

АППРОКСИМАЦИЯ ОДНО-, ДВУ- И ТРЕХМЕРНЫХ ДУГ КРИВЫХ ПАРАМЕТРИЧЕСКИМИ СПЛАЙНАМИ

*Институт проблем математических машин и систем НАН Украины, Киев, Украина

Анотація. Обговорюється алгоритм обчислення коефіцієнтів параметричного сплайну, що апроксимує послідовності експериментальних даних. Наведені узагальнення, які дозволяють використовувати цей алгоритм для одно-, дво- та тривимірних випадків представлення експериментальних даних.

Ключові слова: контур, експериментальні дані, апроксимація, параметричний сплайн, класифікація.

Аннотация. Обсуждается алгоритм вычисления коэффициентов параметрического сплайна, аппроксимирующего последовательности экспериментальных данных. Приведены обобщения, позволяющие использовать данный алгоритм для одно-, дву- и трехмерного случаев представления экспериментальных данных.

Ключевые слова: контур, экспериментальные данные, аппроксимация, параметрический сплайн, классификация.

Abstract. We discuss the algorithm for calculating the coefficients of the parametric spline for approximating sequences of experimental data. We showed the generalizations, which allow using this algorithm for one-, two- and three-dimensional cases of the experimental data submission.

Keywords: circuit, experimental data, approximation, parametric spline, classification.

1. Введение

Очень часто исследователь сталкивается с задачей построения алгоритма классификации тех или иных объектов по их форме. Примеров таких задач можно привести множество. Например, анализ формы временного ряда, анализ формы спектрограмм, анализ формы элементов циклических сигналов в медицине, классификация бинарных и полутоновых изображений по форме элементов контуров этих изображений и т.д. Объединяет все эти задачи одно обстоятельство: все они, в конечном счете, могут быть сведены к задаче классификации формы одной или нескольких дуг кривых в одно-, дву- или трехмерном пространстве координат.

В целом ряде статей мы уже рассматривали подходы к решению таких задач при выделении признаков формы кривых, описывающих те или иные экспериментальные данные [1, 2]. В этой статье решено привести обобщение алгоритмического базиса при предварительной обработке данных, которые могут быть описаны одной или несколькими дугами кривых, для всех случаев пространств координат. Заметим, что описываемые алгоритмы относятся к этапу предварительной обработки исходных данных и построения системы признаков, чувствительных к форме объектов, для решения задачи классификации [3, 4]. Сама же классификация объектов по их форме в данной статье не рассматривается.

2. Постановка задачи

Итак, пусть в результате некоего физического процесса мы получаем дискретные экспериментальные данные $X = \{x(m)\}$, $m = 0, M$ для одномерного, $X = \{x(m)\}$, $Y = \{y(m)\}$ для двумерного и $X = \{x(m)\}$, $Y = \{y(m)\}$, $Z = \{z(m)\}$ для трехмерного случаев, которые исследователем воспринимаются как дискретные отсчеты некоей дуги кривой одномерной, плоской или пространственной. Исследователю важно классифицировать форму этой дуги

кривой, но закон ее построения в общем случае не известен.

Исходные экспериментальные данные могут быть заменены аналитическими кривыми, хорошо совпадающими по форме с кривой исходного контура и в то же время инвариантными относительно изменения масштаба, количества измерений, уровня помех. Такие кривые задаются функциями $x(k)$, $y(k)$, которые представляют собой полиномы r -ной степени от параметра k :

$$x(k) = a_0 k^r + a_1 k^{r-1} + \dots + a_{r-1} k + a_r, \quad (1)$$

$$y(k) = b_0 k^r + b_1 k^{r-1} + \dots + b_{r-1} k + b_r.$$

Выбор r -ной степени полиномов зависит от сложности аппроксимируемых контуров.

Реализация такого подхода позволяет отобразить все существенные особенности формы кривой, исключив влияние помех. Кроме того, вместо описаний контуров в пространстве сигналов можно рассматривать аппроксимирующие их полиномы в пространстве их коэффициентов. Существенным преимуществом такого описания является его инвариантность относительно количества измеренных сигналов в каждом массиве экспериментальных данных.

