М Е Т О Д Ы   С П Л А Й Н - Ф У Н К Ц И Й
1976 год (Вычислительные системы) Выпуск 68

УДК 518.5

ИНТЕРПОЛЯЦИЯ ПЛОСКИХ КРИВЫХ

В.А.Скороспелов

Излагаются два метода интерполяции дуг плоских кривых кубическими сплайнами дефекта 2, основанные на локальных процедурах построения интерполирующих кривых. Достоинством этих методов является, во-первых, тот факт, что для получения коэффициентов сплайна нет необходимости решать систему, во-вторых, полученные кривые не имеют нежелательных осцилляций, которые нередко являются причиной отказа от классических методов интерполяции. Благодаря этим свойствам вместе с простотой реализации вычислительных алгоритмов на ЭВМ эти методы представляют собой достаточно универсальный аппарат для построения плавных кривых и с успехом могут применяться в автоматизированных системах проектирования и технологической подготовки производства в таких отраслях машиностроения, как авиастроение, судостроение, автомобилестроение, турбостроение.

Задача интерполяции рассматривается в следующей постановке. Пусть задана упорядоченная последовательность точек, представленных своими радиусами-векторами относительно некоторой прямоугольной системы координат, в каждой из которых определено направление касательного вектора $\{\bar{r}_i, \bar{H}_i\}$ $(i = 1,2,\ldots,n)$, где $\bar{r}_i = (X_i, Y_i)$, $\bar{H}_i = (H_{xi}, H_{yi})$. Требуется построить непрерывную кривую с непрерывной касательной, проходящую через заданные точки и имеющую заданное направление касательного вектора в них и минимальное число точек перегиба на рассматриваемой дуге, которое полностью определяется заданной системой узлов.

Первые два параграфа посвящены изложению методов построения таких кривых. В третьем параграфе излагается методика прибложенного определения направления касательного вектора в тех узлах, в которых оно не задано.

§I. Построение интерполяционного сплайна с использованием местной системы координат

Следует отметить, что описываемый в этом параграфе метод не допускает абсолютно произвольного выбора узлов интерполяции. Ограничение состоит в том, что углы наклона заданных касательных векторов к соседним хордам не должны достигать значения $\pi/2$. Аналитически это ограничение выражается условиями: $(\bar{H}_i, \bar{h}_i) > 0$; $(\bar{H}_{i+1}, \bar{h}_i) > 0$, где $\bar{h}_i = (\bar{r}_{i+1} - \bar{r}_i)$ – вектор хорды, (,) – скалярное произведение.

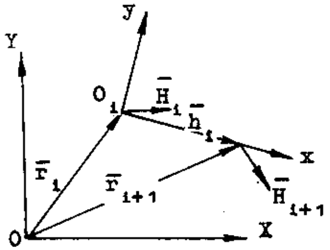

Рис.I

На i-м участке введем местную прямоугольную систему координат $O_i xy$. Ее начало (рис.I) поместим в точку $\bar{r}_i$; ось x направим вдоль хорды так, чтобы ее положительное направление совпало с направлением вектора хорды; направление оси y выбирается так, чтобы система была правая.

Матрица перехода в систему координат $O_i xy$ будет

$$\mu = \frac{1}{L_i}\begin{vmatrix} \Delta X_i, & \Delta Y_i \\ -\Delta Y_i, & \Delta X_i \end{vmatrix},$$

где $\Delta X_i = X_{i+1} - X_i$, $\Delta Y_i = Y_{i+1} - Y_i$, $L_i = |\bar{h}_i|$.

В местной системе координат интерполяционная функция на рассматриваемом участке есть кубический полином

$$y^{(i)}(x) = \sum_{k=0}^{3} a_k^{(i)} x^k, \qquad x \in [0, L_i], \qquad (\mathrm{I.I})$$

коэффициенты которого определяются из условий, что его производные на концах участка принимают заданные значения:

$$y'(0) \equiv y'_0 = \frac{-H_{xi}\,\Delta Y_i + H_{yi}\,\Delta X_i}{H_{xi}\,\Delta X_i + H_{yi}\,\Delta Y_i} \qquad (\mathrm{I.2})$$

$$y'(L_i) \equiv y'_1 = \frac{-H_{x\,i+1}\,\Delta Y_i + H_{y\,i+1}\,\Delta X_i}{H_{x\,i+1}\,\Delta X_i + H_{y\,i+1}\,\Delta Y_i},$$

а сам он в этих точках обращается в нуль. Все дальнейшие рассуждения будут справедливы для любого участка, и поэтому можно опустить его номер i. Выражения для коэффициентов полинома (I.I) с учетом (I.2) имеют вид

$$a_0 = 0; \quad a_1 = y'_0; \quad a_2 = -\frac{2y'_0 + y'_1}{L}; \quad a_3 = \frac{y'_0 + y'_1}{L^2}. \qquad (\mathrm{I.3})$$

Интерес представляют следующие свойства этого полинома.

