

ЕДНО ДОПЪЛНЕНИЕ КЪМ SCATTER-ДИАГРАМИТЕ В EXCEL

Мариана Трифонова¹, Лина Драганова²

¹ Минно-геоложки университет "Св. Иван Рилски", 1700 София, trifonova.m@gmail.com

² студент, lin4eto23@yahoo.com

РЕЗЮМЕ. MS Excel е един от най-популярните продукти за съхраняване, обработка и графично представяне на информация в табличен вид. Често пъти резултатите от инженерната работа се представят в Excel посредством Scatter-диаграми. Независимо от богатата функционалност на продукта, обектният модел на MS Excel не предоставя вградени инструменти за извличане на числова информация от интерполиращата гладка крива извън възлите. В настоящия доклад е дадено сравнение между различни методи за интерполация на резултати от изследвания: Представен е и разработеният от авторите add-ins *Interpolation* за MS Excel, чрез който става възможно да се получат координатите на произволна точка от изглаждащата крива в Scatter-диаграмата.

Ключови думи: MS Excel, Scatter-диаграма, интерполация, апроксимация, кубични сплайни, сплайни на Акима.

AN AMPLIFICATION TO SCATTER DIAGRAMS IN MS EXCEL

Mariana Trifonova¹, Lina Draganova²

¹ University of Mining and Geology "St. Ivan Rilski", 1700 Sofia, trifonova.m@gmail.com

² student, lin4eto23@yahoo.com

ABSTRACT. MS Excel is one of the most popular products aimed at storage, proceeding and graphic presentation of information in the form of tables. The results of engineering works are often represented in Excel as Scatter diagrams. Despite the large functionality of the product, Excel's object model doesn't offer incorporated instruments for obtaining of numeric information from an interpolating curve out of the nodes. The present report offers comparison between various interpolation methods applied to results from studies. An add-ins *Interpolation* developed for Excel by authors is represented as well. It makes possible for the coordinates of any point of the smoothing Scatter diagram's curve to be found.

Keywords: MS Excel, Scatter diagram, interpolation, approximation, cubic splines, Akima splines.

Въведение

Диаграмите в Excel, включително scatter-диаграмите, представляват растерно изображение. Microsoft не обявява вида на използваната интерполираща крива. Независимо че обектният модел на MS Excel е изключително богат (около 200 класа с над 5000 елемента, като около една четвърт от тях са свързани с диаграмите), към момента липсват каквито и да било свойства и/или методи, които да дадат възможност да се получат координатите на точки от интерполиращата крива. За потребителя остават две възможности: ръчно „на око“ да се свалят координатите на точките, които представляват интерес, или да се използват някои от вградените функции за тренд анализ. Очевидно и двата подхода са неудобни.

Първият поради затрудненията, които създава, особено ако трябва да се прилага многократно. Тук не се третира въпроса за точността, тъй като самата интерполация гарантира точност само във възлите, но не и в междинните точки.

Втората възможност пък е направо неприложима за комбиниране със scatter-диаграми, тъй като при тях се използва интерполираща гладка крива, докато вградените тренд функции в MS Excel са базирани единствено на

апроксимация предимно с линейни математически функции.

Например, данните от табл. 1 са представени със scatter-диаграма на фиг. 1.

Във втората колона на табл. 2 са дадени получените стойности за Y с използване на функцията TREND, приложена за възлите от табл. 1.

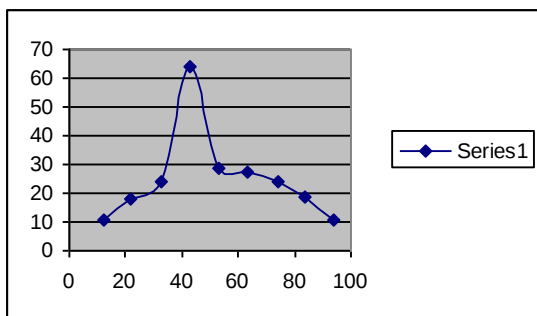
Таблица 1

X	Y
12	11
22	18
32.5	24
43	64
53	29
63	27.5
74	24
83.5	18.5
94	11

Таблица 2

X	Y
18	27.1442
28	26.5951
38	26.0459
48	25.4968
58	24.9477
68	24.3985
78	23.8494
88	23.3003

Очевидно се получават точки, които са доста далече от графиката на функцията.



Фиг. 1

Разбира се, има и трета възможност: да се премине към друг софтуер, където подобни проблеми не съществуват.