В качестве аппроксимирующего параметрического сплайна можно использовать кривую Безье, канонический сплайн (cardinal spline) [5]. В данной работе будет описано применение канонического сплайна третьего порядка.

Канонический сплайн – это последовательность полиномов третьего порядка, каждый из которых может быть записан параметрическим уравнением общего вида (2). Кривые, соответствующие полиномам, начинаются и заканчиваются в управляющих точках, образуя плавную односвязную кривую (рис. 1).

$$x(k) = a_x k^3 + b_x k^2 + c_x k + d_x, \quad (2)$$

$$y(k) = a_y k^3 + b_y k^2 + c_y k + d_y.$$

Для аппроксимации последовательности большого количества экспериментальных данных, являющихся реализацией кривых сложной формы, используют кардинальные кубические сплайны [5, 7]. Кардинальные кубические сплайны являются последовательностью N кубических полиномов, которые определяются управляющими точками $v = (x_n)$ или $v = \{x_n, y_n\}$, или $v = \{x_n, y_n, z_n\}$, $n = 1, N+1, N \ll M$. Каждое значение n в последовательности управляющих точек соответствует некоторому значению m_n в последовательности измерений. Будем рассматривать только корректные последовательности управляющих точек. Под корректной последовательностью будем понимать такую последовательность, при которой выполняется условие $m_n < m_{n+1}$. Количество полиномов в последовательности определяется сложностью формы аппроксимируемой экспериментальной кривой. Последовательность полиномов $X(t) = \{X_n(t_n)\}$, $Y(t) = \{Y_n(t_n)\}$, $Z(t) = \{Z_n(t_n)\}$ соответствует каждой из пространственных координат. Коэффициенты каждого из полиномов в последовательности, например, для координаты x : $X_n(k) = a_{nx} k^3 + b_{nx} k^2 + c_{nx} k + d_{nx}$, определяются значениями в управляющих точках $x_{n-1}, x_n, x_{n+1}, x_{n+2}$:

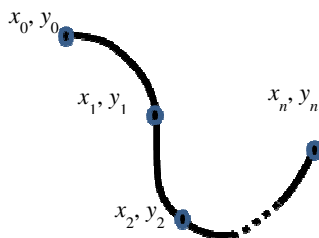


Рис. 1. Канонический сплайн

$$\begin{aligned}
a_{nx} &= T(x_{n+1} - x_{n-1}) + T(x_{n+2} - x_n) + 2x_n - 2x_{n+1}, \\
b_{nx} &= -2T(x_{n+1} - x_{n-1}) - T(x_{n+2} - x_n) - 3x_n + 3x_{n+1}, \\
c_{nx} &= T(x_{n+1} - x_{n-1}), \\
d_{nx} &= x_n,
\end{aligned} \tag{3}$$

где T – натяжение (при $T = 0$ кривая вырождается в прямую). Для большинства практических применений используют значение $T = 0,5$.

Для незамкнутых кривых при $n = 1: x_{n-1} = x_1$, при $n = N: x_{n+1} = x_N$.

Для аппроксимируемой последовательности экспериментальных данных находят такую кривую, которая наилучшим образом отображает ее форму, а это значит, что вычисляют координаты управляющих точек этой кривой и, следовательно, коэффициенты сплайна. В результате последовательность экспериментальных данных будет представлена вектором v , компонентами которого являются коэффициенты канонического сплайна (4).

В качестве меры сходства аппроксимирующего сплайна и последовательности экспериментальных данных будем использовать оценку площади фигуры, ограниченной контуром, образованным расчетной кривой и ломаной линией, последовательно соединяющей экспериментальные данные $S(G, P)$. Тогда поиск сплайна, наилучшим образом аппроксимирующего последовательность экспериментальных данных, можно рассматривать как вычисление оптимальных коэффициентов v_{opt} , при которых минимизируется значение оценки площади (5).

$$v = \{x_n, y_n\}, \tag{4}$$

$$v_{opt} = \arg \min_v S(G, P), \tag{5}$$

где G – последовательность экспериментальных данных $g_m(x_{gm}, y_{gm})$;

P – точки аппроксимирующего сплайна $p_m(x_{pm}, y_{pm})$;

$m = 0, M, M$ – количество отсчетов в последовательности экспериментальных данных.

