

Модуль 1 «Математические модели геометрических объектов»

Лекция 2 «Методы интерполяции кубическими сплайнами»

к.ф.-м.н., доц. каф. ФН-11, Захаров Андрей Алексеевич,
ауд.: 930а(УЛК)
моб.: 8-910-461-70-04,
email: azaharov@bmstu.ru



МГТУ им. Н.Э. Баумана

19 февраля 2016 г.

Интерполяция кубическими сплайнами наиболее часто используется для настройки путей движения объектов или получения представления существующего объекта или рисунка, хотя интерполирующие сплайны также иногда применяются для разработки форм объектов. Кубические сплайны предлагают разумный компромисс между гибкостью и скоростью вычислений. По сравнению с полиномами более высоких порядков кубические сплайны требуют меньше расчётов и памяти, они более устойчивы. Полиномы низших порядков могут не обеспечивать достаточной точности. Кроме того, параметрические кубические кривые — это кривые самого низкого порядка, которые могут занимать произвольное положение (не лежать на плоскости) в трёхмерном пространстве. Действительно, положение кривой, описываемой полиномом второй степени, задаётся тремя точками, а три точки всегда определяют плоскость. Построение кубических сплайнов сводится к решению системы линейных уравнений с трехдиагональной матрицей, имеющей доминирующую главную диагональ. Решение таких систем легко реализуется численно.

Чтобы интерполировать заданный набор контрольных точек кубическими сплайнами, по входным данным подбирается кусочно-кубическая полиномиальная кривая, проходящая через все контрольные точки.

Далее сплайновые представления обобщаются на случай пространственных поверхностей.

Постановка задачи

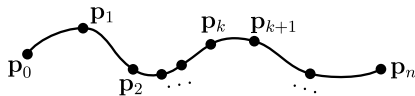


Рис.: Кусочно–непрерывная интерполяция кубическими сплайнами $n + 1$ контрольных точек

Предположим, что даны $n + 1$ контрольных точек, заданных координатами

$$\mathbf{p}_k = (x_k, y_k, z_k), \quad k = 0, 1, 2, \dots, n.$$

Параметрический кубический полином, связывающий каждую пару контрольных точек, можно описать следующей системой уравнений:

$$\mathbf{r}(t) = \mathbf{a}t^3 + \mathbf{b}t^2 + \mathbf{c}t + \mathbf{d}, \quad (t_k \leq t \leq t_{k+1}),$$
$$\mathbf{r}(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} d_x \\ d_y \\ d_z \end{pmatrix}. \quad (1)$$

Нужно определить значения четырёх коэффициентов \mathbf{a} , \mathbf{b} , \mathbf{c} , и \mathbf{d} для каждого из n участков кривой. Для этого задаётся достаточное число граничных условий в контрольных точках между участками кривой, на основе которых находятся численные значения всех коэффициентов.

Одной из первых сплайновых кривых, разработанных для графических приложений, стал *естественный кубический сплайн*. Чтобы определить естественный кубический сплайн, нужно указать, что два соседних участка кривой имеют равные первые и вторые производные по параметру в их общей точке, т.е. он непрерывен в C^2 .

Если даны $n + 1$ контрольных точек, то имеется n участков кривой и всего нужно определить $4n$ коэффициентов полиномов. В каждой из $n - 1$ внутренних контрольных точек задаётся 4 граничных граничных условия: два участка кривой с обеих сторон контрольной точки должны иметь одинаковые первые и вторые производные по параметру в данной контрольной точке, и каждая кривая должна проходить через эту контрольную точку. В результате имеем $4n - 4$ уравнений. Ещё одно уравнение даёт первая контрольная точка \mathbf{p}_0 , начало кривой, и ещё одно — контрольная точка \mathbf{p}_n , которая должна быть последней точкой кривой. Однако, чтобы определить значения всех коэффициентов, недостаёт ещё двух уравнений. Одним из способов получения двух дополнительных условий является установка вторых производных в точках \mathbf{p}_0 и \mathbf{p}_n равными 0.

