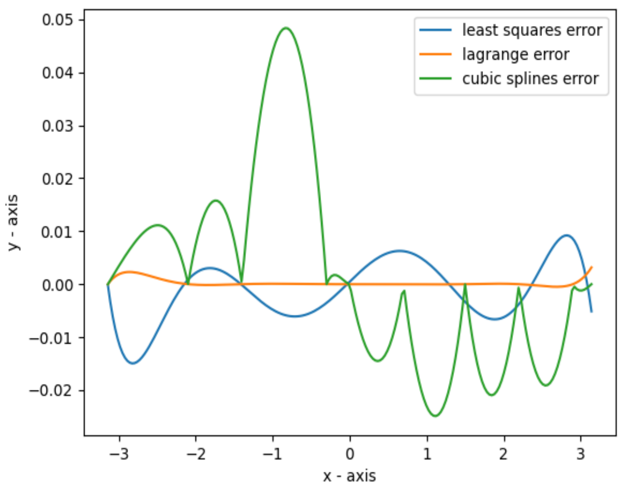
$$y = Xa$$
$$X^T y = X^T X a$$
$$a = (X^T X)^{-1} X^T y$$

Η συνάρτηση Least Squares Όπως οι άλλες συναρτήσεις, και αυτή, δέχεται ως ορίσματα τις συντεταγμένες (x, y) και το x για το οποίο θέλουμε να βρούμε μία προσεγγιστική τιμή y. Αρχικά δημιουργείται μια λίστα στην οποία τοποθετούνται όλες οι τιμές y. Στην συνέχεια αρχικοποιείται ένας πίνακας $10x6$ όπου το 10 αντιστοιχεί στον αριθμό των δεδομένων σημείων και το 6 στους άγνωστους συντελεστές. Αμέσως μετά, με έναν εμφωλευμένο βρόγχο γίνεται ο υπολογισμός του κάθε στοιχείου του πίνακα αυτού. Ο τρόπος υπολογισμού είναι ο εξής:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 & x_1^5 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 & x_2^5 \\ 1 & x_3 & x_3^2 & x_3^3 & x_3^4 & x_3^5 \\ 1 & x_4 & x_4^2 & x_4^3 & x_4^4 & x_4^5 \\ 1 & x_5 & x_5^2 & x_5^3 & x_5^4 & x_5^5 \\ 1 & x_6 & x_6^2 & x_6^3 & x_6^4 & x_6^5 \\ 1 & x_7 & x_7^2 & x_7^3 & x_7^4 & x_7^5 \\ 1 & x_8 & x_8^2 & x_8^3 & x_8^4 & x_8^5 \\ 1 & x_9 & x_9^2 & x_9^3 & x_9^4 & x_9^5 \\ 1 & x_{10} & x_{10}^2 & x_{10}^3 & x_{10}^4 & x_{10}^5 \end{pmatrix}$$

Τέλος, βηματικά, υπολογίζει $a = (X^T X)^{-1} X^T y$ και επιστρέφει το ζητούμενο αποτέλεσμα.

Αν πάρουμε τις 3 συναρτήσεις και υπολογίσουμε τα σφάλματα που παράγουν οι διαφορετικοί μέθοδοι στο $[-\pi, \pi]$ με βάση 200 σημεία, παρατηρούμε ότι την μικρότερη απόκλιση παρουσιάζει η μέθοδος Lagrange, ακολουθεί η μέθοδος ελαχίστων τετραγώνων και τελευταία η μέθοδος των Natural Bicubic Splines.



Υπολογίζοντας το μέσο σφάλμα και για τις 3 μεθόδους για 200 σημεία στο $x \in [-\pi, \pi]$ λαμβάνουμε:

Μέσο Σφάλμα	
Lagrange	0.0003241619316334156
Least Squares	0.004740284068199873
Natural Bicubic Splines	0.014262364092823278

Με βάση αυτά τα στοιχεία συμεραίνουμε ότι $\frac{1}{2}10^{-2} > 0.004740284068199873$ άρα για Least Squares έχουμε κατά προσέγγιση 2 ψηφία ακρίβειας.