Минимизация указанной меры сходства (5) достигается методом градиентного спуска [6].

Общий алгоритм итерационной процедуры аппроксимации экспериментальных данных параметрическими сплайнами с использованием итерационного метода градиентного спуска представлен на рис. 2. На старте алгоритма имеем произвольное приближение сплайна к последовательности экспериментальных точек. Оцениваем меру сходства и производим шаг градиентного спуска по одному из возможных направлений. Снова оцениваем меру сходства и делаем шаг в другом направлении градиентного спуска. Такая процедура выполняется до тех пор, пока величина изменения меры сходства становится меньше некоторого заранее заданного значения ε . Возможная ситуация с отсутствием сходимости итерационного алгоритма блокируется заранее заданным максимальным количеством итераций.

3. Аппроксимация одномерной последовательности экспериментальных данных

Одномерную последовательность получим в том случае, если экспериментальные данные будут представлены в виде последовательности измерений или таблицы, в которой с каждым отсчетом (или индексом в таблице) связано одно значение интересующей величины. Примером таких данных могут быть измерения температуры воздуха или атмосферного давления через равные промежутки времени, измерения оптической плотности через рав-

ные интервалы длин волн и т.п. Ось ординат тогда будет отображать интересующую величину, а ось абсцисс – номера отсчетов (или моменты времени измерений, длины волн для вышеприведенных примеров). Характерным для такой дуги является ее однозначность относительно оси абсцисс.

В этом случае вектор управляющих точек сплайна будет выглядеть как (6)

$$v \rightarrow \{y_n, t_n\}. \quad (6)$$

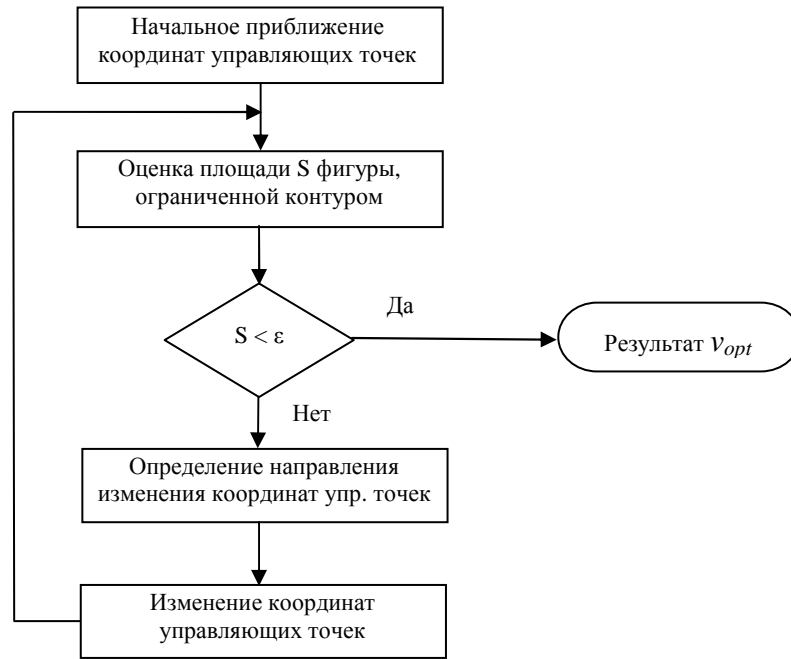


Рис. 2. Схема алгоритма аппроксимации экспериментальных данных параметрическими сплайнами

На рис. 3 представлена иллюстрация работы алгоритма аппроксимации одномерной последовательности экспериментальных данных в наглядном виде.

