УДК 550.8.05:519.652

А. С. Костинский

О принципах сплайн-экстраполяции геофизических данных

(Представлено академиком НАН Украины В. И. Старостенко)

Возможные приложения сплайновой математики обсуждаются применительно к геофизическим наблюдениям, когда построить физическую динамическую модель либо невозможно, либо слишком сложно, нерационально. В подобных ситуациях простая идея сплайн-экстраполяции оказывается единственной: сетка узлов на заданном сегменте дополняется прогнозируемой точкой, строится "прогностический" сплайн на расширенной сетке, необходимо обеспечить минимум интеграла квадратичного отклонения, зависящего от ординаты добавочной точки как от параметра. Для равномерной сетки структурные единицы алгоритма экстраполяции представляются в виде последовательности разложений по координатам заданных точек, коэффициенты разложений доступны аналитически. Показано, что ордината прогнозируемой точки не зависит от шага сетки, это существенно для оценки ближайшего следующего в серии регулярных измерений, когда принципиальна не величина интервала между измерениями, а его неизменность.

Логика сплайн-экстраполяции, если речь идет о приближении известной, "физически разумной" функции, в основном сводится к следующему. Пусть на отрезке $[a, b]$ в узлах сетки $\Delta\colon a = x_0 < x_1 < \cdots < x_N = b$ заданы значения $y_i = f(x_i)$, $i = 0, \ldots, N$, а мы хотим оценить функцию f в некоторой точке $x = x_c$ справа от $[a, b]$, $x_c > x_N$. Обозначим через $y_c^{(e)}$ искомое значение переменной ординаты y_c, то число, которое мы стремимся получить и которое, надо полагать, будет близко к $f(x_c)$. Построим, например, кубический интерполяционный сплайн $S^{(N)}(f; x)$ на сетке Δ [1]:

$$\begin{gathered}S^{(N)}(f;x) = y_i(1-t)^2(1+2t) + y_{i+1}t^2(3-2t) + m_i h_i t(1-t)^2 - m_{i+1}h_i t^2(1-t),\\ x \in [x_i, x_{i+1}], \qquad t = \frac{(x - x_i)}{h_i}, \qquad h_i = x_{i+1} - x_i, \qquad i = 0, \ldots, N-1,\end{gathered} \tag{1}$$

$$\begin{gathered}S^{(N)}(f;x)\big|_{x=x_i} = y_i, \qquad \frac{d}{dx}S^{(N)}(f;x)\bigg|_{x=x_i} = m_i, \qquad i = 0, \ldots, N,\\ S^{(N)}(f;x) \in C^2[a,b],\end{gathered} \tag{2}$$

а затем, временно соглашаясь считать точку (x_c, y_c) известной, аналогичный сплайн $S^{(N+1)}(f; x)$ на расширенной сетке $\delta\colon a = x_0 < x_1 < \cdots < x_N < x_{N+1} = x_c$. Дополнительное условие в узле $x = x_c$ —

$$S^{(N+1)}(f;x)\big|_{x=x_c} = y_c,$$

в силу нелокального характера сплайна приводит к тому, что $S^{(N+1)}$ не совпадает с $S^{(N)}$, различие, если не касаться частностей, "затухает" по мере смещения влево: от точки $x = x_b$

к точке $x = x_a$. "Расстояние" между $S^{(N+1)}$ и $S^{(N)}$ на $[a, b]$ измеряется, строго говоря, тремя интегралами квадратичных отклонений

$$J_e^{(k)} = \int\limits_a^b \left| \frac{d^k}{dx^k} S^{(N+1)}(f; x) - \frac{d^k}{dx^k} S^{(N)}(f; x) \right|^2 dx, \qquad k = 0, 1, 2,$$

зависящими от y_c как от параметра, для определенности договоримся рассматривать пока только $J_e^{(0)} \equiv I_e$.