Естественные кубические и бикубические сплайны играют центральную роль в практических методах аппроксимации сплайнами. Достаточные для многих приложений свойства гладкости таких сплайнов сочетаются с простотой их компьютерной реализации и высокой точностью получаемых результатов. В ряде случаев, однако, поведение интерполяционных сплайнов не согласуется с качественными характеристиками исходных данных. Визуально это проявляется в присутствии выбросов, осцилляций, различных отклонений, не характерных для исходного набора точек, а математически может быть выражено как немонотонность и наличие точек перегиба на участках монотонности и выпуклости исходных данных. Добиться «правильного» поведения сплайна можно путем увеличения числа точек интерполяции. Если этого сделать нельзя, то следует использовать свободные от перечисленных выше недостатков монотонные и выпуклые сплайны, описание методов построения которых будет дано в следующих лекциях.

Естественные кубические сплайны обладают ещё одним серьезным недостатком. Если поменять положение любой контрольной точки, изменится вся кривая. Таким образом, естественные кубические сплайны не допускают *локального управления*.

Эрмитовым сплайном (названным в честь французского математика Шарля Эрмита (Charles Hermite), который подробно исследовал их свойства) является интерполирующий кусочно-кубический полином с заданной касательной в каждой контрольной точке. В отличие от естественных кубических сплайнов эрмитовы сплайны можно настраивать локально, поскольку каждый участок кривой зависит только от условий в конечных точках.

Пусть даны:

$$\begin{aligned}\mathbf{r}(t_k) &= \mathbf{p}_k, \\ \mathbf{r}(t_{k+1}) &= \mathbf{p}_{k+1}, \\ \mathbf{r}'(t_k) &= \mathbf{p}'_k, \\ \mathbf{r}'(t_{k+1}) &= \mathbf{p}'_{k+1},\end{aligned}\tag{2}$$

где \mathbf{p}'_k и \mathbf{p}'_{k+1} задают значения производных по параметру (наклона кривой) в контрольных точках \mathbf{p}_k и \mathbf{p}_{k+1} соответственно.

Запишем вид функций $\mathbf{r}(t)$ и $\mathbf{r}'(t)$ в матричном виде:

$$\mathbf{r}(t) = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \\ \mathbf{d} \end{bmatrix}, \quad \mathbf{r}'(t) = \begin{bmatrix} 3t^2 & 2t & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \\ \mathbf{d} \end{bmatrix}. \tag{3}$$

Подставляя в (3) вместо параметра t значения конечных точек t_k и t_{k+1} , эрмитовы граничные условия можно выразить в матричной форме:

$$\begin{bmatrix} \mathbf{p}_k \\ \mathbf{p}_{k+1} \\ \mathbf{p}'_k \\ \mathbf{p}'_{k+1} \end{bmatrix} = \begin{bmatrix} t_k^3 & t_k^2 & t_k & 1 \\ t_{k+1}^3 & t_{k+1}^2 & t_{k+1} & 1 \\ 3t_k^2 & 2t_k & 1 & 0 \\ 3t_{k+1}^2 & 2t_{k+1} & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \\ \mathbf{d} \end{bmatrix}.$$

Решая это уравнение относительно коэффициентов полиномов, получаем:

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} t_k^3 & t_k^2 & t_k & 1 \\ t_{k+1}^3 & t_{k+1}^2 & t_{k+1} & 1 \\ 3t_k^2 & 2t_k & 1 & 0 \\ 3t_{k+1}^2 & 2t_{k+1} & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{p}_k \\ \mathbf{p}_{k+1} \\ \mathbf{p}'_k \\ \mathbf{p}'_{k+1} \end{bmatrix}.$$

Если $t_k = 0$, а $t_{k+1} = 1$, то

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 3 & 2 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{p}_k \\ \mathbf{p}_{k+1} \\ \mathbf{p}'_k \\ \mathbf{p}'_{k+1} \end{bmatrix} = \begin{bmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{p}_k \\ \mathbf{p}_{k+1} \\ \mathbf{p}'_k \\ \mathbf{p}'_{k+1} \end{bmatrix}. \quad (4)$$

