Тема 4. Аппроксимация функций и смежные вопросы (продолжение).

Среднеквадратичная аппроксимация функций (метод наименьших квадратов, МНК). Постановка задачи.

Критерий интерполирования, предполагающий совпадение исходной и аппроксимирующей функций в узлах таблицы, не является единственным. Обратимся к экспериментальным данным, представленным на рис. 1а. Если выполнить по ним интерполяцию, то получится кривая на рис. 1б. Маловероятно, чтобы ее вид отвечал исходной зависимости, положенной в основу таблицы. Вероятнее всего, что этой зависимости отвечает кривая, похожая на рис. 1в, а отклонение экспериментальных данных от нее продиктовано сравнительно большой погрешностью измерений. В таком случае целесообразно использовать среднеквадратичный критерий

$$\rho^2 = \sum_{k=1}^{N} (f(x_k) - g(x_k))^2$$

или

$$\rho^2 = \int_a^b (f(x) - g(x))^2 dx,$$

если аппроксимируемая функция задана на $[a, b]$ непрерывно.

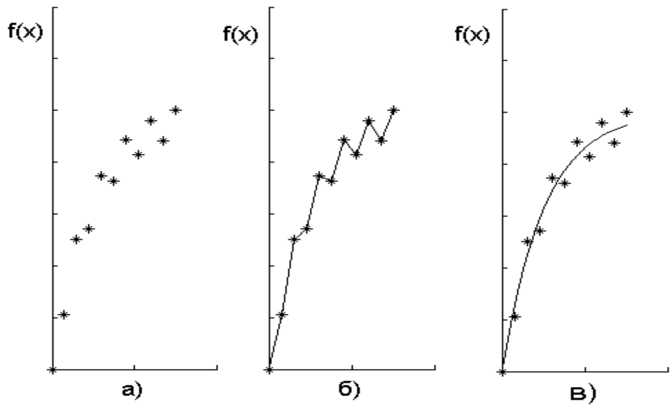

Рис. 1. Аппроксимация экспериментальных данных

Напомним, что, как уже обсуждалось, близость функций по среднеквадратичному критерию еще не гарантирует малой величины их максимальной разности

$$\delta = \max_{x \in [a,b]} \left| f(x) - g(x) \right|.$$

Малое значение интеграла или суммы для ρ^2 свидетельствует лишь о том, что почти на всем отрезке $[a, b]$ значения $f(x)$ и $g(x)$ мало отличаются друг от друга, хотя в отдельных точках или на небольших отрезках разность их значений может быть значительной.

Проблема выбора аппроксимирующей функции решается так же, как и при интерполяции: $Q(x)$ выбирается в виде обобщенного многочлена:

$$Q_m(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \ldots + a_m\varphi_m(x) = \sum_{k=0}^{m} a_k\varphi_k(x), \qquad (1)$$

где $\{\varphi_k\}$ – заданный набор линейно независимых функций, а коэффициенты a_k подлежат определению.

Дискретный случай. Весовые коэффициенты.

Функция $f(x)$ задается на дискретном множестве точек следующей таблицей,

x	x_1	x_2	x_3	…	x_N
$f(x)$	$f(x_1)$	$f(x_2)$	$f(x_3)$	…	$f(x_N)$

а ее аппроксимация $Q_m(x)$ выбирается в виде обобщенного многочлена (1). Коэффициенты a_k выбираются из условия минимума величины ρ^2

$$\rho^2 = \sum_{i=1}^{N} \left(Q_m(x_i) - f(x_i)\right)^2 \to \min. \qquad (2)$$

Рассмотрим три варианта соотношений чисел N и m.

