Сборник

ТГУ

6 февраля 2006 г.

Оглавление

ЗАДАЧА СТОХАСТИЧЕСКОГО ОЦЕНИВАНИЯ СЛУЧАЙНОЙ ФУНКЦИИ И ЕЕ ВАРИАЦИОННЫЙ АНАЛОГ - ПОСТРОЕНИЕ СПЛАЙН-ФУНКЦИИ.

М.В. Дмитриевский
e-mail: mishadm@mail.ru

В работе рассмотрен один стохастический подход к решению задач фильтрации шума, аппроксимации и интерполяции функций. Дан его вариационный аналог, позволяющий в задачах построения сплайн-функций подойти к конструированию функционала энергии исходя из стохастических соображений.

Рассмотрим конечномерное пространство E^N со скалярным произведением

$$(a,b)_{E^N} = a^T b,$$
$$a = (a_1, \ldots, a_N)^T \in E^N$$
$$b = (b_1, \ldots, b_N)^T \in E^N$$

Пусть задан вектор $z \in E^M = Z$, который в дальнейшем будем называть вектором исходных данных:

$$z = (z_1, \ldots, z_M)^T, \tag{1}$$

тогда задачу построения аппроксимационного или интерполяционного сплайна $\sigma \in E^N = X$ можно записать в виде [1]:

аппроксимационный сплайн:

$$\Phi_\alpha(\sigma) = \|A\sigma - z\|_Z^2 + \alpha\|T\sigma\|_Y^2 = \min_{\sigma\in X}, \tag{2}$$

интерполяционный сплайн:

$$\begin{cases} A\sigma = z, \\ \|T\sigma\|_Y^2 = \min\limits_{\sigma\in X} \end{cases}, \tag{3}$$

где Y - N-мерное пространство со скалярным произведением

$$(a,b)_Y = a^T Q\, b \tag{4}$$

с симметричной, положительно определенной матрицей Q размерности $N \times N$, $A : X \to Z$, $T : X \to Y$ - линейные ограниченные операторы, $\alpha > 0$ - параметр "сглаживания".

Предметом данной статьи является конструирование функционала $\|T\sigma\|_Y^2$ с учетом априорной информации о функции σ. Задача сводится к выбору оператора T и матрицы Q, определяющей скалярное произведение в Y.

Задача 1: фильтрация шума.

Рассмотрим задачу восстановления случайной функции

$$f = (f_1, f_2, \ldots, f_N)^T \in H^N$$

в узлах сетки $t = \{t_1, \ldots, t_N\}$, состоящей из N узлов, где H^N - пространство случайных функций со скалярным произведением

$$(\xi, \psi)_H = M\xi^T\psi, \quad \xi, \psi \in H^N, \tag{5}$$

знак M означает математическое ожидание [3]. Пусть известны значения случайной функции $y \in H^N$ в узлах равномерной сетки t:

$$y = (y_1, y_2, \ldots, y_N)^T, \tag{6}$$

и известно, что функция y связана с функцией ε соотношением:

$$y = f + \varepsilon, \tag{7}$$

где случайная функция $\varepsilon \in H^N$ характеризует собой погрешности произведенных измерений:

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)^T.$$

В качестве дополнительной информации о функциях f и ε известно, что их значения распределены по нормальному закону:

$$f \sim N(0, K), \tag{8}$$

$$\varepsilon \sim N(0, \lambda E), \tag{9}$$

где K - ковариационная матрица размерности $(N \times N)$, симметричная и положительно определенная, E - единичная матрица. Число

$\lambda > 0$ можно трактовать как уровень шума. Из уравнений (7), (8) и (9) следует:

$$Mf = 0, \quad M\varepsilon = 0,$$
$$Mff^T = K, \quad M\varepsilon\varepsilon^T = \lambda E. \tag{10}$$

Логично также предположить независимость случайных функций f и ε, что записывается в виде равенства

$$Mf\varepsilon^T = 0. \tag{11}$$

В практических задачах матрица K не задана и ее необходимо конструировать, причем желательно таким образом, чтобы она содержала в себе наши априорные представления о восстанавливаемой функции.

