ЭКОНОМИКА
И МАТЕМАТИЧЕСКИЕ МЕТОДЫ

1971
том VII, вып. 3

СПЛАЙНЫ И ИХ ПРИМЕНЕНИЕ

С. А. СМОЛЯК

(Москва)

1. ОПРЕДЕЛЕНИЕ СПЛАЙНОВ

В данной статье дается общее определение понятия «сплайны», а также описываются многочисленные области их применения.

За рубежом исследованию сплайнов посвящено много работ, однако даваемые в них определения недостаточно общи, плохо переносятся на многомерный случай и, по-видимому, не могут быть применены в экономико-математических моделях. Чтобы прийти к наиболее точному определению сплайна, возьмем за основу проблему восстановления функций при отсутствии априорной информации. Ее наиболее типичная постановка такова: найти функцию $f(x)$, если известны только ее значения в точках $x = x_1, \ldots, x_N$. С точки зрения формально-математической эта задача некорректна, однако инженерам часто приходится проводить графики *неизвестных* функций по отдельным точкам, ничего больше об этих функциях не зная. Очевидно, в этих условиях неопределенности человеческий мозг руководствуется каким-то критерием, выдавая ответ в виде некоторой кривой. Можно сформулировать следующую весьма общую гипотезу о механизме решения данной задачи человеком *.

В ходе исторической эволюции человечества и данного человека в его мозгу сформировался некоторый (или некоторые) функционал J «хорошести» (плавности, гладкости) кривых $y = f(x)$, такой, что кривые, проходящие через данные точки, для которых значение $J(f)$ велико, отбрасываются, а те, для которых оно мало или минимально,— оставляются.

По-видимому, функционал $J(f)$ таков, что отбрасываемые кривые в человеческой практике не встречаются (или встречаются очень редко).

Вопрос о том, каким функционалом J пользуется *человек*, должен решаться путем *психологических* экспериментов, мы же рассмотрим математический аспект данного подхода, т. е. те объекты, которые получаются в результате решения задачи восстановления функций методом, моделирующим поведение человека. Именно эти объекты и будем называть сплайнами. Дадим им общее определение для случая восстановления элементов некоторого линейного топологического пространства M. Здесь роль значений функции в точке играют произвольные линейные непрерывные функционалы. Итак, *сплайном* $s = s(c_1, \ldots, c_N)$ *называется такой элемент линейного топологического пространства* M — $s \in M$, *для которого*

$$L_i(s) = c_i, \ i = 1, \ldots, N,$$
$$J(s) = \min. \tag{1}$$

Будем говорить, что неизвестный элемент s восстановлен по значениям линейных функционалов $L_1, \ldots, L_N$ на нем с помощью сплайна.

---

* Эта гипотеза излагалась Е. А. Александровым в докладе на IV симпозиуме по кибернетике (Тбилиси, 1968).

В принципе, восстанавливать элементы $s \in M$ можно и по значениям нелинейных фукционалов, но этот подход ввиду своей чрезмерной общности не представляет большого интереса. Рассмотрим частный случай такого восстановления, когда нелинейные функционалы просто выражаются через линейные. Подобные задачи имеют вид

$$F_k(L_1(s), \dots, L_n(s)) = 0, \; k = 1, \dots, m,$$
$$J(s) = \min, \tag{1'}$$

и если функции F_k достаточно «хорошие» (например, выпуклые), то их решение немногим сложнее решения задачи типа (1).

2. ПРИМЕРЫ СПЛАЙНОВ

Будем восстанавливать с помощью сплайнов функции одной или нескольких переменных, определенные во *всем* пространстве. Это условие непривычно для многих экономистов и специалистов по прикладной математике. Действительно, сторонники иной точки зрения обычно приводят в качестве основного довода то, что наблюденные значения аргументов лежат в определенных интервалах. Но одно дело использовать при восстановлении функции конкретные интервалы изменения аргументов, а другое — требовать, чтобы вне этих интервалов функция не могла быть продолжена и, значит, чтобы она была там не определена (для этого восстанавливаемая функция должна обладать особенностями на концах интервалов или, например, обращаться в бесконечность вне этих интервалов.) Случаи, когда из *теоретических* соображений известно, что восстанавливаемая функция определена не во всем пространстве, весьма редки и обычно сводятся заменой переменных к случаям, когда функция определена всюду. Кроме того, возможность различного выбора области определения привносит в получаемое решение элемент субъективизма.

Обратим внимание лишь на то, что, восстанавливая функцию во всем пространстве, мы делаем это в каждой точке с разной «надежностью». В точках, близких к наблюденным, значение функции восстанавливается обычно точнее, но мера этой точности меняется более или менее плавно от точки к точке, так что нет такой жесткой границы, до которой функция восстановлена достаточно точно, а вне ее пользоваться расчетными значениями уже нельзя.