Не будем пока касаться вопроса о краевых условиях, принимая их во внимание, мы, быть может, только добавим варьируемые параметры или усложним пробную функцию I_e. Существенно, что при любом выборе краевых условий для равномерных сеток последовательность интерполяционных кубических сплайнов всегда сходится к интерполируемой непрерывной функции. В общем случае произвольного разбиения существуют возможности интерполяции с регулируемой точностью [1], и можно утверждать поэтому, что сплайн $S^{(N)}$ на достаточно подробной сетке в некотором смысле близок функции $f(x)$. Свойства f на $[a, b]$ заимствуются "реперной" интерполирующей функцией $S^{(N)}$ и воспроизводятся (но с дополнительным искажением) интерполирующей функцией $S^{(N+1)}$. Качество интерполяции посредством $S^{(N+1)}$ на $[a, b]$ почти наверное хуже из-за, возможно, слишком крупного шага $h_N = x_c - x_N$ и неопределенной ординаты y_c. Минимум I_e по y_c, таким образом, устанавливает естественный нижний предел сходства $S^{(N+1)}$ с аппроксимацией $S^{(N)}$, и это обстоятельство уже само по себе достаточное основание для того, чтобы определить точку минимума как "пробное" экстраполирующее значение $y_c^{(e)}$ и исследовать разность $f(x_c) - y_c^{(e)}$ в различных вариантах конструкций $S^{(N)}$ и $S^{(N+1)}$.

Общий замысел решения задачи экстраполяции в смысле поиска минимума интеграла I_e заключается в разложении сплайна $S^{(N+1)}$ по системе фундаментальных сплайнов на сетке δ (возможно, конечно, использование B-сплайнов с конечным носителем [1]). Это позволяет записать в явном виде простую систему линейных уравнений относительно параметров и ее аналитическое решение, избежав вычисления собственных значений матриц численными методами в ситуации, когда достоверно не подтверждена невырожденность. Соответственно краевые условия для $S^{(N+1)}$ диктуют характер разложения по базису и прогнозируемые параметры в точке $x = x_c$.

Простейший вариант — прогноз только по ординате — возникает, если потребовать непрерывности третьей производной $S^{(N+1)}$ в узлах x_1, x_N [1]. Мы имеем одну варьируемую переменную $\zeta = y_c$, сплайн $S^{(N+1)}$ в этом случае представляется в виде линейной комбинации

$$S^{(N+1)} = \sum_{i=0}^{N} y_i F_i(x) + y_c F_{N+1}(x), \tag{3}$$

где $F_i(x)$, $i = 0, \ldots, N, N+1$, — фундаментальные сплайны на сетке δ, удовлетворяющие условию нормализации

$$\sum_{i=0}^{N+1} F_i(x) = 1, \qquad x \in [a, x_c],$$

условиям интерполяции вида

$$F_i(x_j) = \delta_{ij}, \qquad i = 0, \ldots, N, N+1, \qquad j = 0, \ldots, N, N+1,$$

и условиям непрерывности третьей производной в узлах x_1, x_N.

Пусть два произвольных сплайна $S^{(A)}$, $S^{(B)}$, записанных в виде (1), с коэффициентами $m_i^{(A)}$, $m_i^{(B)}$, $i = 0, \ldots, N$, интерполируют нулевые значения на сетке Δ. Интеграл по отрезку $[a, b]$ от произведения $S^{(A)} S^{(B)}$ выражается суммой

$$\frac{1}{105} \sum_{i=0}^{N-1} h_i^3 \{ m_i^{(A)} m_i^{(B)} + m_{i+1}^{(A)} m_{i+1}^{(B)} - \frac{3}{4} (m_i^{(A)} m_{i+1}^{(B)} + m_{i+1}^{(A)} m_i^{(B)}) \},$$

это — "вычислительный шлюз", ведущий к аналитическому представлению минимума. Интегралы такого рода появляются, если подставить разложение (3) в выражение для I_e,

$$I_e = \alpha_{00} + 2\alpha_{01}\zeta + \alpha_{11}\zeta^2,$$

коэффициент α_{ij}, $0 \leqslant i, j \leqslant 1$, есть интеграл по отрезку $[a, b]$ от произведения вспомогательных сплайнов, их два, с ними можно связать числа 0,1:

$$\sum_{i=0}^{N} y_i F_i(x) - S^{(N)}(x) \to 0, \qquad F_{N+1}(x) \to 1.$$

Обсудим краевые условия для сплайна с номером 0. В первую очередь следует договориться о выборе краевых условий для сплайна $S^{(N)}$, это вопрос существенный, если преследуется цель аналитически оценить зависимость прогнозируемой ординаты от характерного масштаба задачи, доминирующего шага "основной" сетки Δ. Пусть предпочтение отдано естественным, наиболее простым по смыслу условиям

$$\frac{d}{dx} S^{(N)}(f; x) \bigg|_{x=a} = z'_a, \qquad \frac{d}{dx} S^{(N)}(f; x) \bigg|_{x=b} = z'_b,$$

где числовые значения z'_a, z'_b получаются, например, конечно-разностной аппроксимацией или как результат дополнительного "сглаживания": $S^{(N)}$ должен в некотором смысле "наименее отличаться" от полинома второй или третьей степени [2].