Για Lagrange έχουμε $\frac{1}{2}10^{-3} > 0.0003241619316334156$ και έτσι υπάρχει ακρίβεια με περίπου 3 δεκαδικά ψηφία.

Τέλος για την μέθοδο Natural Bicubic Splines ισχύει $\frac{1}{2}10^{-1} > 0.014262364092823278$ που μας δίνει 1 περίπου ψηφείο ακρίβειας.

Προσεγγίσεις ολοκληρώματος Στην άσκηση 6 μας ζητείται να υπολογίσουμε ένα ολοκλήρωμα χρησιμοποιώντας τις προσεγγιστικές μεθόδους τραπεζίου και Simpson καθώς και να υπολογίσουμε το σφάλμα που παρουσιάζεται στα αποτελέσματα που βγήκαν.

Η προσεγγιστική μέθοδος τραπεζίου Η μέθοδος τραπεζίου χρησιμοποιείται για την προσέγγιση εμβαδού κάτω από μία καμπύλη. Διαλέγοντας n σημεία x και ενώνοντάς τα $f(x_i)$ διαδοχικά με ευθείες γραμμές δημιουργούνται $n-1$ τραπέζια. Υπολογίζοντας το εμβαδόν του κάθε τραπεζίου και στη συνέχεια προσθέτοντας τα μας δίνει μια προσέγγιση του ολοκληρώματος που θέλουμε να βρούμε.

Η συνάρτηση Trapezoidal Rule Στην αρχή της συνάρτησης, καθώς μας ζητείται να γίνει προσέγγιση με 11 σημεία στο $[0, \frac{\pi}{2}]$ χρησιμοποιώντας την linspace αποθηκεύονται σε μία λίστα 11 ομοιόμορφα διαμοιρασμένα σημεία που ανοίκουν στον συγκεκριμένο διάστημα. Επιπροσθέτως, διαιρόντας το $\frac{\pi}{2}$ με το πλήθος των υποδιαστημάτων βρίσκουμε το πλάτος του κάθε τραπεζίου. Στην συνέχεια, χρησιμοποιώντας έναν βρόγχο και μία βοηθητική μεταβλητή sum υπολογίζεται:

$$sum = f(x_0) + 2f(x_1) + 2f(x_2) + ... + 2f(x_9) + f(x_{10})$$

Τέλος πολλαπλασιάζεται ο κοινός παράγοντάς και έτσι έχουμε το αποτέλεσμα το οποίο και επιστρέφεται:

$$\int_0^{\frac{\pi}{2}} sin(x)dx \approx \frac{w(f(x_0) + 2f(x_1) + 2f(x_2) + ... + 2f(x_9) + f(x_{10}))}{2}$$

Όπου w το πλάτος του τραπεζίου.

Η προσεγγιστική μέθοδος του Simpson Η μέθοδος του Simpson είναι άλλη μία μέθοδος προσέγγισης εμβαδού κάτω από μία καμπύλη. Η προσέγγιση γίνεται με το άθροισμα των εμβαδών των παραβολών που δημιουργούνται ανάμεσα σε κάθε $f(x_i), f(x_{i+1}), f(x_{i+2})$. Για αυτόν το λόγο, για να λειτουργήσει σωστά η συγκεκριμένη μέθοδος απαιτείται ο αριθμός υποδιαστημάτων να είναι ζυγός. Καθώς εμπλέκονται πολυώνυμα δευτέρου βαθμού, το σφάλμα που περιμένουμε είναι μικρότερο από αυτό της μεθόδου τραπεζίου.