Постановка и решение на задачата

Да се създаде модел в MS Excel, допълнен със съответно програмно осигуряване, при който зададени точки да се интерполират с гладка крива, максимално близка до scatter-диаграмите и да има възможност да се получават ординатите на произволни точки от тази линия между началния и крайния възел.

Понятията интерполация и апроксимация в математиката са известни и изследвани още преди повече от два века (интерполационният полином на Лагранж е познат от осемнадесети век). Тъй като scatter-диаграмите *интерполират* зададени точки, в настоящата статия ще се спрем само на този въпрос. Има много известни начини за интерполиране: с полиноми на Лагранж, на Чебишев, по метода на Ермит, със сплайни от различен ред и др. Разнообразието е главно в теоретичен аспект. На практика масово за интерполация в софтуерните продукти се използват сплайн функции, и по-конкретно сплайни от трети ред, наречени кубични сплайни. Задачата е следната:

Дадени са точките $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ при условие $x_0 < x_1 < \dots < x_n$ (не се изисква абсцисите да са равномерни, т.е. допустимо е $x_i - x_{i-1} \neq x_{i+1} - x_i$ за всяко i от 1 до $n-1$). По-нататък в изложението тези точки често пъти ще се наричат *възли*. Търси се функция $f(x)$, отговаряща на следните условия:

1. Във всеки подинтервал $[x_i, x_{i+1}]$ функцията $f(x)$ да е полином от степен максимум 3.
2. Функцията $f(x)$, както и нейните първа и втора производна да са непрекъснати във възлите.

Както се вижда от постановката на задачата, функцията $f(x)$ в различни части от дефиниционния си интервал представлява полином от степен максимум трета, при това полиномите в различните подинтервали в общия случай са различни. Такива функции се наричат *частично полиномиални функции* или *сплайни*. В случаят, сплайнът е от трети ред. Може да се постави задачата в различните подинтервали функцията да се представя с полиноми от по-висока степен, например степен k . Тогава се добавят и условия за непрекъснатост във възлите на всички производни на въпросната функция от ред 1, 2, ..., $k-1$. Наблюденията показват, че в практиката се използват предимно сплайни от трети ред, в изключително редки

случаи – от четвърти. На авторите не са известни софтуерни продукти, в които да са реализирани и да се използват сплайни от ред, по-голям от 4. Това е обяснимо, тъй като по-голямата степен предполага не само повече изчисления, но и генерира крива с повече осцилации.

Описаните по-горе изисквания за функцията $f(x)$ не са достатъчни за еднозначното ѝ определяне. Необходимо е да се добавят още две условия.

За получаването на функцията $f(x)$ се постъпва по следния начин. Да положим $f''(x_i) = s_i$ и $h_i = x_i - x_{i-1}$. Тъй като във всеки подинтервал $[x_{i-1}, x_i]$ функцията $f(x)$ е полином от трета степен, то в същия подинтервал $f''(x)$ е линейна и се представя с

$$f''(x) = \frac{s_{i-1}(x_i - x) + s_i(x - x_{i-1})}{h_i} \quad (1)$$

След двукратно интегриране на (1) се получава

$$f(x) = \frac{s_{i-1}(x_i - x)^3 + s_i(x - x_{i-1})^3}{6h_i} + c_1x + c_2 \quad (2)$$

където константите c_1 и c_2 се определят от условията

$$f(x_{i-1}) = y_{i-1} \text{ и } f(x_i) = y_i.$$

Поради непрекъснатостта на първата производна на $f(x)$ във възлите след диференциране на (2) и кратки преобразования се получава

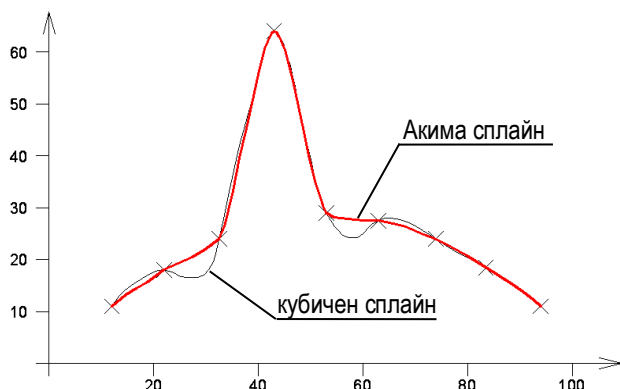
$$\begin{aligned} h_i s_{i-1} + 2(h_i + h_{i+1}) s_i + h_{i+1} s_{i+1} = \\ = \frac{6(y_{i+1} - y_i)}{h_{i+1}} - \frac{6(y_i - y_{i-1})}{h_i} \end{aligned} \quad (3)$$