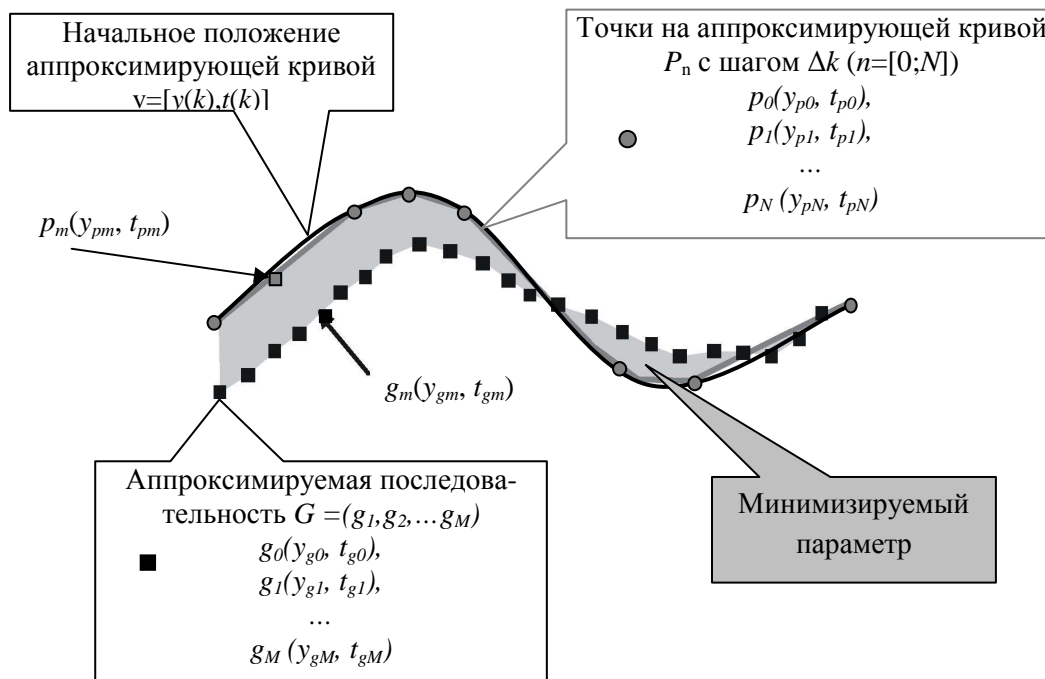


Рис. 3. Аппроксимация одномерной последовательности экспериментальных данных

4. Экспериментальная проверка алгоритма

Алгоритм был экспериментально проверен для более 1000 реальных графиков спектрограмм медицинских препаратов крови, полученных по методу «Онкотест» [2]. На рис. 4 представлены примеры аппроксимации спектрограмм. Как видно из рисунка, по оси абсцисс отображена длина волны, причем интервалы между отдельными отсчетами по этой оси равны, что позволяет заменить значения длины волны на номер отсчета и рассматривать спектрограмму как пример одномерной последовательности экспериментальных данных.

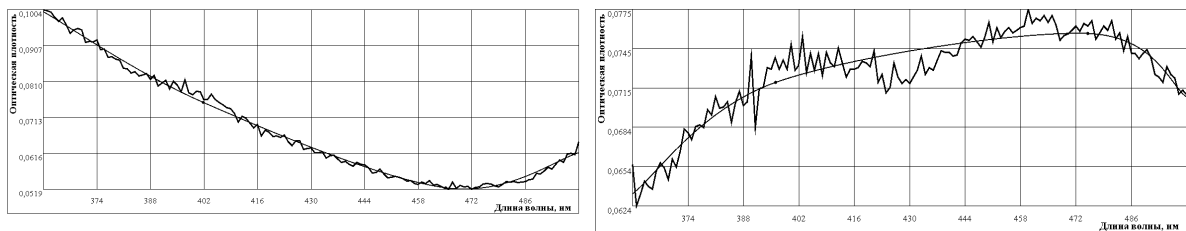


Рис. 4. Примеры аппроксимации спектрограмм каноническими сплайнами

5. Аппроксимация двумерной последовательности экспериментальных данных

Двумерную последовательность получим, если экспериментальные данные представлены в виде таблицы, в которой с каждым отсчетом (или индексом) связана пара значений интересующих величин. Как пример такой последовательности может быть приведен плоский контур (контур бинарного изображения).

В этом случае вектор управляющих точек сплайна будет выглядеть как (7)

$$v \rightarrow \{x_n, y_n, t_n\}. \quad (7)$$

На рис. 5 представлена иллюстрация работы алгоритма аппроксимации двумерной последовательности экспериментальных данных в наглядном виде.