Η συνάρτηση Simpsons Rule Η συνάρτηση αρχίζει αποθηκεύοντας 11 ομοιόμορφα διαμοιρασμένα σημεία σε μία λίστα χρησιμοποιώντας την linspace και υπολογίζει το πλάτος του κάθε υποδιαστήματος. Στην συνέχεια, με την βοήθεια ενός βρόγχου γίνεται ο εξής υπολογισμός:

$$sum = f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + ... + 4f(x_9) + f(x_{10})$$

Τέλος, πολλαπλασιάζοντας τον κοινό παράγοντα παίρνουμε το αποτέλεσμα το οποίο και επιστρέφεται απο την συνάρτηση:

$$\int_0^{\frac{\pi}{2}} sin(x)dx \approx \frac{w(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + ... + 4f(x_9) + f(x_{10}))}{3}$$

Όπου w το πλάτος του υποδιαστήματος.

Τα αριθμητικά σφάλματα που εμφανίζονται στις δύο μεθόδους είναι τα εξής:

Αριθμητικό Σφάλμα	
Τραπεζίου	0.0020570136456427024
Simpson	$3.3922209004000337 \cdot 10^{-6}$

Για τον υπολογισμό του θεωρητικού σφάλματος του τραπεζίου παίρνουμε τον εξής τύπο:

$$\frac{K(b-a)^3}{12n^2}, K \geq max(|f''(x)| : x \in [a,b])$$

Έχουμε $f''(x) = -sin(x)$ και έτσι $K = 1$ καθώς $|-1|$ είναι η μέγιστη τιμή. Το τελικο αποτέλεσμα είναι:

$$errorT \leq \frac{3.87578}{1200} = 0.00322\ldots$$

Για τον υπολογισμό του θεωρητικού σφάλματος του Simpson παίρνουμε τον εξής τύπο:

$$\frac{K(b-a)^5}{180n^4}, K \geq max(|f''(x)| : x \in [a,b])$$

Έχουμε $f''(x) = -sin(x)$ και έτσι $K = 1$ καθώς $|-1|$ είναι η μέγιστη τιμή. Το τελικο αποτέλεσμα είναι:

$$errorS \leq \frac{9.56311}{1800000} = 0.00000531284\ldots$$

Εκτίμηση τιμών μετοχών Στην άσκηση αυτή μας ζητείται να επιλέξουμε δύο εταιρείες, και με βάση τις τιμές κλεισίματος ημέρας των μετοχών τους από 10 διαφορετικές ημέρες, να υπολογίσουμε προσεγγιστικά την επόμενη τιμή κλεισίματος ημέρας καθώς και 5 τιμές μετά. Οι προσεγγίσεις αυτές πρέπει να γίνουν με την μέθοδο ελαχίστων τετραγώνων 2ου, 3ου και 4ου βαθμού. Οι 10 αυτές μέρες επιλέγονται πριν τα γενέθλια μου που είναι στις 14/09. Οι εταιρείες που διάλεξα είναι η ΕΛΛΗΝΙΚΆ ΠΕΤΡΈΛΑΙΑ και η TITAN CEMENT INTERNATIONAL με τιμές:

Τιμές κλεισίματος ημέρας		
Ημερομηνία	ΕΛΠΕ	TITC
31/08	5.1100	11.3800
1/09	5.2000	11.3800
2/09	5.2400	11.2600
3/09	5.1700	11.2400
4/09	5.1800	11.1600
7/09	5.2500	11.1200
8/09	5.1200	10.7800
9/09	5.1100	10.8800
10/09	5.1100	10.8000
11/09	5.0600	10.8200

Χρησιμοποιώντας αυτά τα δεδομένα για να προσεγγίσουμε την επόμενη τιμή κλεισίματος ημέρας που πέφτει στα γενέθλιά μου (14/09) με πολυώνυμα 2ου,3ου και 4ου βαθμού παίρνουμε:

Προσέγγιση τιμής κλεισίματος ημέρας (14/09/2020)		
Βαθμός Πολ.	ΕΛΠΕ	TITC
2	4.87896240601507	10.61870676691737
3	4.954783999782925	10.84183223982999
4	4.7244285487611215	11.551881588281635

