

**ОБЩИЕ
ЧИСЛЕННЫЕ МЕТОДЫ**

УДК 519.16

**ЗАДАЧА МИНИМИЗАЦИИ СУММЫ РАЗНОСТЕЙ
ВЗВЕШЕННЫХ СВЕРТОК¹⁾**

© 2020 г. **А. В. Кельманов^{1,2}**, **Л. В. Михайлова^{1,*}**,
П. С. Рузанкин^{1,2,}**, **С. А. Хамидуллин^{1,***}**

¹ 630090 Новосибирск, пр-т Акад. Коптюга, 4, Ин-т матем. им. С.Л. Соболева, Россия

² 630090 Новосибирск, ул. Пирогова, 2, Новосибирский гос. ун-т, Россия

*e-mail: mikh@math.nsc.ru

**e-mail: ruzankin@math.nsc.ru

***e-mail: kham@math.nsc.ru

Поступила в редакцию 30.07.2019 г.
Переработанный вариант 30.07.2019 г.
Принята к публикации 04.08.2020 г.

В работе рассматривается неизученная экстремальная задача суммирования элементов числовых последовательностей Y длины N и U длины $q \leq N$. В задаче требуется минимизировать сумму разностей взвешенных свертки последовательностей переменной длины (не менее q). В каждой разности первая невзвешенная свертка – автосвертка растянутой на переменную длину последовательности U (путем кратных повторов ее элементов), вторая – взвешенная свертка этой растянутой последовательности с подпоследовательностью из Y . Анализируется вариант задачи с оптимизируемым числом суммируемых разностей. Показано, что задача эквивалентна одной из проблем аппроксимации последовательности Y элементом X из экспоненциального по мощности множества последовательностей. Это множество объединяет все последовательности длины N , которые в качестве подпоследовательностей включают переменное число M допустимых квазипериодических (флуктуационных) повторов последовательности U . Каждый квазипериодический повтор порождается допустимыми преобразованиями последовательности U . Этими преобразованиями являются: 1) сдвиг U на переменную величину, которая между соседними повторами не превышает $T_{\max} \leq N$, 2) переменное растягивающее отображение U в последовательность переменной длины, которое определяется в виде повторов элементов из U , кратность этих повторов – переменная величина. Критерием аппроксимации является минимум суммы квадратов расстояний между элементами последовательностей. Доказано, что рассматриваемая экстремальная задача и вместе с ней задача аппроксимации разрешимы за полиномиальное время. А именно, показано, что существует точный алгоритм, который находит решение задачи за время $\mathcal{O}(T_{\max}^3 N)$. Если T_{\max} – фиксированный параметр задачи, то время работы алгоритма линейно. Примерами численного моделирования проиллюстрирована применимость алгоритма к решению модельных прикладных задач помехоустойчивой обработки ECG-подобных и PPG-подобных квазипериодических сигналов (electrocardiogram-like and photoplethysmogram-like signals). Библ. 13. Фиг. 4.

Ключевые слова: числовые последовательности, разность взвешенных свертки, переменная длина свертки, минимум суммы, полиномиальная разрешимость, линейно-временной алгоритм, численное моделирование, ECG-подобный сигнал, PPG-подобный сигнал.

DOI: 10.31857/S0044466920120054

1. ВВЕДЕНИЕ

В работе рассматривается неизученная экстремальная задача суммирования элементов числовых последовательностей. Цель работы – анализ вычислительной сложности задачи и обоснование алгоритма, обеспечивающего решение задачи с доказуемыми оценками качества (точно-

¹⁾ Работа выполнена при финансовой поддержке РФФИ (19-07-00397 и 19-01-00308), программы ФНИ РАН, проект 0314-2019-0015, а также программы Тор-5-100 Минобрнауки РФ.

сти и трудоемкости). Исследование мотивировано новизной математической задачи и отсутствием каких-либо эффективных (полиномиальных) вычислительных алгоритмов с априорно гарантированными оценками точности для ее решения. Практическая мотивировка обусловлена важностью задачи, в частности, для помехоустойчивого дистанционного мониторинга природных объектов, состояние которых квазипериодически флуктуирует во времени (см. разд. 3, 7).

Статья имеет следующую структуру. В разд. 2 дана формулировка задачи, приведены ее истоки и некоторые интерпретации. В следующем разделе рассмотрены частные случаи задачи и наиболее близкие по постановке задачи. В разд. 4 анализируется размер (мощность) множества допустимых решений задачи. Вспомогательные утверждения, необходимые для алгоритмического решения задачи, доказаны в разд. 5. Основной результат работы – точный полиномиальный алгоритм – представлен и обоснован в разд. 6. В разд. 7 на примерах численного моделирования продемонстрирована применимость алгоритма к решению задач помехоустойчивой обработки ECG- и PPG-подобных сигналов.