Рассмотрим сначала восстановление функций одного переменного по значениям их в отдельных точках. Простейший результат состоит в том, что решением задачи

$$f(x_i) = f_i, \qquad i = 1, \dots, N, \tag{2}$$

$$J(f) = \int_{-\infty}^{\infty} \Phi(f'(x))\, dx = \min$$

для любой выпуклой функции $\Phi(z)$, удовлетворяющей условию $\Phi(0) = 0$, будет непрерывная ломаная линия, соединяющая точки (x_1, f_1), $(x_2, f_2), \dots, (x_N, f_N)$ при $x_1 < x_2 < \dots < x_N$, равная f_1 при $x < x_1$ и f_N при $x > x_N$. Таким образом, простейшими сплайнами являются ломаные линии. Чтобы получить более гладкие сплайны, надо использовать функционалы, содержащие производные высших порядков, например,

$$J(f) = \int_{-\infty}^{\infty} f''^2(x)\, dx. \tag{3}$$

Сплайн, минимизирующий этот функционал при условиях (2), оказывается, имеет очень интересную природу. Его формальное описание таково: на отрезках (x_i, x_{i+1}) функция $f(x)$ является многочленом третьей степени, а при $x > x_N$ и $x < x_1$ — многочленом первой степени. Эти многочлены склеены в точках $x = x_1, \ldots, x_N$ так, что получаемая кусочно-полиномиальная кривая имеет две непрерывные производные [1, 2].

Однако представляет интерес дать и неформальное описание полученного сплайна. Легко видеть, что $J(f)$ пропорционален энергии изгиба тонкого бесконечного стержня, прогиб которого в точке x равен $f(x)$. Поэтому тонкий бесконечный стержень, шарнирно закрепленный в точках (x_i, f_i), под действием внутренних сил изогнется сам, и кривая, определяющая его изгиб, и будет рассматриваемым нами сплайном. Таким приемом построения графиков с помощью гибкой рейки пользуются часто инженеры. Само слово «сплайн» происходит от названия гибкой рейки (spline), применяемой американскими кораблестроителями. Этот термин и описание кубических сплайнов были введены впервые, по-видимому, в [1].

В настоящее время имеется большое количество методов расчета таких сплайнов. Опишем наиболее понятный из них. Если уравнение сплайна на отрезке (x_i, x_{i+1}) есть

$$y = a_i + b_i x + c_i x^2 + d_i x^3, \quad i = 0, 1, \ldots, N^*,$$

то $4(N+1)$ величин a_i, b_i, c_i, d_i определяются из системы уравнений

$$a_i + b_i x_i + c_i x_i^2 + d_i x_i^3 = f_i, \qquad i = 1, \ldots, N,$$

$$a_{i-1} + b_{i-1} x_i + c_{i-1} x_i^2 + d_{i-1} x_i^3 = f_i, \qquad i = 1, \ldots, N,$$

$$b_{i-1} + 2c_{i-1} x_i + 3d_{i-1} x_i = b_i + 2c_i x_i + 3d_i x_i, \qquad i = 1, \ldots, N,$$

$$2c_{i-1} + 6d_{i-1} x_i = 2c_i + 6d_i x_i, \qquad i = 1, \ldots, N,$$

$$c_0 = 0,\ d_0 = 0,\ c_N = 0,\ d_N = 0.$$

Подобные системы уравнений легко решаются методом прогонки.

В [2—4] рассмотрены сплайны, минимизирующие функционалы вида

$$J(f) = \int |L_n(D) f(x)|^2 \, dx,$$

где $L_n(D)$ — произвольный полином степени n от оператора дифференцирования. Такие сплайны так же, как и рассмотренные, обладают достаточной гладкостью (имеют $2n-2$ непрерывные производные) и такие же «кусочные».

Перейдем к двумерному случаю. Разумным обобщением стержня является пластинка, так что соответствующие сплайны есть решение задачи** $f(x_i, y_i) = f_i$, $i = 1, \ldots, N$,

$$J(f) = \iint_{-\infty}^{\infty} |\Delta f(x, y)|^2 \, dx\, dy = \min. \tag{4}$$

Оказывается, имеет место следующее представление этого сплайна [5]

$$f(x, y) = A + Bx + Cy + \sum_{i=1}^{N} D_i h(x - x_i, y - y_i),$$

---

* Мы считаем $x_0 = -\infty$, $x_{N+1} = +\infty$.

** Чрезвычайно любопытно, что задача (4) сформулирована в [3], но пределы интегрирования в $J(f)$ не указаны, вследствие чего получено неверное решение.

где $h(u, v) = (u^2 + v^2) \ln (u^2 + v^2)$, а $N + 3$ неизвестных A, B, C, $D_1, \ldots$ $\ldots, D_N$ определяются из условий (4) и трех дополнительных условий

$$\sum_{i=1}^{N} D_i = 0, \quad \sum_{i=1}^{N} D_i x_i = 0, \quad \sum_{i=1}^{N} D_i y_i = 0. \qquad (5)$$

Механическая аналогия здесь также совершенно прозрачна — величины D_i есть реакции пластинки в точках (x_i, y_i), а тот факт, что сумма этих реакций и сумма моментов от них равна нулю, записан условиями (5). Величины же A, B, C отражают перенос и поворот пластинки в пространстве как твердого тела. Для практики важно, что (4) и (5) являются линейными уравнениями относительно определяющих сплайн параметров A, B, C и всех D_i.