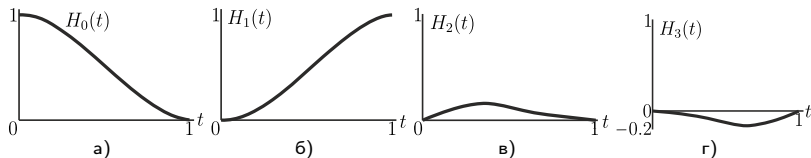


Рис.: Эрмитовы стыковочные функции

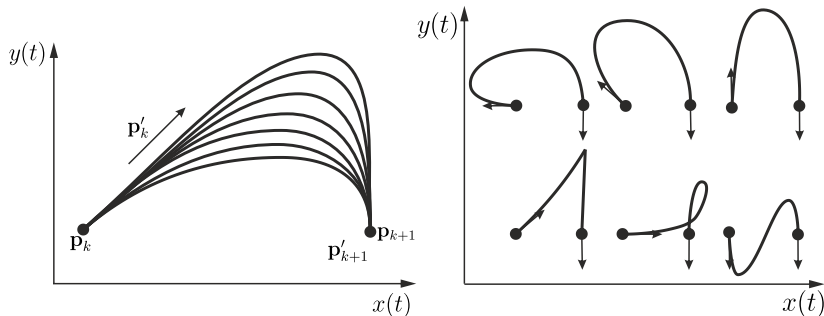
Из (4) можно получить следующую полиномиальную форму:

$$\begin{aligned} \mathbf{r}(t) = & \mathbf{p}_k (2t^3 - 3t^2 + 1) + \mathbf{p}_{k+1} (-2t^3 + 3t^2) + \mathbf{p}'_k (t^3 - 2t^2 + t) + \\ & + \mathbf{p}'_{k+1} (t^3 - t^2) = \mathbf{p}_k H_0(t) + \mathbf{p}_{k+1} H_1(t) + \mathbf{p}'_k H_2(t) + \mathbf{p}'_{k+1} H_3(t). \end{aligned}$$

В начальной точке ($t = 0$) только одна кривая $H_0(t)$ отлична от нуля. В начальной точке на форму кривой $\mathbf{r}(t)$ оказывает влияние только одна кривая $H_0(t)$. С ростом параметра t влияние функций $H_1(t)$, $H_2(t)$, $H_3(t)$ становится более заметным.

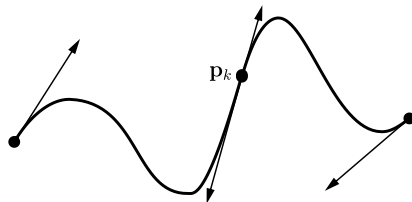
Полиномы Эрмита могут быть полезны в некоторых сферах оцифровки, где задать или аппроксимировать наклон кривой не слишком сложно.

Эрмитова интерполяция

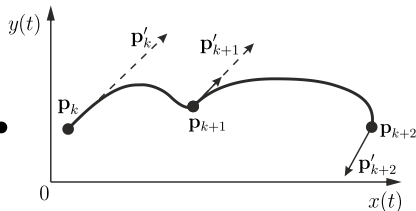


- а) Различная форма кривых определяется вариациями длины касательного вектора p'_{k+1} . Направление этого вектора и остальные величины, влияющие на форму кривых, остаются без изменения. Чем длиннее касательный вектор, тем большее влияние он оказывает на форму кривой. Чем меньше длина касательного вектора, тем более спрямленную форму имеет кривая
- б) Семейство кривых Эрмита, полученное в результате изменения направления касательного вектора одной длины

Рис.: Семейство кривых Эрмита

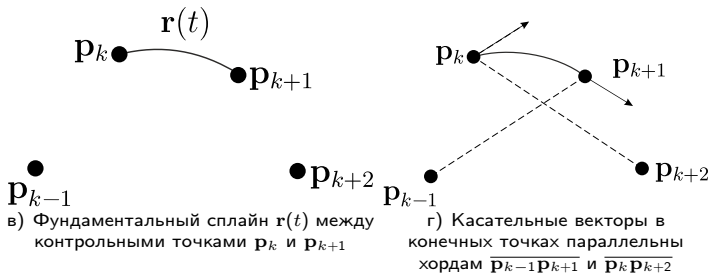


а) Интерактивное управление формой кривой Эрмита



б) Пример соединения двух фрагментов кривых Эрмита класса G^1

Многие интерактивные графические системы позволяют управлять формой кривых Эрмита посредством настройки положения конечных точек и величин касательных векторов. Эти объекты выводятся на экране дисплея и выполняют функции регуляторов формы. Оператор может при помощи мышки перемещать конечные точки и буксировать окончания касательных, программа отслеживает сделанные изменения, рассчитывает и рисует новую форму кривой. Гладкость точки сочленения объектов обычно задаётся специальной командой.