УТВЕРЖДЕНИЕ I. Для того чтобы кубический полином (I.I) с коэффициентами (I.3) сохранял постоянный знак кривизны на отрезке $[0,L]$, необходимо и достаточно, чтобы выполнялись условия:

$$\left.\begin{aligned} &\text{а)}\quad y'_0\, y'_1 < 0 \quad \text{или} \quad y'_0 = y'_1 = 0; \\ &\text{б)}\quad \tfrac{1}{2}|y'_1| \le |y'_0| \le 2\,|y'_1|. \end{aligned}\right. \qquad (\mathrm{I.4})$$

ДОКАЗАТЕЛЬСТВО этого утверждения сводится к исследованию поведения $y''(x)$ на отрезке $[0,L]$ и ввиду простоты опускается.

УТВЕРЖДЕНИЕ 2. Если условие "б" из (I.4) не выполняется и $y'_0, y'_1 \ne 0$, тогда существует единственный кубический полином:

а) удовлетворяющий заданным условиям на конце участка с большей по модулю производной,

б) сопрягающийся до второй производной в некоторой точке $x_0 \in [0,L]$ с прямой, проходящей через противоположный конец участка с наклоном, равным заданной там производной.

Построенная таким образом кривая на отрезке $[0,L]$ сохраняет постоянный знак кривизны.

ДОКАЗАТЕЛЬСТВО этого утверждения осуществляется путем вычисления коэффициентов полинома и точки x_0. Обозначим через d меньшую из данных по модулю производную, а через c – большую.

Пусть $c \equiv y_0'$. Искомый полином запишем по степеням $(x - x_0)$:

$$y(x) = \sum_{k=0}^{3} a_k (x - x_0)^k, \quad x \in [0,x_0]. \tag{I.5}$$

Он должен удовлетворять условиям:

$$\left.\begin{aligned} &y(0) = 0, \qquad y'(0) = c, \\ &y(x_0) = d(x_0 - L), \quad y'(x_0) = d, \quad y''(x_0) = 0. \end{aligned}\right\} \tag{I.6}$$

Подставляя (I.5) в (I.6), получим систему линейных алгебраических уравнений относительно неизвестных a_k, x_0. Ее решение:

$$\left.\begin{aligned} &a_0 = d(x_0 - L), \quad a_1 = d, \quad a_2 = 0, \\ &a_3 = \frac{c-d}{3x_0^2}, \qquad x_0 = -\frac{3dL}{c-d}. \end{aligned}\right\} \tag{I.7}$$

Так как $d(c-d) < 0$ и $|3d| < |c| + |d|$, то $0 < x_0 < L$.

Пусть $c \equiv y_1'$. В этом случае искомый полином должен удовлетворять условиям:

$$y(L) = 0, \quad y'(L) = c, \quad y'(x_0) = d,$$
$$y(x_0) = dx_0, \qquad y''(x_0) = 0.$$

Решение получаемой системы на этот раз выглядит так:

$$a_0 = dx_0, \quad a_1 = d, \quad a_2 = 0,$$
$$a_3 = \frac{c-d}{3(L-x_0)^2}, \qquad x_0 = L + \frac{3dL}{c-d}.$$

Нетрудно убедиться, что и в этом случае справедливо неравенство $0 < x_0 < L$.

УТВЕРЖДЕНИЕ 3. Если условие "а" из (I.4) не выполняется, то на отрезке $[0,L]$ полином (I.I) имеет одну и только одну точку перегиба.

Это утверждение очевидно, поскольку вторая производная от кубического полинома есть линейная функция, принимающая в этом случае на концах отрезка значения разного знака.