Задача построения сплайна $S^{(N)}$ разрешима для любых вещественных z'_a, z'_b, будем пока считать z'_a, z'_b известными параметрами. Затем, без ограничения общности, можно выбрать сетку Δ равномерной с шагом h и записать $x_c - x_N = h h_c$, безразмерный множитель h_c исчисляется в долях h. Для примера, регулярные измерения химического состава, уровня и температуры грунтовых вод, концентрации радона в скважинах, в сущности, не предполагают точного знания времени измерения. Не подвергается сомнению только то, что промежуток времени между последовательными измерениями одинаков.

Обозначим через D_k, $k = 0, \ldots, N$, разрывы третьей производной сплайна $S^{(N)}$ в узлах равномерной сетки Δ [1]. Потребуем, чтобы сплайн $S^{(N)}$ доставлял минимум по переменным z'_a, z'_b сумме

$$\sum_{k=0}^{N} D_k^2, \tag{4}$$

иначе говоря, $S^{(N)}$ в смысле гладкости выбирается как наименее отличающийся от полинома третьей степени. Представим $S^{(N)}$ в виде [2]:

$$S^{(N)} = \sum_{i=0}^{N} y_i F_i(x) + z'_a F_a(x) + z'_b F_b(x), \tag{5}$$

где $F_i(x)$, $i = 0, \dots, N$, — фундаментальные сплайны на сетке Δ, удовлетворяющие условиям интерполяции вида

$$F_i(x_j) = \delta_{ij}, \qquad i = 0, \dots, N, \qquad j = 0, \dots, N,$$

и краевым условиям

$$F'_i(a) = F'_i(b) = 0,$$

появляются дополнительные базисные функции, подчиненные условиям

$$F_a(x_i) = F_b(x_i) = 0, \qquad i = 0, \dots, N, \qquad F'_a(a) = F'_b(b) = 1, \qquad F'_a(b) = F'_b(a) = 0.$$

Из уравнения (5) следует выражение через разрывы базисных сплайнов

$$D_k(S^{(N)}) = \sum_{i=0}^{N} y_i D_k(F_i) + z'_a D_k(F_a) + z'_b D_k(F_b), \qquad k = 0, \dots, N, \tag{6}$$

подставляя его в (4), имеем квадратичную форму по z'_a, z'_b. Условие минимума формы дает систему линейных уравнений, из которой, в свою очередь, получаем "оптимальные" краевые значения первых производных как линейные функции ординат y_i. Вычисляем коэффициенты линейного разложения

$$z'_a = \sum_{i=0}^{N} p_i y_i, \qquad z'_b = \sum_{i=0}^{N} q_i y_i, \tag{7}$$

$$p_i = \frac{1}{\Omega} \sum_{k=0}^{N} D_k(F_i)[\omega_{12} D_k(F_b) - \omega_{22} D_k(F_a)],$$

$$q_i = \frac{1}{\Omega} \sum_{k=0}^{N} D_k(F_i)[\omega_{12} D_k(F_a) - \omega_{11} D_k(F_b)], \qquad i = 0, \dots, N,$$

упрощает ситуацию свойство $q_i = -p_{N-i}$. Элементы матрицы системы, обозначенные как ω_{ij}, зависят только от разрывов базисных функций $F_a(x)$, $F_b(x)$,

$$\omega_{11} = \sum_{k=0}^{N} D_k(F_a)^2, \qquad \omega_{22} = \sum_{k=0}^{N} D_k(F_b)^2, \qquad \omega_{12} = \omega_{21} = \sum_{k=0}^{N} D_k(F_a) D_k(F_b),$$

все это достаточно громоздкие выражения. Как промежуточный итог выделим зависимость от шага сетки

$$\omega_{11}, \omega_{22}, \omega_{12} \sim \frac{1}{h^4}, \qquad \Omega = \mathrm{Det}(\omega_{ij}) = \omega_{11}\omega_{22} - \omega_{12}^2 \sim \frac{1}{h^8}, \qquad p_i, q_i \sim \frac{1}{h}.$$

Из выражений (5) и (7) следует линейное разложение коэффициента:

$$m_1^{(S^{(N)})} = \sum_{i=0}^{N} y_i\{m_1(F_i) + p_i m_1(F_a) + q_i m_1(F_b)\}$$

и аналогично для коэффициента

$$m_{N-1}^{(S^{(N)})} = \sum_{i=0}^{N} y_i\{m_{N-1}(F_i) + p_i m_{N-1}(F_a) + q_i m_{N-1}(F_b)\},$$