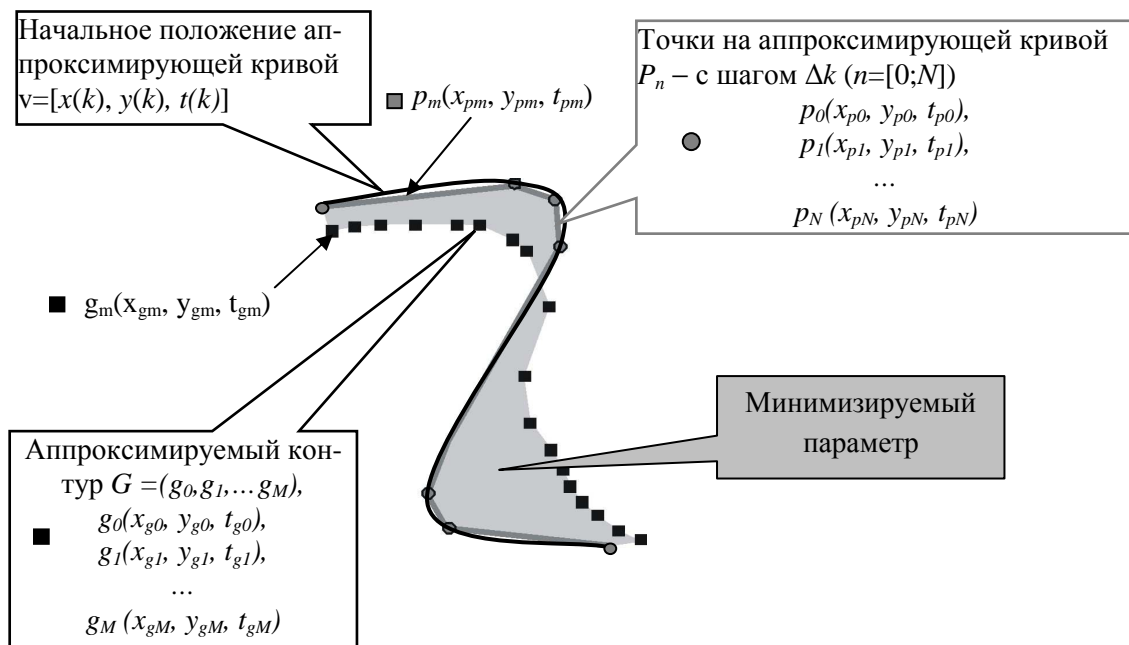


Рис. 5. Аппроксимация двумерной последовательности экспериментальных данных

6. Экспериментальная проверка алгоритма

Алгоритм был использован для аппроксимации двумерных контуров, некоторые результаты представлены на рис. 6. В данном случае каждый отсчет последовательности экспериментальных данных характеризуется двумя значениями – по оси абсцисс и по оси ординат, кроме того, такой контур не является односвязным по оси абсцисс, поэтому его можно рассматривать при аппроксимации как пример двумерной последовательности.

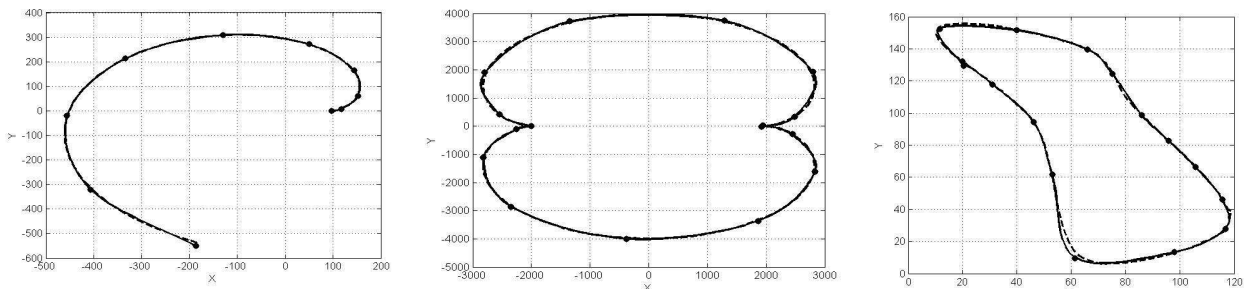


Рис. 6. Примеры аппроксимации двумерных контуров каноническими сплайнами

7. Аппроксимация трехмерной последовательности экспериментальных данных

Трехмерную последовательность получим, если экспериментальные данные представлены в виде таблицы, в которой с каждым отсчетом (или индексом) связана тройка значений интересующих величин. Как пример можно привести электрокардиограмму в трех ортогональных отведениях [8].