Окончательно алгоритм построения интерполирующей функции на участке можно представить в таком виде:

I. Если условия "а" и "б" из (I.4) выполнены либо нарушено условие "а", то коэффициенты определяются по формулам (I.3).

2. Если условие "а" из (I.4) выполнено, а "б" нарушено, то принимаем

$$a_0 = d(x_0 - E), \quad a_1 = d, \quad a_2 = 0, \quad x_0 = \left|\tilde{E} + \frac{3\,dL}{c-d}\right|,$$

$$a_3^0 = E\frac{c-d}{(\tilde{E} - x_0)^2 L}, \quad a_3^1 = \tilde{E}\frac{c-d}{3(\tilde{E} - x_0)^2 L}, \quad a_3 = \begin{cases} a_3^0 & \text{для } x < x_0, \\ a_3^1 & \text{для } x \geq x_0; \end{cases}$$

где

$$E = \begin{cases} 0, & \text{если } d = y_0', \\ L, & \text{если } d = y_1'; \end{cases}$$

$$\tilde{E} = \begin{cases} L, & \text{если } d = y_0', \\ 0, & \text{если } d = y_1'. \end{cases}$$

Описанный метод наиболее эффективен для интерполяции кривых, имеющих точки, в окрестности которых наблюдается резкое изменение кривизны. Такими точками, например, являются точки сопряжения отдельных дуг, образующих кривую. Обычный метод интерполяции кубическими параметрическими сплайнами дефекта I [I] в этом случае дает нежелательные осцилляции и не может быть применен без специального выбора узлов.

Следует заметить, что если на кривой имеются точки перегиба, нежелательно выбирать их в качестве узлов интерполяции.

§2. Интерполяция локальными параметрическими кубическими сплайнами дефекта 2

В отличие от метода, изложенного в предыдущем параграфе, излагаемый ниже допускает совершенно произвольный выбор узлов и может применяться для приближения произвольных непрерывных кривых, имеющих непрерывную касательную. Интерполирующая кривая на участке $(i,i+1)$ определяется в виде:

$$\bar{r}^{(i)}(t) = \sum_{k=0}^{3} \bar{A}_k^{(i)} t^k, \quad t \in [0,1], \quad i = 1,2,\ldots,n-1, \tag{2.I}$$

где

$$\bar{A}_k^{(i)} = (a_k^{(i)},\ b_k^{(i)}).$$

Коэффициенты определяются из условий:

$$\bar{r}^{(i)}(0) = \bar{r}_i, \qquad \bar{r}^{(i)}(1) = \bar{r}_{i+1},$$
$$\frac{\partial \bar{r}^{(i)}}{\partial t}\bigg|_{t=0} = c_i \bar{H}_i, \qquad \frac{\partial \bar{r}^{(i)}}{\partial t}\bigg|_{t=1} = d_i \bar{H}_{i+1}. \tag{2.2}$$

Здесь c_i, $d_i \geq 0$ – пока неизвестные параметры, выбору которых фактически посвящено остальное содержание параграфа. Будем считать, что $\bar{H}_i$ $(i = 1,\ldots,n)$– единичные векторы.

Из (2.2) с учетом (2.I) получим выражения коэффициентов:

$$\bar{A}_0^{(i)} = \bar{r}_i, \qquad \bar{A}_1^{(i)} = c_i \bar{H}_i,$$
$$\bar{A}_2^{(i)} = 3\bar{h}_i - 2c_i\bar{H}_i - d_i\bar{H}_{i+1}, \quad \bar{A}_3^{(i)} = -2\bar{h}_i + c_i\bar{H}_i + d_i\bar{H}_{i+1}. \tag{2.3}$$

Поскольку все рассуждения будут справедливы для любого участка, в дальнейшем для удобства записи опускается номер участка i.

Введем обозначения:

$$\bar{r}_0 = \bar{r}(0), \quad \bar{r}_1 = \bar{r}(1), \quad L = |\bar{h}| \text{ – длина хорды}. \tag{2.4}$$

Запишем уравнение кривой (2.I) с учетом (2.2), (2.3) и обозначений (2.4) в виде $\bar{r} = \bar{r}(t, \bar{r}_0, \bar{r}_1, c\bar{H}_0, d\bar{H}_1)$.

Естественно потребовать выполнения тождества

$$p\bar{r}(t, \bar{r}_0, \bar{r}_1, c\bar{H}_0, d\bar{H}_1) \equiv \bar{r}(t, p\bar{r}_0, p\bar{r}_1, e\bar{H}_0, f\bar{H}_1), \tag{2.5}$$

где $p > 0$ – скалярная величина, $e = pc$, $f = pd$. Это тождество означает, что если плоскость подвергнуть равномерному растяжению (сжатию) с коэффициентом p, то вид интерполирующей кривой не изменяется.

Пусть $p = 1/L$, т.е. вместо исходного участка с длиной хорды, равной L, мы будем рассматривать участок с единичной хордой.