В [2] предлагается рассматривать сплайны, минимизирующие при условиях типа (4) функционал $J(f) = \iint\limits_G \left| \frac{\partial^4 f}{\partial x^2 \partial y^2} \right|^2 dx\,dy$ с прямоугольной областью G (бикубичные сплайны). Так же, как и в одномерном случае, они оказываются кусочно-полиномиальными, а их построение осуществляется путем решения не очень сложной системы линейных уравнений. И тем не менее использование таких сплайнов в задачах *восстановления* функций не кажется нам достаточно разумным по следующим причинам:

1) функционал J не имеет физического смысла;

2) при фиксированных узлах интерполяции (x_i, y_i) область G должна быть вполне определенной и не допускает расширения, если сплайны строить в соответствии с [2];

3) решение задачи определяется неоднозначно, так как если $f(x, y)$ — решение, а $\varphi(x)$ — любая функция, удовлетворяющая условиям $\varphi(x_i) = 0$, $i = 1, \ldots, N$, то $f(x, y) + \varphi(x)$ тоже является решением $(J(f(x, y) + \varphi(x)) = J(f(x, y))$ и $f(x_i, y_i) + \varphi(x_i) = f(x_i, y_i))$;

4) решение задачи и минимальное значение функционала не непрерывно зависят от исходной информации.

Чтобы убедиться в последнем, заметим, что существуют такие N, x_1, y_1, $f_1, \ldots, x_N$, y_N, f_N, при которых для соответствующего сплайна $J(f) > 1$. Проварьируем теперь абсциссы и ординаты так, чтобы среди них не оказалось двух одинаковых, и построим произвольные гладкие функции $\varphi(x)$ и $\psi(y)$ так, чтобы для всех $i = 1, \ldots, N$ было $\varphi(x_i) = 0$, $\psi(y_i) = f_i$. Тогда $f(x, y) = \varphi(x) + \psi(y)$ будет удовлетворять условиям (4) и для нее будет $J(f) = 0$*.

По-видимому, ситуация будет аналогичной для всех сплайнов, построенных на основе гиперболических дифференциальных операторов.

В [2, 3, 6] предлагаются также специальные конструкции сплайнов, предназначенные для проведения плавных кривых на плоскости, проходящих через данные точки в данной последовательности. (По этому поводу см. также п. 4.)

---

* В [2] налагаются условия не только на значения функции в узлах, но и на значения производных. Это не спасает положения, так как в нашем примере достаточно потребовать, чтобы производные от $\varphi(x)$ или $\psi(y)$ приняли необходимые значения. В [3] получено разумное решение, но ценой добавления к $J(f)$ регуляризующих интегралов по контуру G.

3. ПРИМЕНЕНИЕ СПЛАЙНОВ В РЕГРЕССИОННОМ АНАЛИЗЕ

В большом количестве работ анализ корреляционных зависимостей между, например, экономическими показателями проводится на основе регрессионной модели $y = a_1 + a_2x + \ldots + a_nx^{n-1} + \xi$ или немного более общей модели $y = a_1\varphi_1(x) + \ldots + a_n\varphi_n(x) + \xi$, где ξ обычно считается нормально распределенной случайной величиной со средним значением ноль и неизвестным среднеквадратическим отклонением σ. Такого типа модели будем называть *полиномиальными*.

Применение полиномиальной модели к анализу некоторого явления оправдано прежде всего тогда, когда она теоретически обоснована исходя из существа этого явления, а коэффициенты $a_1, \ldots, a_N$ имеют «физический» смысл. В других случаях использование полиномиальных моделей может привести к серьезным ошибкам. Важно отметить, что эти модели не решают задачу о восстановлении неизвестной зависимости y от x, так как *вид* зависимости в них *известен*, он задан заранее, неизвестны лишь несколько числовых параметров, входящих в нее. Правильнее говорить об определении неизвестных параметров в известной зависимости.

Рассмотрим сложившуюся практику применения полиномиальных моделей при отсутствии априорной теоретической информации о виде уравнения регрессии. Стандартным оправданием для их применения является возможность приближения любой кривой регрессии полиномом с любой точностью. Остановимся на недостатках этих моделей.

Во-первых, не доказано и вряд ли может быть доказано, что любую (пусть даже достаточно простую по своим аналитическим свойствам) кривую регрессии, достоверность которой не вызывает сомнений, можно приблизить с достаточной точностью полиномом, коэффициенты которого, определяемые по выборке известными методами, были бы достоверными. По-видимому, вообще невозможно по выборке объемом в 1000 единиц восстановить достоверно кривую регрессии, если о ней совершенно точно известно, что она является полиномом, скажем, сотой степени.