В этом случае вектор коэффициентов сплайна будет выглядеть как (8)

$$v \rightarrow \{x_n, y_n, z_n, t_n\}. \quad (8)$$

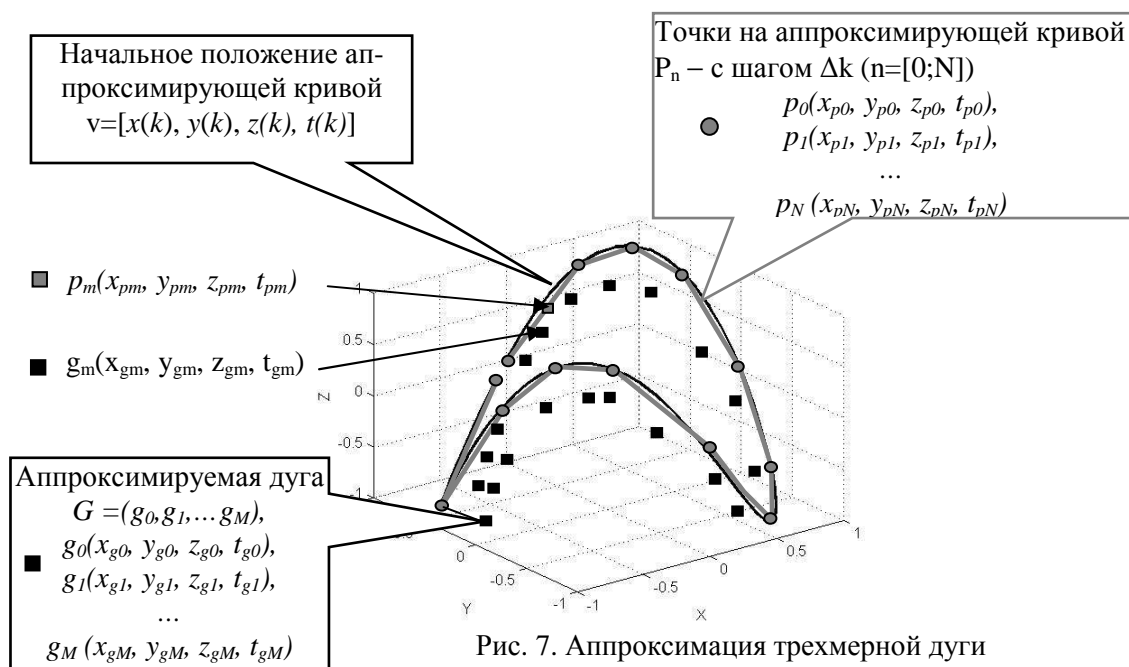


Рис. 7. Аппроксимация трехмерной дуги

На рис. 7 представлен алгоритм аппроксимации трехмерной последовательности экспериментальных данных (трехмерной дуги) в наглядном виде. В случае трехмерной последовательности каждый отсчет экспериментальных данных и каждая точка аппроксимирующего сплайна характеризуются тремя координатами. Иными словами, и эксперимен-

тальные данные, и аппроксимирующий сплайн представляют собой трехмерные дуги. Тогда в качестве меры их сходства нужно использовать оценку площади поверхности, ограниченной этими двумя пространственными кривыми.

8. Экспериментальная проверка алгоритма

Алгоритм был использован для аппроксимации электрокардиограмм в ортогональных отведениях. Так как каждый отсчет такой электрокардиограммы характеризуется тремя значениями, ее можно использовать при аппроксимации как пример трехмерной последовательности экспериментальных данных (трехмерной дуги).

На рис. 8 изображена часть электрокардиограммы в виде трехмерной дуги, построенная по трем ортогональным отведениям, и аппроксимирующий ее канонический сплайн. На рис. 9 приведены две аппроксимированные каноническими сплайнами электрокардиограммы в ортогональных отведениях для различных пациентов.