Матрица K состоит из элементов вида Mf_if_j и содержит в себе средние значения всевозможных произведений элементов случайного вектора f. В геостатистике [2] вместо значений вида Mf_if_j рассматривают значения

$$M\left(f_i - f_j\right)^2 = Mf_if_i - 2Mf_if_j + Mf_jf_j, \tag{12}$$

характеризующее средний квадрат разности элементов вектора f. Значения (12) вычисляются при помощи специальной функции, называемой вариограммой. В эту функцию закладывают априорную информацию об "изменчивости" вектора f.

Нам необходимо найти функцию $\widehat{f} \in H^N$, которая была бы наилучшим приближением к функции f в следующем смысле:

$$S = ||f - \widehat{f}||_H^2 = \min. \tag{13}$$

Эта запись означает, что необходимо подобрать $\widehat{f}$ таким образом, чтобы среднее значение квадрата невязки $f - \widehat{f}$ было минимальным.

Функцию $\widehat{f}$ будем искать в виде:

$$\widehat{f} = Ay = \begin{pmatrix} a_1^T \\ a_2^T \\ \dots \\ a_N^T \end{pmatrix} y, \tag{14}$$

где A - матрица размерности $N \times N$, состоящая из строк a_i^T. Исходя из формулы (14) значение функции $\widehat{f}$ в узле t_p рассчитывается по формуле:

$$\widehat{f}_p = a_p^T y.$$

Уравнение (13) можно записать в следующем виде:

$$S = \sum_{i=1}^{N} S_i = \sum_{i=1}^{N} M(f_i f_i - 2 f_i \widehat{f}_i + \widehat{f}_i \widehat{f}_i). \tag{15}$$

Рассмотрим задачу минимизации слагаемого S_p:

$$S_p = M(f_p f_p) - 2M(f_p a_p^T y) + M(a_p^T y y^T a_p)$$

$$\frac{\partial S_p}{\partial a_{pj}} = -2M f_p y_j + 2M a_p^T y y_j = 0 \Rightarrow$$

таким образом для минимизации S_p получим систему линейных уравнений (СЛУ):

$$M f_p y^T = M a_p^T y y^T,$$

а для минимизации S соответственно:

$$M f y^T = M A y y^T = A M y y^T.$$

Используя формулы (10) и (11) преобразуем полученную СЛУ к виду:

$$K = A(K + \lambda E),$$

откуда получаем формулу для нахождения $\widehat{f}$:

$$\widehat{f} = Ay = K(K + \lambda E)^{-1} y. \tag{16}$$

Для задачи (13) с решением (16) существует вариационная задача, которую можно сформулировать как задачу построения аппроксимационного сплайна:

Теорема 1. *Функция σ, являющаяся решением задачи сплайн-аппроксимации* (2) *в точности совпадает с функцией $\widehat{f}$, описываемой в виде* (14) *и являющейся решением задачи* (13) *при выполнении равенств* (7)-(11)*, если выполнены следующие условия:*

1. $\alpha = \lambda$

2. $z = y$

3. $A = E$

4. $T^T Q T = K^{-1}$

5. $(a,b)_Z = a^T b, \quad a \in Z, b \in Z$

Доказательство. В условиях теоремы задача сплайн-аппроксимации запишется в следующем виде:

$$\Phi_\lambda(\sigma) = y^T y - 2y^T\sigma + \sigma^T\sigma + \lambda\sigma^T K^{-1}\sigma = \min_{\sigma \in E^N}.$$

Продифференцировав $\Phi_\lambda(\sigma)$ по σ_j получим СЛУ для нахождения σ:

$$\frac{\partial\Phi}{\partial\sigma_j} = -2y_j + 2\sigma_j + 2\lambda\sum_k K_{kj}^{-1}\sigma_j = 0, \quad j = 1, \ldots, N,$$