Введем величины $\bar{h}_0 = \frac{\bar{h}}{L}$, $m = [\bar{H}_0, \bar{h}_0]$, $n = [\bar{H}_1, \bar{h}_0]$, $w = [\bar{H}_0, \bar{H}_1]$ и обозначим $M = em$, $N = fn$, $W = efw$.

Здесь

$$[\bar{H}, \bar{h}] = \begin{vmatrix} H_x & H_y \\ h_x & h_y \end{vmatrix}.$$

Числитель формулы кривизны плоской кривой, заданной в параметрической форме, есть $F(t) = x'y'' - y'x''$. Подставив сюда $\bar{r}'$ и $\bar{r}''$ из (2.I) с учетом (2.3), получим $F(t) = Bt^2 + Ct + D$, где $B = 6(-W + M - N)$, $C = 6(W - 2M)$, $D = 2(-W + 3M)$.

Поскольку знак выражения $F(t)$ определяет знак кривизны кривой, интерес представляет вопрос: при каких условиях это выражение сохраняет постоянный знак на отрезке $[0,I]$?

Нетрудно видеть, что $F(t)$ можно записать в виде

$$F(t) = W\left[2t^2\left(-\frac{3N}{W} - 1\right) + 2(1-t)^2\left(\frac{3M}{W} - 1\right) + t(1-t)\right].$$

Отсюда непосредственно следует

УТВЕРЖДЕНИЕ 4. *Для того чтобы $F(t)$ сохраняло знак на отрезке $[0,I]$, необходимо и достаточно, чтобы выполнялись условия:*

$$-\frac{N}{W} \geq \frac{1}{3}, \qquad \frac{M}{W} \geq \frac{1}{3}. \tag{2.6}$$

Перепишем (2.6) с учетом ранее введенных обозначений

$$-\frac{1}{e}\frac{n}{w} \geq \frac{1}{3}, \qquad \frac{1}{f}\frac{m}{w} \geq \frac{1}{3}. \tag{2.7}$$

Если $-\frac{n}{w}$, $\frac{m}{w} > 0$, то можно всегда так подобрать значения величин e и f, чтобы неравенства (2.6) выполнялись. В этом случае интерполирующая кривая будет сохранять постоянный знак кривизны на рассматриваемом участке.

На основании условий (2.7) и тождества (2.5) для выбора модулей касательных векторов на произвольном участке можно воспользоваться следующим правилом:

$$e = \begin{cases} 1, & \text{если } 3|n| \geq |w|; \\ 3\left|\frac{n}{w}\right|, & \text{если } 3|n| < |w|; \end{cases}$$

$$f = \begin{cases} 1, & \text{если } 3|m| \geq |w|; \\ 3\left|\frac{m}{w}\right|, & \text{если } 3|m| < |w|. \end{cases}$$

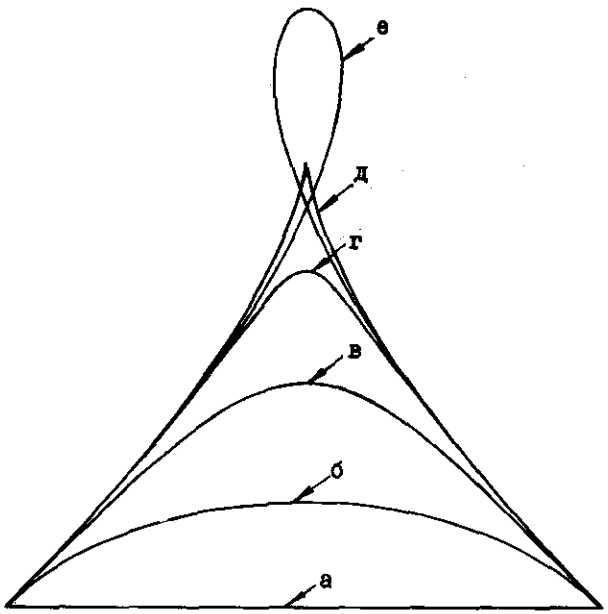

Рис. 2

Рис.2 иллюстрирует характер поведения полученной кривой в зависимости от значений величин c и d. Для этого примера взяты следующие данные: $\bar{r}_0 = (0;0)$, $\bar{r}_1 = (1;0)$, $\bar{H}_0 = (0{,}5\sqrt{2};0{,}5\sqrt{2})$, $\bar{H}_1 = (0{,}5\sqrt{2};\ -0{,}5\sqrt{2})$.