Как и эрмитовы сплайны, *фундаментальные сплайны* — это интерполирующие кусочно-кубические полиномы с заданными касательными в конечных точках. Отличие заключается в том, что теперь значения касательных в конечных точках вводить не нужно.

Участок фундаментального сплайна полностью задаётся положениями четырёх последовательных контрольных точек. Две средние контрольные точки \mathbf{p}_k и \mathbf{p}_{k+1} являются конечными точками участка, другие две \mathbf{p}_{k-1} и \mathbf{p}_{k+2} используются при расчёте касательных в конечных точках:

$$\begin{aligned} \mathbf{r}(0) &= \mathbf{p}_k, & \mathbf{r}(1) &= \mathbf{p}_{k+1}, \\ \mathbf{r}'(0) &= \frac{1}{2}(1-u)(\mathbf{p}_{k+1} - \mathbf{p}_{k-1}), & \mathbf{r}'(1) &= \frac{1}{2}(1-u)(\mathbf{p}_{k+2} - \mathbf{p}_k). \end{aligned} \quad (5)$$

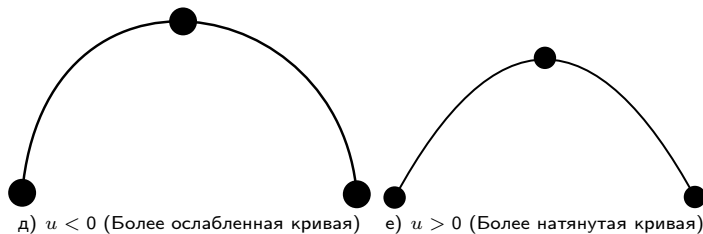


Рис.:

Параметр u называется параметром *натяжения*, поскольку он контролирует то, насколько тесно фундаментальный сплайн соответствует входным контрольным точкам. При $u = 0$ данный класс кривых называется *сплайнами Катмалла–Рома* (Catmull–Rom splines) или *сплайнами Оувергаузера* (Overhauser splines).

Фундаментальные сплайны

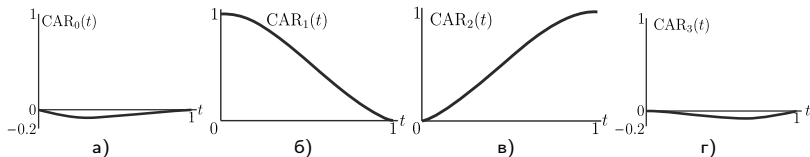


Рис.: Стыковочные функции фундаментальных сплайнов при $u = 0$ ($s = 0.5$)

Используя методы, подобные описанным для эрмитовых сплайнов, граничные условия (5) можно преобразовать в матричную форму:

$$\mathbf{r}(t) = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} -s & 2-s & s-2 & s \\ 2s & s-3 & 3-2s & -s \\ -s & 0 & s & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{p}_{k-1} \\ \mathbf{p}_k \\ \mathbf{p}_{k+1} \\ \mathbf{p}_{k+2} \end{bmatrix}, \quad s = (1-u)/2. \quad (6)$$

Расписывая матричное уравнение (6) в полиномиальной форме, получаем:

$$\begin{aligned} \mathbf{r}(t) &= \mathbf{p}_{k-1} \left(-st^3 + 2st^2 - st \right) + \mathbf{p}_k \left[(2-s)t^3 + (s-3)t^2 + 1 \right] + \\ &+ \mathbf{p}_{k+1} \left[(s-2)t^3 + (3-2s)t^2 + st \right] + \mathbf{p}_{k+2} \left(st^3 - st^2 \right) = \\ &= \mathbf{p}_{k-1} CAR_0(t) + \mathbf{p}_k CAR_1(t) + \mathbf{p}_{k+1} CAR_2(t) + \mathbf{p}_{k+2} CAR_3(t). \end{aligned}$$

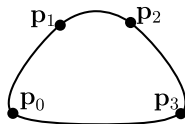


Рис.: Замкнутая кривая из четырёх участков — фундаментальных сплайнов, построенных по контрольным точкам $\{p_0, p_1, p_2, p_3\}$, $\{p_1, p_2, p_3, p_0\}$, $\{p_2, p_3, p_0, p_1\}$ и $\{p_3, p_0, p_1, p_2\}$ (параметр натяжения $u = 0$)