что, в свою очередь, дает возможность линейного разложения по ординатам параметров Z'_a, Z'_b в соотношениях

$$m_0^{(0)} + 2m_1^{(0)} + Z'_a = 0,$$

$$h_c m_{N-1}^{(0)} + (1 + h_c)(m_N^{(0)} + Z'_b) = 0,$$

$$Z'_a \equiv z'_a + 2m_1^{(S^{(N)})} - \frac{1}{h}(2y_1 - \frac{5}{2}y_0 + \frac{1}{2}y_2),$$

$$Z'_b \equiv \frac{h_c}{1 + h_c} m_{N-1}^{(S^{(N)})} + z'_b + \frac{1}{hh_c(1 + h_c)^2} y_N - \frac{h_c(3 + 2h_c)}{h(1 + h_c)^2}(y_N - y_{N-1}),$$

замыкающих систему уравнений для первых производных сплайна с номером 0. Коэффициент α_{01} в разложении по ζ интегрального отклонения I_e есть линейная комбинация Z'_a, Z'_b, коэффициент α_{11} не зависит от ординат, в итоге искомая координата $\zeta^{(\min)}$ точки минимума не зависит от шага сетки h.

Даже простые функциональные зависимости иллюстрируют и "тестируют" логику экстраполяции с помощью кубических сплайнов, уже здесь обнаруживается разнообразие ситуаций, вполне сопоставимых с "многоцветностью" реальности эксперимента. Первое, в чем следует удостовериться, очевидно, это "первичный" аргумент в пользу качественности алгоритма экстраполяции: если для известной зависимости $y = f(x)$ "псевдоэкспериментальные" значения $y_i = f(x_i)$, $i = 0, \ldots, N$, заданы на фиксированном отрезке и число N увеличивается, прогнозируемые ординаты y_i должны приближаться к $f(x_c)$ для всех не слишком больших значений h_i. По крайней мере для равномерной сетки это действительно так, и кривые, построенные по отклонению функции, по отклонению производной, по комбинированному отклонению

$$\frac{I_e}{\min\limits_{\zeta}(I_e)} + \frac{J_e^{(1)}}{\min\limits_{\zeta}(J_e^{(1)})},$$

практически неразличимы (рис. 1). Минимальные по $\zeta = y_c$ значения отклонений $J_e^{(k)}$, $k = 0, 1, 2$, строго говоря, должны быть функциями h_c, но оказывается, что они не зависят от h_c (рис. 2).

На фиксированном отрезке $[0, 2\pi]$, согласно рис. 1, в узлах равномерной сетки

$$x_i = hi, \qquad i = 0, \ldots, N, \qquad h = \frac{2\pi}{N},$$

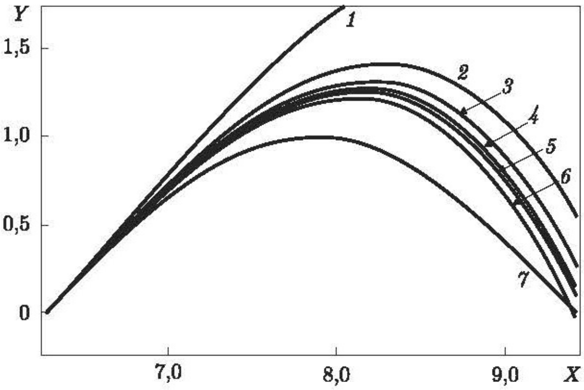

Рис. 1

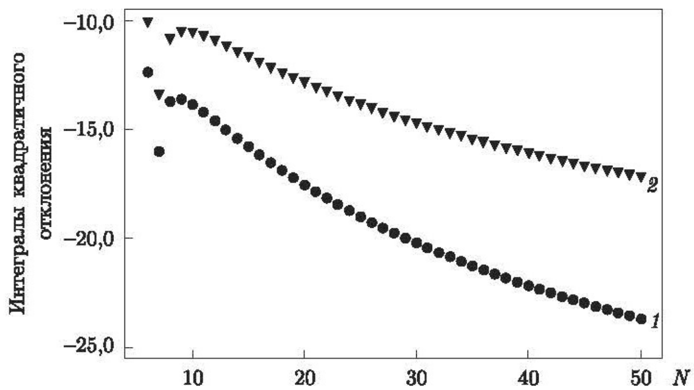

Рис. 2

заданы значения ординат $y_i = \sin(x_i)$. Прогнозируется значение y_c в точке

$$x_c = 2\pi + hh_c$$

для $N = 10$ (кривая *1*), $N = 20$ (*2*), $N = 30$ (*3*), $N = 40$ (*4*), $N = 50$ (*5*), $N = 100$ (*6*) при меняющемся h_c, $h_c^{\min} \leqslant h_c \leqslant h_c^{\max}$, $h_c^{\min} = 0{,}01$, $h_c^{\max} = \pi/h$. Графики зависимости $y_c = y_c(x_c)$ с ростом N все точнее следуют за изгибом синуса (кривая *7*).