1. $N = m+1$. Число коэффициентов a_k равно числу точек таблицы, решение задачи единственное, и им является интерполяционный полином, проходящий через все точки. Минимальное значение ρ^2 оказывается нулевым.
2. $N < m+1$. Минимальное значение ρ^2 также равно нулю, но задача имеет бесконечное множество решений.
3. $N > m+1$. Это типичный случай среднеквадратичной аппроксимации. Более того, на практике часто $N \gg m+1$. Минимальное значение ρ^2 оказывается уже, как правило, ненулевым, а задача имеет единственное решение. В настоящем параграфе ограничимся этим третьим вариантом, так как первый был уже достаточно рассмотрен, а ко второму вернемся позднее.

Записываем необходимое условие экстремума

$$\frac{\partial\rho^2}{\partial a_k} = 0, \qquad k = 0,\ 1,\ 2, \ldots, m,$$

и выполняем операцию дифференцирования:

$$\frac{\partial\rho^2}{\partial a_k} = 2\sum_{i=1}^{N} \left(Q_m(x_i) - f(x_i)\right)\cdot\varphi_k(x_i) = 0.$$

Подставляя в получившуюся формулу выражение для $Q_m(x)$, получаем систему линейных алгебраических уравнений относительно a_k:

$$a_0\sum_{i=1}^{N}\varphi_0(x_i)\cdot\varphi_k(x_i) + a_1\sum_{i=1}^{N}\varphi_1(x_i)\cdot\varphi_k(x_i) + \ldots + a_m\sum_{i=1}^{N}\varphi_m(x_i)\cdot\varphi_k(x_i) = \sum_{i=1}^{N} f(x_i)\cdot\varphi_k(x_i),$$

$$k = 0,1,\ldots m; \qquad (3)$$

Если определитель системы (3) не равен нулю, задача имеет единственное решение. Самой популярной является аппроксимация полиномами, когда $\varphi_k(x) = x^k$, а система (3) приобретает вид

$$\left(\sum_{i=1}^{N} 1\right) a_0 + \left(\sum_{i=1}^{N} x_i\right) a_1 + \ldots + \left(\sum_{i=1}^{N} x_i^{\,m}\right) a_m = \left(\sum_{i=1}^{N} f(x_i)\right),$$

$$\left(\sum_{i=1}^{N} x_i\right) a_0 + \left(\sum_{i=1}^{N} x_i^{\,2}\right) a_1 + \ldots + \left(\sum_{i=1}^{N} x_i^{\,m+1}\right) a_m = \left(\sum_{i=1}^{N} f(x_i) x_i\right), \tag{4}$$

…..

$$\left(\sum_{i=1}^{N} x_i^{\,m}\right) a_0 + \left(\sum_{i=1}^{N} x_i^{\,m+1}\right) a_1 + \ldots + \left(\sum_{i=1}^{N} x_i^{\,2m}\right) a_m = \left(\sum_{i=1}^{N} f(x_i) x_i^{\,m}\right).$$

При выполнении среднеквадратичной аппроксимации (другое название – «метод наименьших квадратов») возможна ситуация, когда исходные данные имеют различную точность. Если к каким-либо экспериментальным значениям доверие выше, т.е. они являются более надежными по сравнению с другими, это может быть учтено введением в критерий (2) положительных весовых коэффициентов p_i

$$\rho^2 = \sum_{i=1}^{N} p_i \left(Q_m(x_i) - f(x_i)\right)^2 \to \min. \tag{5}$$

***Вопрос 12**. Закончите следующую фразу.*

Для тех точек, степень доверия которым выше и к которым аппроксимирующую кривую желательно провести ближе, чем к другим точкам, весовые коэффициенты следует задавать … (больше или меньше?).

Конечно больше. Величину ρ^2 можно трактовать, как своеобразную «функцию штрафа». За отклонение $Q_m(x)$ от $f(x)$ в точке x_i к значению ρ^2 добавляется слагаемое («штраф») $\left(Q(x_i) - f(x_i)\right)^2$ тем большее, чем больше это отклонение. Если какая-то точка является более приоритетной и к ней аппроксимирующую кривую желательно провести ближе, с помощью весового коэффициента за отклонение в этой точке «штраф» должен быть увеличен.
На практике положительные весовые коэффициенты p_i часто задают так, чтобы их сумма была равна, например, единице или 100. Последнее часто удобно, но не обязательно. Если все коэффициенты умножить на одно и то же число, то, хотя ρ^2 и изменится, решение задачи останется прежним. Важными являются отношения p_i друг к другу.

