# Аппроксимация временных рядов ломаной кратчайшей длины

М. Х. Мирхамидов

В настоящей работе рассматривается задача аппроксимации временных рядов ломаными кратчайшей длины с заданной точностью. Временной ряд интерпретируется последовательностью точек плоскости, а задача аппроксимации этой последовательности ломаной приводит к поиску кратчайшего пути между двумя отрезками. Получен алгоритм, решающий поставленную задачу с использованием построения графа видимости и алгоритма Дейкстры.

**Ключевые слова:** аппроксимация точек на плоскости, граф видимости, плоскость с многоугольными препятствиями, кратчайший путь, алгоритм Дейкстры.

## 1. Введение

Под временным рядом мы понимаем конечную последовательность действительных чисел. Алгоритм, полученный в настоящей работе, позволяет аппроксимировать временной ряд с заданной точностью ломанной кратчайшей длины. Таким образом, кодируя временной ряд последовательностью вершин полученной ломанной, можно добиться значительного сжатия исходной информации [1].

Рассматриваемая задача приводит к построению ломаной кратчайшей длины, проходящей через заданное множество отрезков равной длины и лежащих на параллельных прямых. Задача поиска кратчайшего пути между двумя точками на плоскости с многоугольными препятствиями рассматривалась, например, в [2]. Для решения рассматриваемой в этой работе задачи мы, используя сложившийся подход, строим некоторый вспомогательный граф и применяем алгоритм Дейкстры [3] для поиска кратчайшего пути между двумя вершинами в этом графе.

## 2. Определения, обозначения

На плоскости рассмотрим прямоугольную систему координат $Oxy$. Пусть задано положительное число $\varepsilon$. Имеются ломаные $l_1(A_1, A_n) = A_1, A_2, \ldots, A_n$ и $l_2(B_1, B_n) = B_1, B_2, \ldots, B_n$. Точка $A_i$ имеет координаты $(x_i, y_i + \varepsilon)$, а точка $B_i$ — координаты $(x_i, y_i - \varepsilon)$. При этом последовательность абсцисс точек $A_i$ строго возрастает, то есть $x_1 < x_2 < \ldots < x_n$.

Пусть точка $T = T(x, y)$ такова, что: $x_1 \leqslant x \leqslant x_n$. Тогда найдется $i$ такое, что $x_i \leqslant x \leqslant x_{i+1}$. Ординату точки отрезка $[A_i, A_{i+1}]$ с абсциссой $x$ обозначим $y'$,а ординату точки отрезка $[B_i, B_{i+1}]$ с абсциссой $x$ обозначим $y''$. Если при этом $y'' \leqslant y \leqslant y'$, то будем говорить, что точка $T$ лежит между ломаными $l_1(A_1, A_n)$ и $l_2(B_1, B_n)$.

Будем говорить, что фигура $O$ лежит между ломаными $l_1(A_1, A_n)$ и $l_2(B_1, B_n)$, если все точки этой фигуры лежат между $l_1(A_1, A_n)$ и $l_2(B_1, B_n)$. Пусть $Z$ — множество, состоящее из всех ломаных и, в частности, отрезков, лежащих между ломаными $l_1(A_1, A_n)$ и $l_2(B_1, B_n)$.

Множество вершин ломаной $l_1(A_1, A_n)$ обозначим через $P_1$, а множество вершин ломаной $l_2(B_1, B_n)$ — через $P_2$.

## 3. Постановка задачи и формулировка результатов

Требуется составить алгоритм, который по ломаным $l_1(A_1, A_n)$ и $l_2(B_1, B_n)$ строит ломаную кратчайшей длины, соединяющую отрезки $[A_1, B_1]$ и $[A_n, B_n]$ и содержащуюся в $Z$.

Пусть $S_1$ — множество перпендикуляров, опущенных из вершин ломаных $l_1(A_2, A_n)$ и $l_2(B_2, B_n)$ на отрезок $[A_1, B_1]$, а $S_2$ — множество перпендикуляров, опущенных из вершин ломаных $l_1(A_1, A_{n-1})$ и $l_2(B_1, B_{n-1})$ на отрезок $[A_n, B_n]$.