Обозначив $\alpha = \left|\frac{N}{W}\right|$, $\beta = \left|\frac{M}{W}\right|$, рассмотрим различные кривые, получающиеся для

а) $c = d = 0$, кривая вырождается в отрезок прямой, совпадающий с хордой;

б) $c = d = 1$, знак кривизны на всем участке постоянен;

в) $c = 3\alpha$, $d = 3\beta$, в граничных точках кривизна обращается в нуль; внутри рассматриваемого участка кривая сохраняет постоянный знак кривизны;

г) $c = 4\alpha$, $d = 4\beta$, точки перегиба расположены внутри рассматриваемого участка;

д) $c = 6\alpha$, $d = 6\beta$, две точки перегиба слились и образовали точку возврата;

е) $c = 8\alpha$, $d = 8\beta$, точки перегиба исчезли, и образовалась петля; знак кривизны на всем участке постоянен, но противоположен таковому для случая "б".

Следует отметить тот факт, что кривые, которые рассматривались в §I, являются частным случаем кривой (2.I) и могут быть получены путем специального выбора параметров c и d. А именно, если принять

$$c = \frac{L^2}{(\bar{H}_0, \bar{h})}, \qquad d = \frac{L^2}{(\bar{H}_1, \bar{h})},$$

тогда в местной системе координат (O_1xy) (рис. I) уравнение участка кривой будет иметь вид:

$$y = \sum_{k=0}^{3} a_k t^k, \quad x = Lt,$$

или, исключив параметр t, получим уравнение участка в виде кубического полинома

$$y = \sum_{k=0}^{3} b_k x^k.$$

На практике удобно пользоваться иной формой записи кривой (2.I). Подставив (2.3) в (2.I) и собрав степени параметра t при известных величинах $\bar{r}_0$, $\bar{r}_1$, $\bar{r}'_0$, $\bar{r}'_1$, получим

$$\bar{r}(t) = \bar{r}_0 F_0(t) + \bar{r}_1 F_1(t) + \bar{r}'_0 G_0(t) + \bar{r}'_1 G_1(t),$$

где

$$F_0(t) = 1 - 3t^2 + 2t^3, \quad F_1(t) = 3t^2 - 2t^3,$$

$$G_0(t) = t - 2t^2 + t^3, \quad G_1(t) = -t^2 + t^3.$$

Несмотря на несколько большие затраты времени ЭВМ на вычисление множества точек интерполяционной кривой, удобство такой формы очевидно. Во-первых, нет необходимости вычислять и хранить коэффициенты. Во-вторых, за счет аккуратного подсчета значений функций F_0, F_1, G_0, G_1, которые не зависят от исходных данных, можно добиться абсолютной стыковки соседних участков при реализации алгоритмов на ЭВМ.

§3. Приближенное определение направления касательного вектора в узлах плоской кривой

Если в некоторых узлах плоской кривой, выбранной в качестве узлов интерполяции, неизвестны направления касательных векторов, для их приближенного определения можно воспользоваться алгоритмом, описываемом в данном параграфе. Этот алгоритм учитывает характер кривых, применяемых в машиностроении при аналитическом задании плавных контуров. Как правило, это составные кривые, образованные путем сопряжения отрезков прямых, дуг окружностей, кривых второго порядка и др.

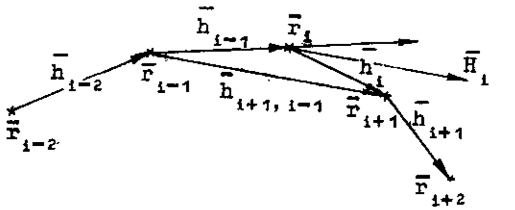

Рис. 3

Рассмотрим пять подряд идущих узлов дуги кривой $\bar{r}_{i-2}$, $\bar{r}_{i-1}$, $\bar{r}_i$, $\bar{r}_{i+1}$, $\bar{r}_{i+2}$ (рис.3). Обозначим через $\bar{h}_{i+1,i-1} = \bar{r}_{i+1} - \bar{r}_{i-1}$ вектор хорды, соединяющий узлы с номерами i - 1 и i + 1. Искомое направление касательного вектора $\bar{H}_i$ определим в виде

$$\bar{H}_i = U_i \bar{h}_{i-1} + V_i \bar{h}_i,$$

где $\quad U_i = L_i^2 u_i, \quad V_i = L_{i-1}^2 v_i,$

$$u_i = \frac{|[\bar{h}_i, \bar{h}_{i+1}]|}{L_{i+2,i} L_i L_{i+1}}, \quad v_i = \frac{|[\bar{h}_{i-2}, \bar{h}_{i-1}]|}{L_{i,i-2} L_{i-2} L_{i-1}}.$$