Во-вторых, полиномиальные модели не позволяют использовать априорную информацию о величине влияния случайных факторов. Действительно, в этих моделях величина σ *определяется*, а не *задается*. Однако в экономических исследованиях бывают этапы, когда из большого перечня факторов отбираются те, которые сильно влияют на величину y, и отбрасываются те, влияние которых мало. При этом можно оценить суммарную величину влияния отбрасываемых факторов, т. е. величину σ. Эта информация в корреляционных моделях не только не используется, но наоборот, получаемое значение σ нередко противоречит такой предварительной оценке. В связи с этим важно отметить, что при повышении степени полинома $n \to \infty$ расчетная величина $\sigma \to 0$, в то время как на самом деле у нее должен быть предел, зависящий от величины влияния на переменную y прочих, не связанных с x, факторов. Обычно в экономике подобные факторы легко указываются. Например, при исследовании корреляционной зависимости себестоимости производства от объема выпуска предельное значение σ должно отражать влияние на себестоимость колебаний в количестве, структуре и возрасте применяемых производственных фондов, в качестве сырья и т. п. Ясно, что как бы ни была высока степень полинома, отражающего изучаемую зависимость, отклонения от этой зависимости будут всегда и притом в среднем *не меньше* определенной величины. Это обстоятельство также не может быть учтено в полиномиальных моделях.

В-третьих, выбор той или иной системы базисных функций $\varphi_1(x), \ldots, \varphi_n(x)$ в полиномиальных моделях является субъективным. Обычно исходят из простоты аналитической записи этих функций. Тем самым в основу как бы кладется гипотеза, что *простые явления отображаются простыми аналитическими формулами*. По существу, выбор базисных функций является самым слабым местом полиномиальных моделей. По-видимому, ни в одной практической задаче не удается грамотно обосновать выбор применяемой системы функций.

Избавиться от указанных недостатков могут помочь сплайны. Возьмем сначала простейший случай, когда случайные факторы отсутствуют, т. е. известно, что $\sigma=0$*. В этом случае исследуемая зависимость $y=f(x)$ удовлетворяет уравнению (2). Но если сторонники полиномиальных моделей стали бы подбирать *наиболее простую по своей аналитической записи* функцию $f(x)$ из всех, удовлетворяющих (2), то мы выберем *наиболее «гладкую», наиболее «хорошую»* из них в том смысле, о котором говорилось выше. Точнее, в качестве искомой зависимости мы выберем сплайн, являющийся решением задачи

$$\begin{gathered} f(x_i)=f_i, \\ J(f(x))=\min. \end{gathered} \tag{6}$$

При этом практически любой функционал J, ограничивающий (хотя бы в среднем) первые или вторые производные функции f, гарантирует нам отсутствие у решения каких-либо особенностей типа неожиданных выбросов или провалов, излишней колебательности и т. п., что почти всегда можно ожидать от любой полиномиальной модели.

Теперь уже ясно, что надо делать при наличии случайных факторов. В этом случае условие (2) надо заменить на условие *неточного* прохождения кривой $y=f(x)$ через заданные точки (x_i, f_i), причем среднеквадратичная ошибка этого прохождения должна быть равна σ. Таким образом, искомую кривую естественно искать как решение задачи

$$\begin{gathered} \frac{1}{N}\sum_{i=1}^{N}(f(x_i)-f_i)^2=\sigma^2, \\ J(f(x))=\min, \end{gathered} \tag{7}$$

являющейся обобщением (6) и частным случаем (1′).

Для решения (7) необходимо знать величину σ. Но именно это обстоятельство следует считать достоинством модели, так как оно позволяет учесть в расчете объективно существующий уровень влияния на переменную y случайных факторов. Тем самым появляется возможность, обработав одни и те же экспериментальные данные — N пар чисел (x_i, f_i), получить различные линии регрессии в зависимости от того, какую часть колеблемости переменной y мы приписываем влиянию x, а какую — прочим факторам. В одном случае мы будем иметь малое σ, хорошее совпадение расчета с экспериментом, но не слишком гладкую кривую, с большим $J(f)$. Такую кривую, поскольку она «на вид» мало правдоподобна, имеет смысл считать решением задачи, только если обосновано принятое в расчете малое σ. В другом случае (σ велико) кривая будет достаточно хорошей, но будет не очень хорошее совпадение расчета и эксперимента. Если влияние неучитываемых факторов действительно велико, то такая кривая является самой гладкой и потому самой правдоподобной, самой приемлемой из всех возможных.

---

* В этом случае мы приходим к задаче интерполяции и экстраполяции. (См. также п. 5).

Дальнейшим обобщением (7) является задача, в которой экспериментальные значения f_i неравноточны и их дисперсии σ_i^2 различны

$$\sum_{i=1}^{N}\left(\frac{f(x_i)-f_i}{\sigma_i}\right)^2 = N, \tag{8}$$
$$J(f(x)) = \min.$$

Задача (8) изучалась в [5, 7], где дана также алгол-процедура решения для функционала типа (3). Задачи (7) и (8) дают хороший пример восстановления функции по значениям нелинейных функционалов от нее.

Остановимся на вопросе — откуда взять фукционал $J(f)$. Сторонники полиномиальных моделей обычно пользуются несколькими семействами функций $\varphi_i(x)$, в результате применения которых получаются функции $f(x)$, не имеющие каких-либо разрывов или иных особенностей и не слишком «колеблющиеся». При использовании же сплайнов желаемое поведение решения прямо отражается в формуле функционала J.