Введем следующие обозначения: $R_1 = \{T : T \in [A_1, B_1], [T, H] \in S_1\}, R_2 = \{T : T \in [A_n, B_n], [T, H] \in S_2\}$.

Определим следующий граф:

Граф $G = (V, W)$ называется графом видимости, если для множества $V$ его вершин и множества $W$ его ребер выполнены следующие равенства:

$V = P_1 \cup P_2 \cup R_1 \cup R_2 \cup \{M, K\}$, где $M, K \notin P_1 \cup P_2 \cup R_1 \cup R_2$
$W = W_1 \cup W_2 \cup W_3 \cup W_4 \cup W_5$, где
$W_1 = \{[A, B] : A \in [A_1, B_1], B \in (P_1 \cup P_2) \setminus \{A_1, B_1\}, [A, B] \perp [A_1, B_1],$
$[A, B] \in Z\},$
$W_2 = \{[A, B] : A \in [A_n, B_n], B \in (P_1 \cup P_2) \setminus \{A_n, B_n\}, [A, B] \perp [A_n, B_n],$
$[A, B] \in Z\},$
$W_3 = \{[A(x_a, y_a), B(x_b, y_b)] : A, B \in P_1 \cup P_2, x_a \neq x_b, [A, B] \in Z\},$
$W_4 = \{[M, B] : B \in R_1\},$
$W_5 = \{[A, K] : A \in R_2\};$

Будем считать, что вес любого ребра из множеств $W_4$ и $W_5$ равен нулю, а веса ребер из $W_1 \cup W_2 \cup W_3$ равны длине соответствующих отрезков.

Построив граф видимости $G$, используя алгоритм Дейкстры, найдем кратчайший путь между вершинами $M$ и $K$. Далее покажем, что ломаная, соответствующая этому пути, является кратчайшей среди всех непрерывных спрямляемых кривых, соединяющих отрезки $[A_1, B_1]$ и $[A_n, B_n]$.

**Теорема 1.** *Для любой спрямляемой непрерывной кривой, соединяющей отрезки $[A_1, B_1]$ и $[A_n, B_n]$ и содержащейся в $Z$, существует ломаная с ребрами из $W$, соединяющая отрезки $[A_1, B_1]$ и $[A_n, B_n]$, длина которой не более, чем длина исходной кривой.*

**Доказательство.** Пусть имеется некоторая кривая L, соединяющая отрезки $[A_1, B_1]$ и $[A_n, B_n]$ и содержащаяся в $Z$. Покажем, что ее можно заменить ломаной, соединяющей отрезки $[A_1, B_1]$ и $[A_n, B_n]$, звенья которой принадлежат $W$ и что длина этой ломаной будет не больше, чем длина $L$.

Пусть $P = \{A_2, A_3, \ldots, A_{n-1}\} \cup \{B_2, B_3, \ldots, B_{n-1}\}$.

Шаг 1.

Пусть $E$ — множество точек, образующих кривую $L$. Обозначим через $R = \{[D', D''] | D' \in [A_1, B_1] \cap L, D'' \in E, D' \neq D''\}$ — множество таких отрезков, что один конец $D'$ лежит на $[A_1, B_1] \cap L$, а другой конец $D''$ лежит на $L$. Тогда если в $R$ содержится отрезок $[D', D''] \in Z$ такой, что $D'' \in [A_n, B_n] \cap L$, то отрезок $[D', D'']$ будет не длиннее, чем кривая $L$ (так как минимальное расстояние между двумя точками равно длине отрезка, соединяющего их).

Иначе, $\exists D \in E : [D', D] \in R \cap Z$ и $[D', D]$ содержит элемент $H_1 \in P$. Тогда рассмотрим новую кривую $L' = [D', D] \cup L(D', D)$, где $L(D', D)$ — часть кривой $L$, заключенной между точками $D'$ и D. Очевидно, что $|L'| \leqslant |L|$.