Рис. 2 показывает, что с ростом N и уменьшением шага h на фиксированном отрезке изменения минимальные значения интегральных отклонений, как и ожидается, резко падают. Кривая *1* соответствует $\log\min(J_e^{(0)})$, кривая 2 — $\log\min(J_e^{(1)})$.

Геофизическая картина Земли — это "островки" нестохастических связей между явлениями, они могут появляться и исчезать и не сливаются в "материк". Как иллюстрация, измерения крипа в точке разлома рядом с очаговой областью не показывали никаких закономерных изменений перед землетрясением с магнитудой 5,9 в центральной Калифорнии 6 августа 1979 г., но в коровых движениях перед землетрясением Тонанкай 7 декабря 1944 г. составляющая-предвестник несомненна [3]. Только физики или химии процессов здесь недостаточно для полноты динамического описания, приемлемо сузить набор динамических переменных не удается главным образом из-за сложности геологической "сцены", на которой происходят события. Решающую роль приобретает, следовательно, формулировка общих "первичных" принципов, позволяющих ограничить множество вариантов выбора модели, принципов, относящихся к совокупности, "россыпи" на плоскости экспериментальных точек. Это в первую очередь в определенном смысле понимаемая оптимальность, отражение всеобщности “бритвы Оккама” [4, *с. 5*]. Изложенную выше математическую схему следует рассматривать именно с такой точки зрения.

1. *Завьялов Ю. С., Квасов Б. И., Мирошниченко В. Л.* Методы сплайн-функций. – Москва: Наука, 1980. – 352 с.
2. *Вершинин В. В., Завьялов Ю. С., Павлов Н. Н.* Экстремальные свойства сплайнов и задача сглаживания. – Новосибирск: Наука, 1988. – 102 с.
3. *Моги К.* Предсказание землетрясений. – Москва: Мир, 1988. – 382 с.
4. *Романовский Ю. М., Степанова Н. В., Чернавский Д. С.* Математическая биофизика. – Москва: Наука, 1984. – 234 с.

Отдел сейсмологии Института геофизики им. С. И. Субботина НАН Украины, Симферополь

Поступило в редакцию 09.09.13

О. С. Костінський

Про принципи сплайн-екстраполяції геофізичних даних

Можливі застосування сплайнової математики обговорюються стосовно до геофізичних спостережень, коли побудувати фізичну динамічну модель або неможливо, або занадто складно, нераціонально. У подібних ситуаціях проста ідея сплайн-екстраполяції виявляється единою: сітка вузлів на заданому сегменті доповнюється прогнозованою точкою, будується “прогностичний” сплайн на розширеній сітці, необхідно забезпечити мінімум інтеграла квадратичного відхилення, залежного від ординати додаткової точки як від параметра. Для рівномірної сітки структурні одиниці алгоритму екстраполяції представляються у вигляді послідовності розкладів за координатами заданих точок, коефіцієнти розкладань доступні аналітично. Показано, що ордината прогнозованої точки не залежить від кроку сітки, це суттєво для оцінки найближчого наступного в серії регулярних вимірювань, коли принциповою є не величина інтервалу між вимірами, а його незмінність.

A. S. Kostinsky

On the principles of a spline extrapolation concerning geophysical data

Possible applications of spline mathematics applied to geophysical observations, when to build a physical dynamic model is either impossible or too complicated and unpractical, are discussed. In situations like this, the simple idea of spline extrapolation is determined uniquely: the net of knots on a specified segment is supplemented by a potentially predictable point, a “prognostic” spline on the augmented net is built, and it is necessary to ensure a minimum of the integral of the quadratic deviation depending on the add-on point ordinate as a parameter. For a uniform net base, structural units of the extrapolation algorithm are represented in the form of a sequence of expansions in terms of coordinates of the specified points, and the expansion coefficients are available analytically. It is found that the forecasted point ordinate does not depend on the net spacing, which is essential for the evaluation of the nearest next event in a series of regular measurements, when the basic thing is not the interval between measurements, but its constancy.