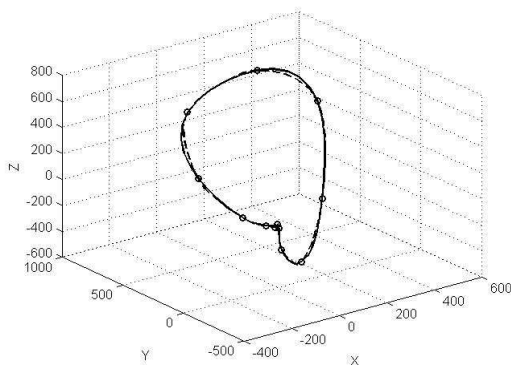


Рис. 8. Пример аппроксимации электрокардиограммы в ортогональных отведениях каноническим сплайном

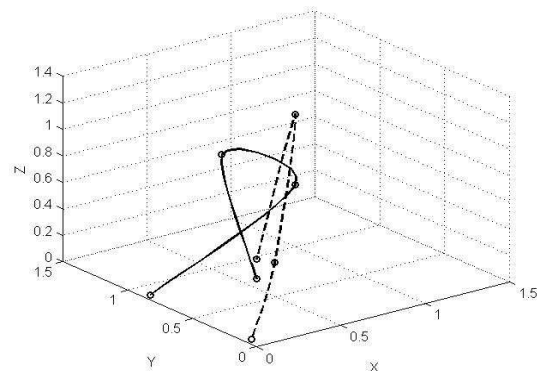


Рис. 9. Результаты аппроксимации электрокардиограмм в ортогональных отведениях двух различных пациентов

9. Заключение

Предложенный алгоритм позволяет вычислить коэффициенты параметрического сплайна, аппроксимирующего последовательности экспериментальных данных и может быть использован для одно-, дву- и трехмерного случаев представления исходных экспериментальных данных.

Применение данного алгоритма для предварительной обработки исходных данных позволяет построить систему признаков, чувствительных к форме объекта, которую удобно применять при решении задач классификации.

СПИСОК ЛИТЕРАТУРЫ

1. Вишневский В.В. Аппроксимация экспериментальных данных кривыми Безье / В.В. Вишневский, В.Г. Калмыков, Т.Н. Романенко // XIII-th International Conference KDS-2007. – Varna, Bulgaria, 2007. – June. – P. 3 – 9.
2. Вишневский В.В. Программно-аппаратный комплекс «Онкотест-WM-01» / В.В. Вишневский, В.А. Владимиров, Т.Н. Романенко // Тезисы доклада II международной школы-семинара «Телемедицина – опыт и перспективы». – Донецк, 2006. – Т. 4, № 2. – С. 62 – 65.
3. Вишневський В.В. Біометрична ідентифікація за допомогою електрокардіограми / В.В. Вишневський, Т.М. Романенко, Л.А. Кізуб // V міжнар. наук.-практ. конф. "Інформаційні технології та комп'ютерна інженерія" ІТКІ 2015. – Івано-Франківськ, 2015. – С. 130 – 131.

4. Вишневський В.В. Біометрична ідентифікація людини за її електрокардіограмою / В.В. Вишневський, Т.М. Романенко, Л.А. Кізуб // III міжнар. наук.-практ. конф. "Обчислювальний інтелект (результати, проблеми, перспективи)" ComInt 2015. – Черкаси, 2015. – С. 295.
5. Петцольд Ч. Программирование для Microsoft Windows на С#: в 2-х т. / Петцольд Ч.; пер. с англ. – М.: Русская редакция, 2002. – Т. 2. – С. 632 – 641.
6. Корн Г. Справочник по математике / Г. Корн, Т. Корн. – М.: Наука, 1974. – С. 660 – 661.
7. Vishnevsky V. Approximation of Planar and Spatial Experimental Curves by Splines that are Defined Parametrically / V. Vishnevsky, T. Romanenko, V. Kalmykov // Proc. of the international conference on applications of information and communication technology and statistics in economy and education ICAICTSEE-2014. – UNVE, Sofia, Bulgaria, 2014. – October 24–25th. – P. 236 – 240.
8. Орлов В.Н. Руководство по электрокардиографии / Орлов В.Н. – М.: Медицинское информационное агентство, 2012. – С. 58.

Стаття надішла до редакції 24.09.2015