где K_{kj}^{-1} - элементы матрицы K^{-1}. Запишем СЛУ в матричном виде:

$$(E + \lambda K^{-1})\sigma = y. \tag{17}$$

Путем несложных преобразований получим формулу для нахождения σ:

$$(KK^{-1} + \lambda K^{-1}KK^{-1})\sigma = y$$
$$(K + \lambda E)K^{-1}\sigma = y$$
$$\sigma = K(K + \lambda E)^{-1}y \tag{18}$$

Равенство σ и $\widehat{f}$ непосредственно следует из сравнения формул (16) и (18). □

Замечание 1. Так как часто нам не будет известна матрица K^{-1}, а будет известна матрица K, то СЛУ для нахождения σ с вычислительной точки зрения удобно записывать в следующем виде:

$$(K + \lambda E)\sigma = Ky. \tag{19}$$

Замечание 2. Условием (4) теоремы можно воспользоваться для выбора матричного оператора T и матрицы Q, определяющей скалярное произведение в Y. Рассмотрим два из возможных вариантов:

1. $T^T T = K^{-1}$, $Q = E$
2. $T = E$, $Q = K^{-1}$,

т.е. априорная информация о поведении функции, заложенная при конструировании матрицы K, может быть использована или при задании скалярного произведения в Y, или путем определения оператора T или каким-либо другим способом, лишь бы выполнялось условие (4) теоремы.

Замечание 3. Так как стохастическая и вариационная постановки эквивалентны, то выбор весового коэффициента α в функционале (2) можно основывать на информации об уровне помех, содержащихся в функции $z \in Z$.

Пример 1. Продемонстрируем, что задание априорной информации о поведении функции оказывает существенное влияние при решении задачи сплайн-аппроксимации. Пусть нам необходимо восстановить функцию

$$f = 4sin(t), \quad t = \{1 : 0.1 : 14\}$$

по замерам

$$z = (z_1, \dots, z_N),$$

содержащим случайную погрешность

$$\varepsilon \sim N(0, 0.5).$$

На рис. 1 изображены два графика: пунктирной линией отображается кривая z, используемая в качестве исходных данных, кривой с квадратиками - истинная кривая f, которую необходимо восстановить, а кривой с кружочками - сплайновая кривая σ, являющаяся результатом расчетов.

На первом графике в качестве функционала $||T\sigma||_Y^2$ использовалась сумма $\sum_{i=1}^{N-1} (\sigma_{i+1} - \sigma_i)^2$. Эта запись соотвествует нашему представлению о том, что кривая должна как можно меньше отличаться от константы. Весовой коэффициент α брался равным 0.5.

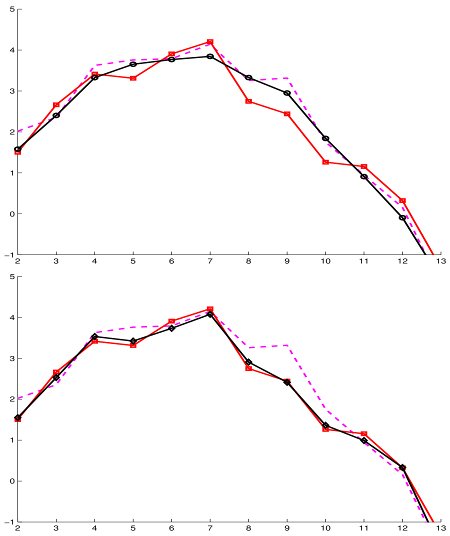

Рис. 1: Влияние учета априорной информации.

Для построения второго графика была сконструрована матрица K, удовлетворяющая условиям теоремы 1. Для этого были использованы 21 случайная реализация функции z. Как видно из сравнения двух графиков, функция σ, построенная с использованием матрицы K и формулы (17) (второй график) более точно повторяет искомую функцию f. На первом графике очевидно более точное повторние заданной функции z.

Задача 2: аппроксимация.