Шаг 2.

Пусть $E'$ — множество точек, образующих часть кривой $L'(H_1, D'')$, где $D'' \in [A_n, B_n] \cap L'$. Обозначим через $R = \{[D', D''] | D' = H_1, D'' \in E', D' \neq D''\}$ — множество таких отрезков, что один конец $D' = H_1$, а другой конец $D''$ лежит на $L'$. Тогда если в $R$ содержится отрезок $[D', D''] \in Z$ такой, что $D'' \in [A_n, B_n] \cap L'$, то ясно, что отрезок $[D', D'']$ будет не длиннее, чем $|L'(D', D'')|$, где $L'(D', D'')$ — часть кривой $L'$, заключенной между точками $D'$ и $D''$. Тогда через $L''$ обозначим кривую $L' \setminus L'(D', D'') \cup [D', D'']$. Так как $|[D', D'']| \leqslant |L'(D', D'')|$, то $|L''| \leqslant |L'|$.

Иначе, $\exists D \in E' : [D', D] \in R \cap Z$ и $[D', D]$ содержит элемент $H_2 \in P \setminus H_1$. Тогда рассмотрим новую кривую $L'' = L' \setminus L'(D', D) \cup [D', D]$. Так как $|[D', D]| \leqslant |L'(D', D)|$, то $|L''| \leqslant |L'|$.

Продолжая выполнять эту процедуру, на шаге $m$ получим:

Шаг $m$.

Пусть $E^{(m-1)}$ — множество точек, образующих часть кривой $L^{(m-1)}(H_{m-1}, D'')$, где $D'' \in [A_n, B_n] \cap L^{(m-1)}$. Обозначим через $R = \{[D', D''] | D' = H_{m-1}, D'' \in E^{(m-1)}, D' \neq D''\}$ — множество таких отрезков, что один конец $D' = H_{m-1}$, а другой конец $D''$ лежит на $L^{(m-1)}$. Тогда если в R содержится отрезок $[D', D''] \in Z$ такой, что $D'' \in [A_n, B_n] \cap L^{(m-1)}$, то ясно, что отрезок $[D', D'']$ будет не длиннее, чем $|L^{(m-1)}(D', D'')|$, где $L^{(m-1)}(D', D'')$ — часть кривой $L^{(m-1)}$, заключенной между точками $D'$ и $D''$. Тогда через $L^m$ обозначим кривую $L^{(m-1)} \setminus L^{(m-1)}(D', D'') \cup [D', D'']$. Так как $|[D', D'']| \leqslant |L^{(m-1)}(D', D'')|$, то $|L^m| \leqslant |L^{(m-1)}|$.

Иначе, $\exists D \in E^{(m-1)} : [D', D] \in R \cap Z$ и $[D', D]$ содержит элемент $H_m \in P \setminus \cup_{s=1}^{m-1} H_s$. Тогда рассмотрим новую кривую $L^m = L^{(m-1)} \setminus L^{(m-1)}(D', D) \cup [D', D]$. Так как $|[D', D]| \leqslant |L'(D', D)|$, то $|L^m| \leqslant |L^{(m-1)}|$.

Очевидно, что число шагов рассмотренной процедуры не превосходит $n - 1$.

Итак, получили ломаную, длина которой не превосходит длины изначальной кривой, а все вершины (быть может, только кроме конечных) принадлежат множеству $P$. Поэтому все звенья полученной ломаной, кроме конечных, принадлежат множеству $W$.

Покажем, что конечные звенья можно заменить ломаными, ребра которых принадлежат $W$ и длины которых не будут превышать длин соответствующих звеньев.

**Случай 1 (исходная кривая $L$ заменена ломаной, состоящей из двух или более отрезков).**

Рассмотрим только последнее звено (для первого звена рассуждения проводятся аналогично).

Шаг 1.

Пусть $[M, N]$ — последний отрезок ломаной, построенной в результате приведенной выше процедуры, причем $M = (x_M, y_M) \in P$, $N = (x_N, y_N) \in [A_n, B_n], A_n = (x_A, y_A), B_n = (x_B, y_B)$.