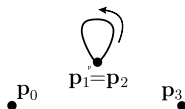


Рис.: Петля фундаментального сплайна, полученная при совпадающих конечных точках кривой. Натяжение равно 0

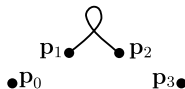


Рис.: Самопересекающаяся кривая (фундаментальный сплайн), порождённая близко расположенными конечными точками кривой. Параметр натяжения равен 0

Сплайны Коханека–Бартелса (Kochanek–Bartels splines) являются расширением фундаментальных сплайнов. В граничные условия вводятся два дополнительных параметра, обеспечивающих повышенную гибкость в настройке форм сегментов кривых:

$$\mathbf{r}(0) = \mathbf{p}_k, \quad \mathbf{r}(1) = \mathbf{p}_{k+1},$$

$$\mathbf{r}'(0) = \frac{1}{2}(1-u)[(1+b)(1-c)(\mathbf{p}_k - \mathbf{p}_{k-1}) + (1-b)(1+c)(\mathbf{p}_{k+1} - \mathbf{p}_k)],$$

$$\mathbf{r}'(1) = \frac{1}{2}(1-u)[(1+b)(1+c)(\mathbf{p}_{k+1} - \mathbf{p}_k) + (1-b)(1-c)(\mathbf{p}_{k+2} - \mathbf{p}_{k+1})],$$

где b — параметр *смещения* (биения), а c — параметр *непрерывности*. Смещение b используется для настройки кривизны в каждом конце участка, так что этот участок кривой можно скосить к одному или другому концу. Параметр c контролирует непрерывность касательного вектора на границах участков. Если параметру c присвоено ненулевое значение, на границах участков наблюдается разрыв касательной кривой.

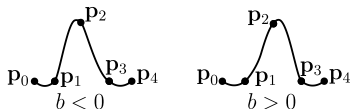


Рис.: Влияние параметра смещения на форму участка сплайна Коханека–Бартелса

Сплайны Коханека–Бартелса разработаны для моделирования путей анимации. В частности, резкие изменения движения объекта можно симитировать, используя ненулевые значения параметра c . Данные изменения движения используют в мультипликации, например, когда анимационный персонаж быстро останавливается, меняет направление или сталкивается с некоторым объектом.

Пусть задана двумерная сетка узлов и первые производные в узлах в обоих параметрических направлениях $\mathbf{r}(i, j)$, $\mathbf{r}_t(i, j)$, $\mathbf{r}_\tau(i, j)$, $\mathbf{r}_{t\tau}(i, j)$; $i = 0, 1, 2, \dots, m$; $j = 0, 1, 2, \dots, n$. Тогда поверхность Эрмита определяется следующим образом:

$$\mathbf{r}(t, \tau) = \begin{pmatrix} H_0(u) & H_1(u) & H_2(u) & H_3(u) \end{pmatrix} \times$$

$$\times \begin{pmatrix} \mathbf{r}(i, j) & \mathbf{r}(i, j+1) & \mathbf{r}_\tau(i, j) & \mathbf{r}_\tau(i, j+1) \\ \mathbf{r}(i+1, j) & \mathbf{r}(i+1, j+1) & \mathbf{r}_\tau(i+1, j) & \mathbf{r}_\tau(i+1, j+1) \\ \mathbf{r}_t(i, j) & \mathbf{r}_t(i, j+1) & \mathbf{r}_{t\tau}(i, j) & \mathbf{r}_{t\tau}(i, j+1) \\ \mathbf{r}_t(i+1, j) & \mathbf{r}_t(i+1, j+1) & \mathbf{r}_{t\tau}(i+1, j) & \mathbf{r}_{t\tau}(i+1, j+1) \end{pmatrix} \times \begin{pmatrix} H_0(v) \\ H_1(v) \\ H_2(v) \\ H_3(v) \end{pmatrix},$$

$$u = t - i; \quad v = \tau - j; \quad 0 \leq t \leq m; \quad 0 \leq \tau \leq n. \quad (7)$$

Здесь через i и j обозначены целые части параметров t и τ соответственно, u и v — местные параметры.