Ниже даются некоторые рекомендации по применению отдельных функционалов $J(f)$ в практических задачах восстановления, полученные на основе опыта ряда авторов.

1. Если зависимость не является плавной кривой, то можно использовать функционал $J(f) = \int f'^2(x)\,dx$. Решение получится в виде ломаной, нахождение которой требует решения простой системы линейных уравнений.

2. Если решение должно быть монотонной, но не обязательно гладкой кривой, можно применять функционал «энтропии» $J(f) = \int f'(x)\ln|f'(x)|\,dx$. Решением также является ломаная.

3. Если решение должно быть гладкой плавной кривой, то применим функционал $J(f) = \int f''^2(x)\,dx$. Решение получается в виде кубичного сплайна, параметры которого определяются из системы линейных уравнений.

4. Если решение должно быть плавной монотонной кривой, вероятно, можно использовать функционалы

$$J(f) = \int \frac{f''^2(x)}{f'^n(x)}\,dx.$$

В частности, при $n = 2$

$$J(f) = \int \left(\frac{d}{dx}\ln f'(x)\right)^2 dx,$$

а при $n = {}^3/_2$ значение $J(f)$ для функции $y = f(x)$ и для обратной функции $x = f^{-1}(y)$ одинаково. С помощью такого функционала, вероятно, удобно восстанавливать интегральные кривые распределения или кривые накопленных частот. Такие сплайны, если окажется достаточно простая программа их построения, можно будет применять для моделирования случайных величин с заданными законами распределения.

5. При восстановлении многофакторных зависимостей $y = f(x^1, \ldots, x^s)$ нужно использовать функционалы от функций нескольких переменных. Единственные функционалы, для которых удается получить приемлемые решения в не слишком сложной аналитической форме, это функционалы вида

$$J(f) = \int_{-\infty}^{\infty}\!\cdots\!\int |\Delta^k f(x^1, \ldots, x^s)|^2\,dx^1 \ldots dx^s,$$

где Δ — оператор Лапласа, $k > s/2$. При этом практически можно использовать лишь значение $k = \left[\frac{s+3}{2}\right]$. Нахождение соответствующих сплайнов снова сводится к решению систем линейных алгебраических уравнений.

6. Если возникает необходимость использовать отличные от приведенных функционалы $J(f)$, рекомендуется пользоваться функционалами, обладающими следующими свойствами, обеспечивающими инвариантность восстанавливаемой зависимости относительно выбора начала координат и масштаба по осям: 1) $J(f(x^1+a^1, x^2+a^2, \ldots, x^s+a^s)) = J(f(x^1, \ldots, x^s))$; 2) $J(f(\lambda x^1, \ldots, \lambda x^s)) = |\lambda|^p J(f(x^1, \ldots, x^s))$; 3) $J(\lambda f) = |\lambda|^q J(f)$, $q > 0$. Чтобы обеспечить единственность восстановления, достаточно потребовать выпуклости функционала J. Задачи типа (7) или (8) при этом становятся задачами отыскания минимума выпуклого функционала на границе выпуклого множества.

В данной статье не случайно рассматривается случай, когда величина σ известна точно. Решение задачи при наличии двусторонних ограничений на σ (см. [7]) не дает ничего нового и приводит к сплайнам уже рассмотренного типа. Действительно, для любых выпуклых функционалов J решение задачи

$$\sigma_1^2 \leqslant S(f) = \frac{1}{N}\sum_{i=1}^{N}(f(x_i) - f_i)^2 \leqslant \sigma_2^2,$$

$$J(f) = \min,$$

если оно существует, совпадает с решением либо задачи (7) при $\sigma = \sigma_2$ (обычно это бывает, когда с увеличением ошибки σ решение становится более гладким), либо этой же задачи при $\sigma = \sigma_1$, либо задачи безусловной минимизации $J(f)$. В последнем случае решение обычно не единственно (например, для функционалов (3) решением является любая прямая линия), но среди возможных решений f имеются и удовлетворяющие неравенству $\sigma_1^2 \leqslant S(f) \geqslant \sigma_2^2$. Практически это означает, что перед восстановлением зависимости с помощью сплайна надо проверить, не даст ли требуемую точность наиболее гладкая из возможных кривых (например, для функционала (3) — наименее уклоняющаяся от данных точек прямая).

4. ПРИМЕНЕНИЕ СПЛАЙНОВ В ФАКТОРНОМ АНАЛИЗЕ

В регрессионном анализе переменные y и x неравноправны — направление причинной связи должно быть указано заранее. Однако в практических задачах нередко бывает неизвестно, какой фактор от какого зависит. В такой ситуации задачу описания взаимосвязи переменных можно понимать двояко: 1) указать такой вид зависимости между факторами $x^1, x^2, \ldots, x^s$, в который эти факторы входили бы равноправно; 2) найти такие (или такой) факторы $u^1, \ldots, u^r$, которые с достаточной точностью объясняли бы изменение факторов $x^1, \ldots, x^s$, или, что то же, такие факторы $u^1, \ldots, u^r$, что $x^1, \ldots, x^s$ можно было бы считать зависящими практически только от них.