Точку $M$ назовем текущей.

Пусть $Q = \{T : T \in [A_n, B_n]$ и $[M, T] \in Z\}$. Очевидно, что $Q \neq \varnothing$. Далее положим:

$y_{max} = \max\{y(T)|T \in Q, y$ — ордината точки $T\}$,
$y_{min} = \min\{y(T)|T \in Q, y$ — ордината точки $T\}$
Тогда

1) если $y_M < y_{min}$ и $y_{min} = y_B$, то строим отрезок $[M, B_n] \in W$; если $y_M < y_{min}$ и $y_{min} > y_B$, то положим: $Q_{min} = (x_B, y_{min})$. $\exists H \in \{B_2, B_3, \ldots, B_{n-1}\} \cap [M, Q_{min}]$. Так как $y_{min} \leqslant y_N$, то $|[M, Q_{min}]| \leqslant |[M, N]|$. Строим отрезок $[M, Q_{min}]$. Но так как $[M, Q_{min}]$ может не содержаться в $W$, то за текущую точку возьмем $H$ и на следующем шаге рассмотрим отрезок $[H, Q_{min}]$.
2) если $y_{min} \leqslant y_M \leqslant y_{max}$, то $\exists T : T \in Q, [M, T] \perp [A_n, B_n]$. Так как $[M, T]$ — отрезок минимальной длины, соединяющий $M$ и $[A_n, B_n]$, то заменим отрезок $[M, N]$ построенной ломанной на отрезок $[M, T]$. Тогда получим ломаную длины не большей, чем длина предыдущей с последним звеном из $W$.
3) если либо $y_M > y_{max}$ и $y_{max} = y_A$, либо $y_M > y_{max}$ и $y_{max} < y_A$, то поступаем аналогично пункту 1.

Шаг 2.

Повторяем предыдущий шаг, но для текущей точки $H$.

Повторяя приведенную процедуру, получаем требуемую ломаную.

Так как число шагов не может превосходить $n - 1$, то данный процесс завершится за конечное число шагов, причем все звенья построенной ломаной будут содержаться в $W$.

**Случай 2 (кривая $L$ заменена ломаной, состоящей из одного отрезка).**

Пусть $[M, N]$ — отрезок, заменивший исходную кривую, причем $M = (x_M, y_M) \in [A_1, B_1], N = (x_N, y_N) \in [A_n, B_n], A_n = (x_A, y_A)$, $B_n = (x_B, y_B)$.

Будем двигать вверх (вниз) отрезок $[M, N]$ параллельно до тех пор, пока хотя бы один элемент $H \in \{A_1, \ldots, A_n\} \cup \{B_1, \ldots, B_n\}$ не будет содержаться в отрезке $[M, N]$. Тогда, взяв за текущую точку $H$, сведем к первому случаю.

Таким образом, мы доказали, что любую кривую $L \in Z$, соединяющую отрезки $[A_1, B_1]$ и $[A_n, B_n]$, можем заменить ломаной, длина которой будет не больше длины кривой $L$, причем ее звенья $\in W$. Теорема доказана.

**Следствие 1.** *Длина кратчайшего пути, соединяющего отрезки $[A_1, B_1]$ и $[A_n, B_n]$ и проходящего через все промежуточные отрезки $[A_i, B_i], i = \overline{2, n-1}$, равна длине кратчайшего пути в графе видимости, соединяющего точки $M$ и $K$.*

Автор выражает благодарность Анатолию Александровичу Часовских за научное руководство.

## Список литературы

[1] Сэломон Д. Сжатие данных, изображения и звука. — М.: Техносфера, 2004.

[2] Lee D. T., Preparata F. P. Euclidean shortest paths in the presence of rectilinear barriers // Networks 14. — 1984.

[3] Dijkstra E. W. A note on two problems in connexion with graphs // Numerische Mathematik. — 1959. V. 1. — P. 269–271.