Непрерывный случай. Понятие ортогональности.

Теперь обратимся к варианту непрерывного задания $f(x)$ на $[a, b]$.

$$\rho^2 = \int_a^b \left(Q(x) - f(x)\right)^2 dx \to \min . \tag{6}$$

Выражение (3) сохранит похожий вид, только вместо сумм появятся интегралы

$$a_0 \int_a^b \varphi_0(x)\cdot\varphi_k(x)dx + a_1 \int_a^b \varphi_1(x)\cdot\varphi_k(x)dx + \ldots + a_m \int_a^b \varphi_m(x)\cdot\varphi_k(x)dx = \int_a^b f(x)\cdot\varphi_k(x)dx\,,$$

$$k = 0,1,\ldots m\,; \tag{7}$$

В качестве примера обратимся к аппроксимации полиномами $\varphi_k(x) = x^k$ на промежутке $[0, 1]$. Определитель системы (7) имеет вид

$$G_m = \begin{vmatrix} 1 & 1/2 & 1/3 & \ldots & 1/(m+1) \\ 1/2 & 1/3 & 1/4 & \ldots & 1/(m+2) \\ 1/3 & 1/4 & 1/5 & \ldots & 1/(m+3) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1/(m+1) & 1/(m+2) & 1/(m+3) & \ldots & 1/(2m+1) \end{vmatrix}. \tag{8}$$

Он называется определителем Гильберта, а соответствующая ему матрица – матрицей Гильберта. Она, как и матрица системы (4), характеризуется крайне неприятным свойством: весьма малые изменения ее элементов приводят к сильному изменению решения системы. Определитель Гильберта стремительно уменьшается с ростом порядка системы:

m:	1	2	3	4	...	8	9
G_m:	1	$8.3*10^{-2}$	$4.6*10^{-4}$	$1.7*10^{-7}$	...	$2.7*10^{-33}$	$9.7*10^{-43}$

Аналогично дискретному случаю критерий ρ^2 может быть обобщен введением положительной весовой функции $p(x)$

$$\rho^2 = \int_a^b p(x) \cdot \left(Q(x) - f(x)\right)^2 dx \to \min$$

Все формулы сохранят похожий вид, а под знаком интеграла появится $p(x)$

$$a_0 \int_a^b p(x) \cdot \varphi_0(x) \cdot \varphi_k(x) dx + \ldots + a_m \int_a^b p(x) \cdot \varphi_m(x) \cdot \varphi_k(x) dx = \int_a^b p(x) \cdot f(x) \cdot \varphi_k(x) dx,$$

$$k = 0,1,\ldots m; \qquad (9)$$

Решение системы (9) значительно упрощается, если вместо произвольных линейно независимых функций $\{\varphi_k(x)\}$ воспользоваться ортогональными функциями $\{g_k(x)\}$.

Последовательность функций $\{g_k(x)\}$ является ортогональной на промежутке $[a,b]$ с весом $p(x)$, если выполняются следующие условия

$$\int_a^b p(x) g_k(x) g_i(x) dx = \begin{cases} 0, \text{ если } i \neq k \\ A > 0, \text{ если } i = k \end{cases}$$

Если в дополнение $A = 1$, то такие функции называются *ортонормированными* (т.е. ортогональными и нормированными). Для ортогональных функций все интегралы, кроме одного, в левой части (9) равны нулю, матрица этой системы оказывается диагональной, и каждое уравнение дает готовое выражение для коэффициента a_k

$$a_k = \frac{\int_a^b p(x) f(x) g_k(x) dx}{\int_a^b p(x) {g_k}^2(x) dx} \qquad (10)$$

Если исходный набор функций $\left\{\varphi_k(x)\right\}$ не является ортогональным, его можно сделать таковым, используя процедуру Грама–Шмидта.