Легко видеть, что оба подхода эквивалентны отысканию s зависимостей

$$x^k \approx f_k(u^1, \ldots, u^r), \quad k = 1, \ldots, s \tag{9}$$

на основе выборочных значений x_i^k. Иными словами, исходными данными для установления зависимостей между факторами x^k являются точные или приближенные равенства

$$x_i^k \approx f_k(u_i^1, \ldots, u_i^r), \quad k = 1, \ldots, s, \; i = 1, \ldots, N, \tag{10}$$

где неизвестны не только функции f_k, но и значения факторов u_i^j.

«Классический» факторный анализ [8] обходит эти трудности так: функции f_k предполагаются линейными; факторы u^j считаются независимыми нормально распределенными случайными величинами с нулевыми средними и единичными дисперсиями.

Ясно, что первое условие эквивалентно требованиям полиномиальных моделей, оно введено лишь *для упрощения* и неадекватно действительности. А вот какое-то условие нормировки для факторов u^j необходимо по существу, ибо одни только условия (9) и (10) не позволят выбрать в качестве наиболее приемлемой *одну* из зависимостей: $x^k \approx f(u^1, \ldots, u^r)$; $x^k \approx \approx g(2u^1, \ldots, 2u^r)$; $x^k \approx h(u^1 + 1, \ldots, u^r + 1)$.

Все они будут одинаково приемлемы.

Нам представляется разумным использовать в подобных задачах приведенное выше условие, которое удобно записать в виде

$$\frac{1}{N} \sum_{i=1}^{N} |u_i^j|^2 = 1, \; j = 1, \ldots, r, \tag{11}$$

или простое ограничение типа

$$0 \leqslant u_i^j \leqslant 1, \; i = 1, \ldots, N; \; j = 1, \ldots, r. \tag{12}$$

Вернемся к условию линейности функций f_k и постараемся заменить его более реалистическим. Рассмотрим, как «классический» факторный анализ проверяет адекватность своих моделей. С этой целью рекомендуется определить закон распределения вектора $(f_1(u^1, \ldots, u^r), \ldots, f_s(u^1, \ldots, u^r))$ — он должен быть достаточно близок к закону распределения вектора $(x^1, \ldots \ldots, x^s)$. Исходя из этого, следует определить количество r неизвестных факторов. Мы же можем поступить наоборот — *задаться количеством факторов* r и степенью близости законов распределения правой и левой части в (10), а *определять вид функций* f_k, удовлетворяющий этим условиям.

Это приводит к задаче типа

$$\begin{gathered} J(f_1, \ldots, f_s) = \min, \\ \frac{1}{N} \sum_{i=1}^{N} |x_i^k - f_k(u_i^1, \ldots, u_i^r)|^2 = \sigma_k^2, \quad k = 1, \ldots, s, \\ \frac{1}{N} \sum_{i=1}^{N} |u_i^j|^2 = 1, \quad j = 1, \ldots, r. \end{gathered} \tag{13}$$

Здесь последним написано условие нормировки (11), число σ_k отражает желаемую степень «достаточной близости» правой и левой части в (10), а $J(f_1, \ldots, f_s)$ есть функционал гладкости *системы* функций $(f_1, \ldots, f_s)$. (Его можно «конструировать» из функционалов гладкости одной функции, например, по формуле

$$J(f_1 \ldots, f_s) = J(f_1) + \ldots + J(f_s).)$$

Постановка типа (13) кажется нам более разумной, чем постановка «классического» факторного анализа. К ней можно прийти и иначе — отыскивая уравнение линии регрессии при наличии ошибок в измерениях (нечто вроде нелинейного варианта задачи Вальда). Обычно эта задача ставится так: найти $f(x)$, если известны с ошибкой выборочные значения x и соответствующих y

$$u_i = x_i + \xi_i,$$
$$v_i = f(x_i) + \eta_i = y_i + \eta_i.$$

Эта постановка, однако, неинвариантна относительно того, какое переменное считать независимым.

Если же кривую $y = f(x)$ задать в параметрической форме, то придем к уравнениям типа $u_i = \varphi(t_i) + \xi_i$; $v_i = \psi(t_i) + \eta_i$. Если задать дисперсию ошибок ξ и η, а также нормировать фактор t, то можно найти φ и ψ как решение задачи $J(\varphi, \psi) = \min$,