Нетрудно убедиться, что вектор $\bar{H}_i$ обладает следующими свойствами:

а) если все пять узлов лежат на окружности, $\bar{H}_i$ совпадает по направлению с касательным вектором к ней в точке $\bar{r}_i$;

б) если угол между хордами $\bar{h}_{i-2}$ и $\bar{h}_{i-1}$ устремить к нулю, направление вектора $\bar{H}_i$ стремится к направлению вектора $\bar{h}_{i-1}$; если устремить к нулю угол между векторами $\bar{h}_i$ и $\bar{h}_{i+1}$, направление вектора $\bar{H}_i$ стремится к направлению вектора $\bar{h}_i$;

в) если $u_i = v_i$, то $\bar{H}_i$ совпадает по направлению с касательным вектором к окружности, проходящей через точки $\bar{r}_{i-1}$, $\bar{r}_i$, $\bar{r}_{i+1}$.

Выражение (3.I) распространяется на внутренние точки с номерами i= 3,4,...,n-2. Для того чтобы воспользоваться им для крайних узлов с номерами i = 1,2,n-1,n следует добавить дополнительные узлы с номерами i = -1,0, n+1, n+2 . С этой целью можно воспользоваться следующим алгоритмом, который определяет дополнительные точки, лежащие на окружности, проходящей через три крайних.

Потребуем, чтобы дополнительные хорды $\bar{h}_0$ и $\bar{h}_n$ (рис. 4) удовлетворяли условиям:

$$\begin{cases} [\bar{h}_0, \bar{h}_1] = [\bar{h}_1, \bar{h}_2], \\ (\bar{h}_0, \bar{h}_1) = (\bar{h}_2, \bar{h}_1), \end{cases} \qquad \begin{cases} [\bar{h}_{n-2}, \bar{h}_{n-1}] = [\bar{h}_{n-1}, \bar{h}_n], \\ (\bar{h}_{n-2}, \bar{h}_{n-1}) = (\bar{h}_{n-1}, \bar{h}_n). \end{cases}$$

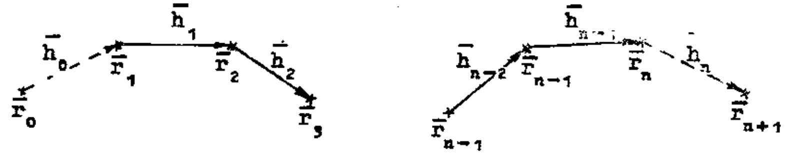

Рис. 4

Отсюда получаем

$$h_{ox}=\frac{[\bar{h}_1,\bar{h}_2]\,h_{1y}+(\bar{h}_1,\bar{h}_2)h_{1x}}{L_1^2};$$

$$h_{oy}=\frac{(\bar{h}_2,\bar{h}_1)h_{1y}-[\bar{h}_1,\bar{h}_2]\,h_{1x}}{L_1^2};$$

$$h_{nx}=\frac{[\bar{h}_{n-1},\bar{h}_{n-2}]\,h_{n-1,y}+(\bar{h}_{n-2},\bar{h}_{n-1})h_{n-1,x}}{L_{n-1}^2};$$

$$h_{ny}=\frac{(\bar{h}_{n-2},\bar{h}_{n-1})h_{n-1,y}-[\bar{h}_{n-1},\bar{h}_{n-2}]\,h_{n-1,x}}{L_{n-1}^2}.$$

По этому же правилу найдутся узлы $\bar{r}_{-1}$ и $\bar{r}_{n+2}$, если считать, что узлы $\bar{r}_0$ и $\bar{r}_{n+1}$ уже определены.

Если известно, что кривая замкнутая и не имеет точек излома, то в качестве дополнительных узлов следует взять $\bar{r}_0\equiv\bar{r}_{n-1}$, $\bar{r}_{n+1}\equiv\bar{r}_2$, $\bar{r}_{-1}\equiv\bar{r}_{n-2}$, $\bar{r}_{n+2}\equiv\bar{r}_3$.

Литература

I. АЛБЕРГ Дж., НИЛЬСОН Э., УОЛШ Дж. Теория сплайнов и ее приложения. М., "Мир", 1972.

Поступила в ред.-изд.отд.
I5 ноября I976 года