Процедура ортогонализации Грама-Шмидта

Задан набор линейно независимых функций

$\varphi_0(x), \varphi_1(x), \varphi_2(x), \ldots, \varphi_m(x), \ldots$

Требуется построить набор ортогональных функций

$g_0(x), g_1(x), g_2(x), \ldots, g_m(x), \ldots$

которые не только ортогональны, но и являются линейной комбинацией функций $\varphi_k(x)$. Аппроксимация, таким образом, будет выполняться в том же классе функций.

Введем также следующее обозначение.

$\tilde{g}_k(x)$ – функции ортогональные, но еще не нормированные.

Очередная функция $\tilde{g}_k(x)$ строится так, чтобы она была ортогональной всем $\tilde{g}_i(x)$, построенным до нее.

$\tilde{g}_0(x) = \varphi_0(x)$. Нормируем ее.

$$\int_a^b p(x)\tilde{g}_0^2 dx = \alpha_0^2 \ , \ g_0(x) = \frac{\tilde{g}_0}{\alpha_0} .$$

Функция $\tilde{g}_m(x)$ строится с привлечением новой $\varphi_m(x)$ и добавлением линейной комбинации функций $g_k(x)$, построенных на предыдущих шагах.

$$\tilde{g}_m(x) = \varphi_m(x) - \sum_{k=0}^{m-1} C_{m,k} \cdot g_k(x)$$

Из условия ортогональности $\tilde{g}_m(x)$ и $g_i(x)$ получаем выражения для коэффициентов $C_{m,i}$

$$\int_a^b p(x)\cdot \tilde{g}_m(x)\cdot g_i(x)dx = \int_a^b p(x)\cdot \varphi_m(x)\cdot g_i(x)dx - \sum_{k=0}^{m-1} C_{m,k}\cdot \int_a^b p(x)\cdot g_k(x)\cdot g_i(x)dx = 0 .$$

Все интегралы под знаком суммы, кроме одного, равны нулю и для $C_{m,i}$ имеем

$$C_{m,i} = \frac{\int_a^b p(x)\cdot \varphi_m(x)\cdot g_i(x)dx}{\int_a^b p(x)\cdot g_i^2(x)dx}, \quad \int_a^b p(x)\tilde{g}_m^2 dx = \alpha_m^2, \quad g_m(x) = \frac{\tilde{g}_m}{\alpha_0} .$$

Примеры ортогональных полиномов

Неотъемлемыми атрибутами понятия ортогональности являются промежуток интегрирования и весовая функция. Проблема различных промежутков, возникающих на практике, решается легко. Полиномы для стандартных промежутков (обычно это $[-1, 1]$ или $[0, 1]$) приводятся в справочниках и учебниках, а к произвольному промежутку переходят обычной заменой переменных. Примером, является следующая замена

$$x = \frac{a+b}{2} + \frac{b-a}{2}t, \quad x \in [a,b], \quad t \in [-1,1].$$

В приводимых примерах остановимся на стандартном промежутке $[-1,+1]$. Тогда главной отличительной особенностью различных полиномов будет весовая функция.

Ортогональные полиномы Лежандра. Для этих полиномов весовая функция имеет популярный вид: $p(x) \equiv 1$ и сами они могут быть вычислены по формуле

$$L_n(x) = \frac{(-1)^n}{n!2^n}\frac{d^n}{dx^n}\left[\left(1-x^2\right)^n\right], \qquad x \in [-1,1]. \tag{11}$$

Легко заметить, что $L_0(x)=1$, $L_1(x)=x$, и имеет место следующее рекуррентное соотношение:

$$(n+1)L_{n+1}(x) - (2n+1)xL_n(x) + nL_{n-1}(x) = 0 . \tag{12}$$

Тогда для последующих полиномов легко получаем

$$L_2(x) = \frac{3x^2-1}{2}, \quad L_3(x) = \frac{5x^3-3x}{2}, \quad L_4(x) = \frac{35x^4-30x^2+3}{8}, \ldots$$

В такой форме полиномы Лежандра ортогональны на $[-1,+1]$, но не нормированы. Квадрат их нормы имеет вид

$$\int_{-1}^{1} L_n^2(x)dx = \frac{2}{2n+1},$$

а графики первых четырех полиномов представлены на рис. 2а.