$$\frac{1}{N}\sum_{i=1}^{N}(u_i - \varphi(t_i))^2 = \sigma_\xi^2,$$

$$\frac{1}{N}\sum_{i=1}^{N}(v_i - \psi(t_i))^2 = \sigma_\eta^2,$$

$$\frac{1}{N}\sum_{i=1}^{N} t_i^2 = 1.$$

5. СПЛАЙНЫ И ТЕОРИЯ ПРИБЛИЖЕНИЙ

Между сплайнами и теорией приближений существует двоякая связь. С одной стороны, сплайны можно использовать для того, чтобы приближать ими известные функции, а с другой — они оказываются тесно связанными с решением некоторых задач наилучшего приближения. Для того, чтобы продемонстрировать эти связи, используем понятие поперечника. Обычно в задачах приближения функции f из некоторого класса K приближаются линейными комбинациями некоторых фиксированных n элементов пространства M, т. е. точками линейного подпространства $P_n \subset M$ *размерности* n. В других задачах эти функции приближаются такими элементами M, на которые наложено n линейных однородных связей (например, заданы нулевые значения функций в некоторых n точках или потребована ортогональность их к n заданным функциям), т. е. точками линейного подпространства $P_{-n} \subset M$ *коразмерности* n. *Точность приближения* в линейных топологических пространствах удобно определять с помощью центрально-симметричного выпуклого тела B, считая, что «расстояние» между элементами x и y

$$\rho(x, y) = \inf_{\lambda \geqslant 0\ x - y \in \lambda B} \lambda$$

(такое «расстояние» может быть и бесконечным, если соответствующих λ не существует). Поэтому точность самого лучшего способа приближения элементов K при самом лучшем выборе линейного подпространства может быть охарактеризована величиной

$$d_{\pm n}(K, B) = \inf_{P_{\pm n}} \sup_{y \in K} \inf_{x \in P_{\pm n}} \rho(x, y) = \inf_{P_{\pm n}} \inf_{\substack{\lambda \geqslant 0 \\ K \subset \lambda B + P_{\pm n}}} \lambda.$$

Эта величина и называется $\pm n$-мерным поперечником (в «метрике» B) множества K (подробнее см. [9]).

Сплайны, минимизирующие однородные выпуклые функционалы J, тесно связаны с поперечниками d_{-n}. Действительно, если принять за «единичный шар метрики» $B = \{f \in M\colon J(f) \leqslant 1\}$ и восстанавливать элементы

$f \in K$ по значениям n линейных функционалов с помощью сплайнов s

$$L_i(s) + L_i(f), \qquad J(s) = \min, \tag{14}$$

то *ошибка восстановления* $\varphi = f - s$ будет решением задачи наилучшего приближения (в «метрике» B) f точкой гиперплоскости P_{-n}: $L_i(\varphi) = 0$, $i = 1, \ldots, n$, коразмерности n. Поэтому гладкость самого негладкого сплайна из всех, порожденных элементами $f \in K$, будет равна уклонению множества K от гиперплоскости P_{-n}. Отыскание же набора n линейных функционалов (линейно независимых!), наиболее пригодных для восстановления элементов K (т. е. дающих при восстановлении *самые гладкие* результаты), эквивалентно отысканию гиперплоскости P_{-n}, наименее уклоняющейся от K, т. е. отысканию поперечника $d_{-n}(K, B)$ и той гиперплоскости P_{-n}, на которой этот поперечник достигается.

Для установления связи между сплайнами и поперечниками d_n используем сплайн (14) для приближения элемента $f \in K$. Чтобы убедиться, насколько хорош такой способ приближения, надо оценить его ошибку и сравнить ее с n-мерным поперечником K.

В качестве примера классов K рассмотрим множества $W_\alpha(K)$ $(\widetilde{W}_\alpha(K))$ α раз дифференцируемых функций $f(x)$, определенных на $(0,1)$ (имеющих период 1), для которых

$$\| f^{(\alpha)}(x) \|^2_{L_2} = \int_0^1 | f^{(\alpha)}(x) |^2 \, dx \leqslant K^2.$$

Поперечник $\widetilde{W}_\alpha(K)$ в метрике L_2 был вычислен А. Н. Колмогоровым (см. также в [9]), а в метрике C оценивался, например, в [10]. Из указанных работ вытекает, что

$$d_n(W_\alpha(K), L_2) \geqslant d_n(\widetilde{W}_\alpha(K), L_2) > C_1 \frac{K}{n^\alpha}, \tag{15}$$

$$d_n(W_\alpha(K), C) \geqslant d_n(\widetilde{W}_\alpha(K), C) > C_2 \frac{K}{n^{\alpha-0,5}},$$

где C_1, C_2 — абсолютные постоянные.

Функции $f(x)$ этих классов будем приближать сплайнами $s(x)$, определяемыми из условий

$$s(x_i) = f(x_i), \quad i = 1, \ldots, n,$$

$$J(s) = \int_0^1 | a_0(x) s^{(r)}(x) + \ldots + a_{r-1}(x) s'(x) + a_r(x) s(x) |^2 \, dx = \min,$$

или периодическими сплайнами $\tilde{s}(x)$, удовлетворяющими этим же условиям. В [4] показано, что при естественных ограничениях на $a_i(x)$ и при достаточно регулярном расположении узлов интерполяции x_i ($\max\{x_2 - x_1, x_3 - x_2, \ldots, x_n - x_{n-1}, x_1 + 1 - x_n\} \leqslant \frac{c_3}{n}$) точность приближения сплайнами $s(x)$ функций из $W_r(K)$ будет

$$\| f(x) - s(x) \|_{L_2} \leqslant C_4 \frac{K}{n^r} \| f(x) - s(x) \|_C \leqslant C_5 \frac{K}{n^{r-0,5}}.$$