Тъй като равенство (3) е валидно за всеки подинтервал, то фактически (3) представлява система от $n-1$ уравнения с $n+1$ неизвестни $s_0, s_1, \dots, s_n$. За определянето на тези неизвестни, а оттам и на функцията $f(x)$ е необходимо да се добавят още две условия. Тези условия могат да бъдат:

- нулева кривина в крайните възли, т.е. $s_0 = s_n = 0$, в този случай се говори за естествен сплайн;
- задаване на стойности на първата производна в крайните възли (напр. така се построява сплайна в AutoCAD);
- задаване на стойностите на първата и втората производна в началния възел (малко изкуствен и невинаги приложим подход, но пък изчисленията се опростяват, тъй като се избягва решаването на голяма система линейни уравнения);
- избор на квадратична функция, а не полином от трета степен в крайните подинтервали, т.е. $s_0 = s_1$ и $s_{n-1} = s_n$ и др.

В литературата са известни и модификации на кубичните сплайни, например сплайни на Ермит, за чието генериране е необходимо освен възлите, да се зададат и стойностите на първата производна за всяко x_i , т.е. да се определи наклона на допирателната във всеки възел.

През 1970г. японският учен Хироши Акима публикува още един подход за интерполация, който вече носи неговото име – Акима сплайн. Тези сплайни са създадени да неутрализират един от недостатъците на класическите кубични сплайни, които са склонни към колебание в околностите на точките, при които се постига някакъв пик. На фиг. 2 е представен резултатът от интерполирането на точките от таблица 1 с класически кубичен сплайн и сплайн на Акима.



Фиг. 2

Както се вижда от фигурата, Акима сплайнът е много по-стабилен и практически отсъстват признаци на осцилация в подинтервалите, съседни на пиковата точка.

И така, нека са дадени отново точките $(x_0, y_0), (x_1, y_1) \dots (x_n, y_n)$. Акима налага ограничението във всеки възел (x_i, y_i) наклонът t_i на кривата да е равен на

$$t_i = \frac{m_{i-1} |m_{i+1} - m_i| + m_i |m_{i-1} - m_{i-2}|}{|m_{i+1} - m_i| + |m_{i-1} - m_{i-2}|}, \quad (4)$$

където всеки от коефициентите m_i в (4) представлява

наклона на линейните сегменти около възела и се пресмята по формулата

$$m_i = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \quad (5)$$

В случай, че знаменателят в (4) е равен на 0, то

$$t_i = 0.5(m_i + m_{i-1}) \quad (6)$$

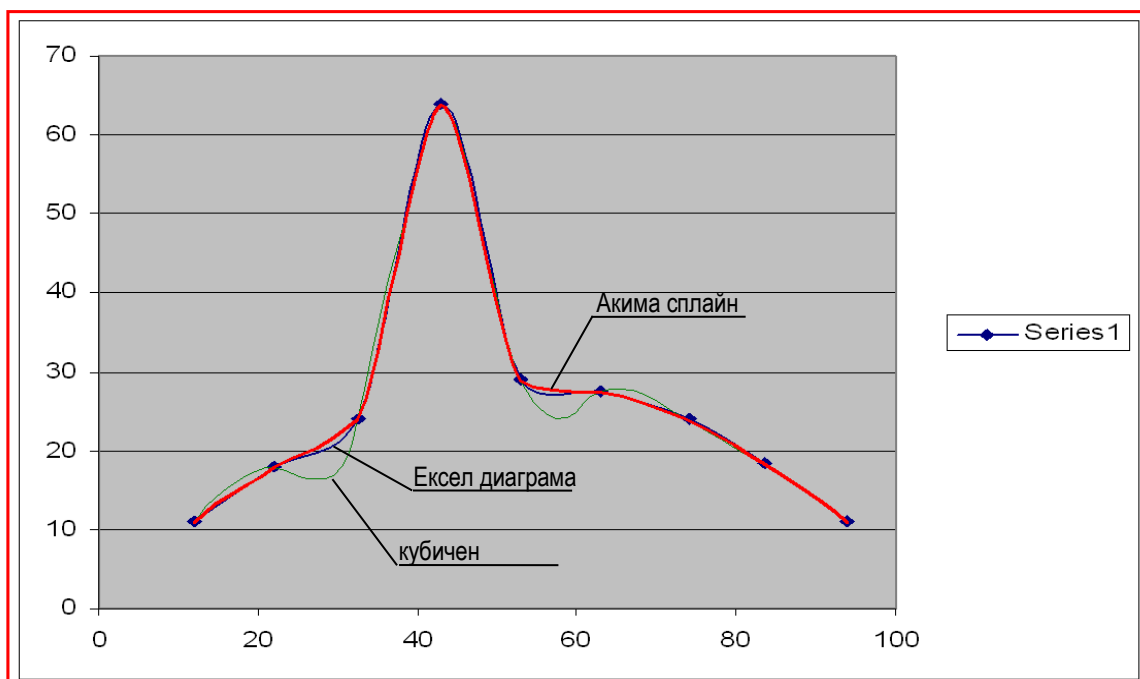
След определяне на коефициентите t_i може да се пристъпи към генериране на сплайна във всеки подинтервал $[x_i, x_{i+1}]$ по формулата