Рассмотрим задачу построения функции $\widehat{f}$ для случая, когда вектор y задан не во всех узлах сетки t, т.е. будем считать, что вместо вектора y задан вектор Y, представленный P значениями $(P < N)$, т.е.

$$Y = (Y_1, \dots, Y_P)^T. \tag{20}$$

Без ограничения общности будем рассматривать случай, когда вектор Y задан в первых P узлах сетки t:

$$Y_i = f_i + \varepsilon_i, \quad i = 1, \ldots, P \tag{21}$$

Как и ранее будем полагать, что функции f и ε распределены по нормальному закону (8) и (9) и в дополнение к формулам (10) запишем формулу

$$MYY^T = \widehat{E}(K + \lambda E)\widehat{E}^T, \tag{22}$$

где $\widehat{E}$ матрица размерности $(P \times N)$:

$$\widehat{E} = \begin{pmatrix} eye(P) & zeros(P, N-P) \end{pmatrix}.$$

$eye(P)$ - единичная матрица размерности $(P \times P)$, $zeros(P, N-P)$ - нулевая матрица размерности $(P \times N - P)$.

Так же, как и в задаче фильтрации шума, будем искать функцию $\widehat{f} \in H^N$, которая была бы наилучшим приближением к функции f в смысле (13). Функцию $\widehat{f}$ будем искать в виде

$$\widehat{f} = \widehat{A}Y, \tag{23}$$

где $\widehat{A}$ - матрица размерности $(N \times P)$, состоящая из строк $\widehat{a}_p^T$. Используя запись (15), рассмотрим минимизацию слагаемого S_p:

$$S_p = Mf_pf_p - 2M(f_p\,\widehat{a}_p^TY) + M(\widehat{a}_p^TYY^T\widehat{a}_p),$$

для минимизации которого необходимо решить СЛУ

$$\frac{\partial S_p}{\partial \widehat{a}_{p,j}} = -2Mf_pY_j + 2M\widehat{a}_p^TYY_j = 0, \quad j = 1, \ldots, P$$

которую можно записать в матричной форме:

$$Mf_pY^T = M\widehat{a}_p^TYY^T.$$

Таким образом, для минимизации S необходимо решить СЛУ

$$MfY^T = \widehat{A}MYY^T,$$

которую путем несложных преобразований запишем в виде, удобном для вычисления $\widehat{f}$:

$$Mf\left(\widehat{E}(f+\varepsilon)\right)^T = \widehat{A}\widehat{E}(K+\lambda E)\widehat{E}^T$$

$$K\widehat{E}^T = \widehat{A}\widehat{E}(K+\lambda E)\widehat{E}^T$$

$$\widehat{f} = \widehat{A}Y = K\widehat{E}^T\left(\widehat{E}(K+\lambda E)\widehat{E}^T\right)^{-1}Y. \tag{24}$$

Для задачи (13) аппроксимации исходных данных (20),(21) с решением (24) существует вариационная задача, которую можно сформулировать как задачу построения аппроксимационного сплайна:

Теорема 2. *Функция σ, являющаяся решением задачи сплайн-аппроксимации* (2) *в точности совпадает с функцией $\widehat{f}$, описываемой в виде* (23) *и являющейся решением задачи* (13) *при выполнении равенств* (21), (8)-(11), *если выполнены следующие условия:*

1. $\alpha = \lambda$
2. $z = Y$
3. $A = \widehat{E}$
4. $T^T Q T = K^{-1}$
5. $(a,b)_Z = a^T b, \quad a \in Z, b \in Z.$

Доказательство. В условиях теоремы функционал (2) примет вид:

$$\Phi_\lambda(\sigma) = \sum_{j=1}^{P}(Y_j - \sigma_j)^2 + \lambda\|T\sigma\|_Y^2 = \min_{\sigma\in X}.$$