Для точности приближения периодическими сплайнами $\tilde{s}(x)$ функций из $\widetilde{W}_{2r}(K)$ в [4] получены оценки

$$\| f(x) - s(x) \|_{L_2} \leqslant C_6 \frac{K}{n^{2r}}, \quad \| f(x) - s(x) \|_C \leqslant C_7 \frac{K}{n^{2r-0,5}}.$$

Сопоставляя эти оценки с неравенствами (15), можно увидеть, что рассматриваемый способ приближения является по порядку точности оптимальным. В данном примере сплайны давали приближение, оптимальное только по порядку. В [11] исследовались наилучшие приближения на других классах функций, при этом было показано, что приближение с помощью специально устроенных кусочно-полиномиальных сплайнов позволяет на этих классах обеспечить наименьшую возможную ошибку, в точности равную поперечнику этих классов в метрике C.

Из [2, 4] видно, что при данном способе аппроксимации не только сама функция, но и ее производные будут приближены с оптимальной по порядку точностью.

Изложенную процедуру аппроксимации с большим эффектом применяют при решении уравнений теории упругости и строительной механики («метод конечных элементов»), аппроксимируя сплайнами линии (поверхности) прогиба. Поскольку данный метод является разновидностью метода Ритца, он сводит задачу к решению конечной системы уравнений, но использование сплайнов обеспечивает бо́льшую точность решения ввиду того, что реальные прогибы по своему характеру ближе к сплайнам, чем к полиномам.

Если известно асимптотическое поведение функции $f(x)$ вблизи границ ее области определения, то точность аппроксимации можно существенно повысить, применив подходящие функциональные преобразования по координатным осям, т. е. используя сплайн для представления зависимости переменной $H(f)$ от переменной $G(x)$.

В [12, 13] применен другой метод аппроксимации, при котором известную функцию $f(x)$, заданную на густой сетке точек x_i, аппроксимируют кубичным (в двумерном случае — бикубичным) сплайном $s(x)$, построенным по значениям $s(\xi_j) = s_j$ в фиксированных равноотстоящих узлах ξ_j, значительно более редких. Величины s_j неизвестны и подбираются из условия наилучшей аппроксимации

$$\sum_i (f(x_i) - s(x_i))^2 = \min.$$

Решение этой задачи сводится к решению системы линейных уравнений и позволяет задавать довольно сложные линии и поверхности в кусочно-полиномиальном виде (к тому же с полиномами невысокой степени), что очень удобно для использования в машиностроении и строительстве.

Из приведенных примеров видно, что в ряде интересных для практики случаев сплайны с успехом выступают как простой и эффективный аппарат для получения приближений функций.

ЛИТЕРАТУРА

1. I. J. Shoenberg. Contribution to the Problem of Approximation of Equidistant Data by Analytic Functions. Quar. Appl. Math., 1946, v. 4.
2. J. H. Ahlberg, E. N. Nilson, J. H. Walsch. The Theory of Splines and Their Applications. Academic Press, 1967.
3. G. Birkhoff, C. R. de Boor. Piecewise Polynomial Interpolation and Approximation. In: Approximation of Functions. Elsevier, 1965.
4. M. H. Schultz, R. S. Warga. L — Splines. Numerische Mathematik, 1967, v. 10, N 4.
5. С. А. Смоляк. Оптимальное восстановление функций и связанные с ним геометрические характеристики множеств. В сб. Труды третьей зимней школы по математическому программированию и смежным вопросам. Т. 3. М., 1971 (ЦЭМИ АН СССР).
6. A. H. Fowler, C. W. Wilson. Cubic Spline, a Cure Fitting Routine. Report Y — 1400. Oar Ridge, 1963.

7. C. H. Reinsch. Smoothing by Spline Functions. Numerische Mathematik, 1967, v. 10, N 3.
8. Д. Н. Лоули, А. Э. Максвелл. Факторный анализ как статистический метод. М., «Мир», 1967.
9. А. Д. Иоффе, В. М. Тихомиров. Двойственность выпуклых функций и экстремальные задачи. Успехи матем. наук, 1968, т. 23, вып. 6.
10. И. Ф. Шарыгин. Оценки снизу в теории интегрирования и приближения на классах функций. Канд. дисс., М., 1965 (Моск. гос. ун-т).
11. В. М. Тихомиров. Наилучшие методы приближения и интерполирования дифференцируемых функций в пространстве C [—1,1]. Матем. сб., 1969, т. 80 (122).
12. А. Д. Добыш. Кусочно-полиномиальная аппроксимация гладких кривых и поверхностей. В сб. XXIX научно-техническая конференция. Тезисы докладов, аннотации. Секция физико-математических наук. М., 1970 (Моск. инженерно-строит. ин-т им. В. В. Куйбышева).
13. А. Д. Добыш. Построение интерполирующих кусочно-полиномиальных функций. В сб. Труды третьей зимней школы по математическому программированию и смежным вопросам, Т. 2. М., 1971, (ЦЭМИ).

Поступила в редакцию
27 VII 1970