$$y = p_0 + p_1(x - x_i) + p_2(x - x_i)^2 + p_3(x - x_i)^3 \quad (7)$$

Където:

$$\begin{aligned} p_0 &= y_i, & p_1 &= t_i \\ p_2 &= \frac{3 \frac{y_{i+1} - y_i}{x_{i+1} - x_i} - 2t_i - t_{i+1}}{x_{i+1} - x_i} \\ p_3 &= \frac{t_i + t_{i+1} - 2 \frac{y_{i+1} - y_i}{x_{i+1} - x_i}}{(x_{i+1} - x_i)^2} \end{aligned} \quad (8)$$

Както се вижда от формули (4)-(8), сплайнът на Акима притежава свойство „локалност“ в известен смисъл. За да може да се построи са необходими не по-малко от 5 възли. По данни на Акима във вътрешните подинтервали грешката е от порядъка на $O(h^2)$ и генерирането на самия сплайн се извършва по-бързо в сравнение с класическите кубични сплайни.



Фиг. 3

Разработеният add-ins **Interpolation** за MS Excel дава възможност поредица от точки да се *интерполира* с класически кубични сплайни и със сплайни на Акима и да се изведат координатите на произволни точки от получения сплайн. На фиг. 3 са представени три графики, построени по точките от таблица 1. Едната е вградената scatter-диаграма в Excel, а другите две са получени с помощта на функциите от Interpolation, като са наложени една върху друга в средата на векторна система за сравнение. Както се вижда от фигурата, scatter-диаграмата и графиката на Акима сплайна почти съвпадат. Графиката на класическия кубичен сплайн се различава от графиките на другите две в околност на „пиковата точка“.

Експериментите показват:

- в случай, че точките са разположени монотонно, scatter-диаграмата и графиката на Акима сплайна се припокриват, а графиката на класическия кубичен сплайн се различава от тях пренебрежимо малко;
- само в случай на значително отклонение на една от точките, има пренебрежимо малка разлика между scatter диаграмата в Excel и графиката на Акима сплайна, която разлика за практически цели може да се пренебрегне. В този случай класическият кубичен сплайн се различава съществено от другите две графики в подинтервалите от двете страни на точката с отклонението.

Приложението **Interpolation** е разработено на VBA (Visual Basic for Application). Зарежда се като стандартен add-ins в MS Excel и диалогът с неговите функции се подчинява на принципите на работа в MS Excel. Единствената разлика е, че към момента в неговите функции няма добавена вградена помощна информация.

Заклучение

Разработено е допълнение към MS Excel, което дава възможност да се получават координати на точки от гладка крива, *интерполираща* набор възли. Графиката на самата интерполираща крива може да се изчертае както чрез вградената scatter диаграма, така и чрез функция от допълнително заредения add-ins **Interpolation**. При scatter-диаграмата обаче, възможностите за форматиране на елементите ѝ са по-мощни.

При анализ на точки от измервания и необходимост от получаване на координати на междинни точки, според характера на решаваната задача и поставените цели, въпрос на избор от страна на потребителя е да избере един от дава подхода:

- Апроксимация на точките: В този случай потребителят може да ползва някои от вградените функции на MS Excel, напр. TREND функцията.
- Интерполация на точките: Тъй като MS Excel не предоставя вградени функции за решаване на проблема, биха могли да се ползват инструментите на създадения add-ins **Interpolation**.

Литература

- Hiroshi Akima, A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures, Journal of the Association for Computing Machinery, Vol. 17, № 4, October 1970, pp. 590-602.
- А. Фокс, М. Прат, Вычислительная геометрия, Применение в проектировании и на производстве, Москва, „Мир“, 1982
- Роджерс Д., Дж. Адамс, Математические основы машинной графики, Москва, 1980.

Статията е препоръчана за публикуване от кат. „Информатика“