Для того, чтобы найти минимум $\Phi_\lambda(\sigma)$ необходимо решить СЛУ:

$$\begin{cases} \dfrac{\partial\Phi_\lambda}{\partial\sigma_k} = -2(Y_k - \sigma_k) + 2\lambda\displaystyle\sum_{j=1}^{N} K_{j,k}^{-1}\sigma_k = 0, \quad k = 1,\ldots,P \\ \dfrac{\partial\Phi_\lambda}{\partial\sigma_k} = 2\lambda\displaystyle\sum_{j=1}^{N} K_{j,k}^{-1}\sigma_k = 0, \quad k = P+1,\ldots,N \end{cases} \tag{25}$$

Запишем полученную СЛУ в матричной форме:

$$\left(\widehat{E}^T \widehat{E} + \lambda K^{-1}\right) \sigma = \widehat{E}^T Y$$

Найдем выражение для σ:

$$\left(\widehat{E}^T \widehat{E} K K^{-1} + \lambda K^{-1} K K^{-1}\right) \sigma = \widehat{E}^T Y$$

$$\left(\widehat{E}^T \widehat{E} K + \lambda E\right) K^{-1} \sigma = \widehat{E}^T Y$$

$$\sigma = K \left(\widehat{E}^T \widehat{E} K + \lambda E\right)^{-1} \widehat{E}^T Y \tag{26}$$

Покажем теперь, что $\widehat{f}$, вычисляемая по формуле (24) в точности совпадает с σ, вычисляемой по формуле (26). Для этого распишем равенство $\widehat{f} = \sigma$:

$$K \left(\widehat{E}^T \widehat{E} K + \lambda E\right)^{-1} \widehat{E}^T Y = K \widehat{E}^T \left(\widehat{E}(K + \lambda E)\widehat{E}^T\right)^{-1} Y$$

умножая слева на K^{-1} получим:

$$\left(\widehat{E}^T \widehat{E} K + \lambda E\right)^{-1} \widehat{E}^T Y = \widehat{E}^T \left(\widehat{E}(K + \lambda E)\widehat{E}^T\right)^{-1} Y \Rightarrow$$

$$\widehat{E}^T Y = \left(\widehat{E}^T \widehat{E} K + \lambda E\right) \widehat{E}^T \left(\widehat{E}(K + \lambda E)\widehat{E}^T\right)^{-1} Y \Rightarrow$$

$$\left(\widehat{E}(K + \lambda E)\widehat{E}^T\right) \widehat{E}^T = \left(\widehat{E}^T \widehat{E} K + \lambda E\right) \widehat{E}^T \Rightarrow$$

$$\widehat{E}^T \widehat{E} K \widehat{E}^T + \lambda \widehat{E}^T \widehat{E} E \widehat{E}^T = \widehat{E}^T \widehat{E} K \widehat{E}^T + \lambda \widehat{E}^T \widehat{E} E \widehat{E}^T.$$

$\square$

Задача 3: интерполяция.

Рассмотрим предыдущую задачу для случая, когда отсутствуют помехи, т.е. $\varepsilon = 0$. Тогда (21) запишется так:

$$Y_i = f_i, \quad i = 1, \ldots, P. \tag{27}$$

Заметим, что в этом случае

$$MYY^T = \widehat{E}K\widehat{E}^T.$$

Как и раньше, $\widehat{f}$ будем искать в виде (23) такую, которая бы минимизировала функционал (13). Функция $\widehat{f}$, доставляющая минимум этому функционалу находится из решения СЛУ:

$$MfY^T = \widehat{A}MYY^T$$

$$Mf\widehat{E}f^T = \widehat{A}\widehat{E}K\widehat{E}^T$$

$$K\widehat{E}^T = \widehat{A}\widehat{E}K\widehat{E}^T$$

$$\widehat{f} = \widehat{A}Y = K\widehat{E}^T\left(\widehat{E}K\widehat{E}^T\right)^{-1}Y. \tag{28}$$

Для задачи (13) аппроксимации исходных данных (20),(27) с решением (28) существует вариационная задача, которую можно сформулировать как задачу построения интерполяционного сплайна:

Теорема 3. *Функция σ, являющаяся решением задачи сплайн-интерполяции* (3) *в точности совпадает с функцией $\widehat{f}$, описываемой в виде* (23) *и являющейся решением задачи* (13) *при выполнении равенств* (27), (8)-(11), *если выполнены следующие условия:*

1. $z = Y$
2. $A = \widehat{E}$
3. $T^TQT = K^{-1}$
4. $(a,b)_Z = a^Tb, \quad a \in Z, b \in Z.$

Доказательство. В условиях теоремы, задача сплайн-интерполяции (3) запишется следующим образом:

$$\begin{cases} \sum_{j=1}^{P}(Y_i - \sigma_i)^2 = \min, \\ \sigma^T K^{-1}\sigma = \min \end{cases} \tag{29}$$

Перепишем эту задачу в виде задачи минимизации одного функционала, используя метод неопределенных множителей Лагранжа:

$$S = \sigma^T K^{-1} \sigma + \mu \left(Y - \widehat{E} \sigma \right) = \min_{\sigma},$$

где $\mu = (\mu_1, \ldots, \mu_P)^T$ - множители Лагранжа. Продифференцировав S по σ_j, получим СЛУ для нахождения σ:

$$\begin{cases} \dfrac{\partial S}{\partial \sigma_j} = 2 \displaystyle\sum_{k=1}^{N} K_{j,k}^{-1} \sigma_k - 2\mu_j = 0, \quad j = 1, \ldots, P \\ \dfrac{\partial S}{\partial \sigma_j} = 2 \displaystyle\sum_{k=1}^{N} K_{j,k}^{-1} \sigma_k = 0, \quad j = P+1, \ldots, N \\ \dfrac{\partial S}{\partial \mu_j} = Y_j - \sigma_j = 0, \quad j = 1, \ldots, P \end{cases} \tag{30}$$

Запишем полученную СЛУ в матричной форме. Используя второе матричное уравнение, выразим вектор μ и подставим его в первое уравнение:

$$\begin{cases} K^{-1}\sigma + \widehat{E}^T \mu = 0 \\ \widehat{E}\sigma = Y \end{cases}$$

$$\sigma + K\widehat{E}^T \mu = 0$$

$$Y + \widehat{E}K\widehat{E}^T \mu = 0 \quad \Rightarrow \quad \mu = -\left(\widehat{E}K\widehat{E}^T\right)^{-1} Y \Rightarrow$$

$$\sigma = K\widehat{E}^T \left(\widehat{E}K\widehat{E}^T\right)^{-1} Y. \tag{31}$$

Доказательство теоремы следует из сравнения формул (28) и (31). □

Выводы:

В работе предложен способ учета априорной информации при построении сплайн-функций. Он может быть использован в тех областях знаний, где информации для построения функции σ всегда недостаточно и есть возможность доопределять ее из статистических соображений. К таким областям знаний, например, относится геология и задача построения карт геологических параметров. Остается открытым вопрос конструировании матрицы K^{-1}. Ясно лишь, что она должна быть построена таким образом, чтобы с одной стороны хорошо отражала закладываемую статистическую гипотезу, а с другой стороны имела простую структуру и была разреженной.

Список литературы

[1]. Василенко В.А. Сплайн-функции: теория, алгоритмы, программы. Новосибирск: Наука, 1983.

[2]. М. Давид. Геостатистические методы при оценке запасов руд, Л.: Недра, 1980, 360 стр.

[3]. Розанов Ю.А. Случайные процессы (краткий курс), Главная редакция физико-математической литературы изд-ва "Наука", 1971.

[4]. Fomel, S. Three-dimensional seismic data regularization. Stanford Exploration Project, http://sepwww.stanford.edu/sep/prof

[5]. Wahba G. An Introduction to Model Building with Reproducing Kernel Hilbert Spaces. Technical Report NO. 1020, April 18, 2000, Departament of Statistics, University of Wisconsin.