Η τιμή κλεισίματος της ΕΛΠΕ για τις 14/09 είναι 5,1800 που σημαίνει ότι η καλύτερη προσέγγιση έγινε από το πολυώνυμο τρίτου βαθμού ενώ η χειρότερη προσέγγιση παρατηρείται στο πολυώνυμο τετάρτου βαθμού. Σε κάθε περίπτωση, παρ΄ όλο που η απόκλιση των τιμών είναι μικρή, παρουσιάζεται μείωση σε σχέση με την προηγούμενη τιμή κάτι που μπορεί να θεωρηθεί προβληματικό στην ακρίβεια προσέγγισης μετοχών. Στην περίπτωση της TITC με τιμή κλεισίματος ημέρας 11,3000, παρατηρούμε ότι το πολυώνυμο 4ου βαθμού μας δίνει ένα σχετικά ακριβές αποτέλεσμα, ενώ η χειρότερη προσέγγιση γίνεται από του 2ου βαθμού πολυώνυμο που όχι μόνο έχει την μεγαλύτερη απόκλιση άλλα και φαίνεται να μειώνεται η τιμή κλεισίματος ημέρας.

Υπολογίζοντας την τιμή κλεισίματος ημέρας για 5 συνεδριάσεις μετά εμφανίζονται τα εξής αποτελέσματα:

Προσέγγιση τιμής κλεισίματος ημέρας (18/09/2020)		
Βαθμός Πολ.	ΕΛΠΕ	TITC
2	4.54654135338348	10.427689223057708
3	4.841960336958779	11.297039195035989
4	3.3587512879035764	15.868893754342455

Στις 18/09/2020 η τιμή κλεισίματος ημέρας για τις μετοχές της ΕΛΠΕ ήταν 5,1600. Παρατηρείται ότι, όπως και στην προηγούμενη περίπτωση, την μικρότερη απόκλιση για τις συγκεκριμένες τιμές μετοχών την πετυχαίνει το πολυώνυμο 3ου βαθμού, μετά του 2ου και τέλος του 4ου που παρουσιάζει μια αναλογικά πολυ μεγάλη απόκλιση τιμής. Το ίδιο φαινόμενο παρουσιάζεται και στις μετοχές της TITC με τιμή κλεισίματος ημέρας στις 18/09/2020, 11,1000. Παρότι το ζητούμενο y είναι πιο μακριά, και έτσι περιμένουμε χειρότερη προσέγγιση, βλέπουμε ότι το πολυώνυμο 3ου βαθμού μας επιστρέφει μία σχετικά ακριβής εκτίμηση του πραγματικού αποτελέσματος. Στου 2ου βαθμού εμφανίζεται μία μεγαλύτερη απόκλιση ενώ στου 4ου φαίνεται η προσέγγιση να αρχίζει να ξεφεύγει απο την πραγματική τιμη.

Τελικές παρατηρήσεις Σε γενικές γραμμές παρατηρείται ότι οι καλύτερες προσεγγίσεις παράγονται με την χρήση 3ου βαθμού πολυωνύμου ενώ η μεγαλύτερη απόκλιση φαίνεται στου 4ου βαθμού. Τα αποτελέσματα αυτά γίνονται φανερά καθώς στις τιμές των συγκεκριμένων μετοχών σημειώνονται πολύ μικρές αυξομειώσεις. Διαλέγοντας ένα πολυώνυμο με γραμμική παλινδρόμηση μας δίνει πιο ακριβής εκτιμήσεις από πολυώνυμα μεγαλύτερου βαθμού. Αυτό γίνεται αντιληπτό παίρνοντας αποτελέσματα προσεγγίσεων για τις μετοχές βαθμού μεγαλύτερου του 4 όπου οι τελικές τιμές αρχίζουν να γίνονται όλο και χειρότερες.