Ортогональные полиномы Чебышева. Для этих полиномов весовая функция выглядит следующим образом: $p(x) = \frac{1}{\sqrt{1-x^2}}$.

При $x \in [-1,1]$ они могут быть вычислены по формуле

$$T_n(x) = \cos(n \cdot \arccos(x)); \quad T_0(x) = 1, \quad T_1(x) = x. \tag{13}$$

Как и для полиномов Лежандра, здесь имеет место рекуррентное соотношение

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \tag{14}$$

на основе которого легко получаются последующие полиномы

$$T_2(x) = 2x^2 - 1, \quad T_3(x) = 4x^3 - 3x, \quad T_4(x) = 8x^4 - 8x^2 + 1, ...$$

В справедливости (14) легко убедиться. Подставляя $\varphi = \arccos(x)$ в очевидное соотношение

$$\cos(n+1)\varphi + \cos(n-1)\varphi = 2\cos\varphi\cos(n\varphi),$$

приходим к (14). Полиномы Чебышева могут быть представлены и в ином виде, отличном от (13)

$$T_n(x) = \frac{(-2)^n n!}{(2n)!}\sqrt{1-x^2}\,\frac{d^n}{dx^n}\left[\left(1-x^2\right)^{n-1/2}\right]. \tag{15}$$

Как и полиномы Лежандра, полиномы Чебышева в форме (13) ортогональны, но не нормированы. Квадраты их нормы имеют вид

$$(T_0, T_0) = \pi, \qquad (T_n, T_n) = \int_{-1}^{1} \frac{T_n^2(x)}{\sqrt{1-x^2}}dx = \frac{\pi}{2}, \qquad n>0,$$

а графики первых четырех полиномов представлены на рис. 2б.

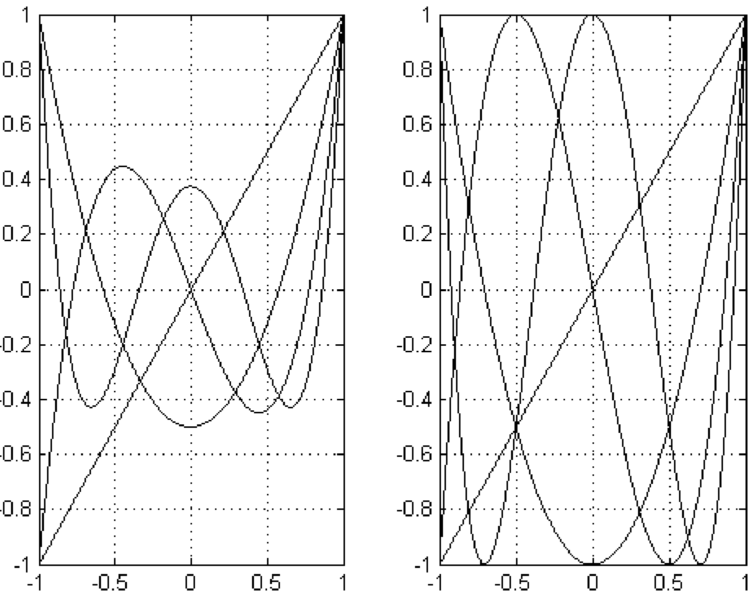

Рис. 2а. Полиномы Лежандра   Рис. 2б. Полиномы Чебышева