2. ФОРМУЛИРОВКА ЗАДАЧИ, ЕЕ ИСТОКИ И ТРАКТОВКИ

Рассматриваемая экстремальная задача имеет следующую формулировку.

Задача 1. Дано: числовые последовательности $Y = (y_1, \dots, y_N)$, $U = (u_1, \dots, u_q)$, и натуральные числа T_{\max} , ℓ . Найти: набор $\mathcal{M} = \{n_1, \dots, n_m, \dots\}$ номеров последовательности Y , набор $\mathcal{P} = \{p^{(1)}, \dots, p^{(m)}, \dots\}$ натуральных чисел, набор $\mathcal{J} = \{J^{(1)}, \dots, J^{(m)}, \dots\}$ сжимающих отображений, в котором $J^{(m)} : \{1, \dots, p^{(m)}\} \rightarrow \{1, \dots, q\}$, а также размерность M этих наборов, которые минимизируют целевую функцию

$$F(\mathcal{M}, \mathcal{P}, \mathcal{J}) = \sum_{m=1}^M \sum_{i=1}^{p^{(m)}} \{u_{J^{(m)}(i)}^2 - 2y_{n_m+i-1}u_{J^{(m)}(i)}\}, \quad (2.1)$$

при ограничениях

$$q \leq p^{(m)} \leq \ell \leq T_{\max} \leq N, \quad m = 1, \dots, M, \quad (2.2)$$

$$p^{(m-1)} \leq n_m - n_{m-1} \leq T_{\max}, \quad m = 2, \dots, M, \quad (2.3)$$

на элементы искомых наборов \mathcal{M} , \mathcal{P} и при ограничениях

$$\begin{aligned} J^{(m)}(1) &= 1, \quad J^{(m)}(p^{(m)}) = q, \\ 0 \leq J^{(m)}(i) - J^{(m)}(i-1) &\leq 1, \quad i = 2, \dots, p^{(m)}, \\ m &= 1, \dots, M, \end{aligned} \quad (2.4)$$

на элементы искомых сжимающих отображений.

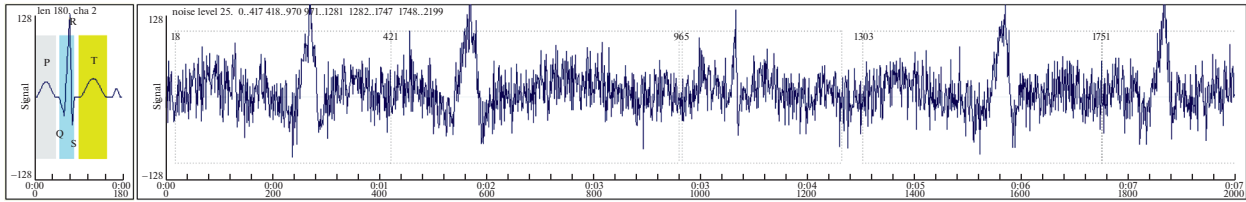
Приведем несколько трактовок задачи 1 и ее истоки. Далее будем считать, что номерам элементов входных последовательностей Y и U соответствуют дискретно-временные отсчеты непрерывных сигналов. На фиг. 1 в виде графиков изображены пример последовательности U (слева) и пример последовательности Y (справа). Раскрашенные участки последовательности U нам понадобятся для одной из интерпретаций задачи.

Очевидно, что задачу 1 можно трактовать как задачу об оптимальном (в смысле минимума (2.1)) суммировании элементов двух последовательностей. Эта трактовка видна непосредственно из правой части формулы (2.1), в которой оптимизируемыми переменными являются индексы только двух последовательностей U и Y .

Целевую функцию (2.1) можно записать в эквивалентном виде:

$$F(\mathcal{M}, \mathcal{P}, \mathcal{J}) = \sum_{m=1}^M \left\{ \sum_{i=1}^{p^{(m)}} u_{J^{(m)}(i)}^2 - 2 \sum_{i=1}^{p^{(m)}} y_{n_m+i-1} u_{J^{(m)}(i)} \right\}.$$

В правой части этой формулы в скобках видим: автосвертку последовательности длины $p^{(m)}$, образованной из элементов U , за вычетом удвоенной свертки этой последовательности с подпо-



Фиг. 1. Пример входных последовательностей задачи 1: последовательности U слева и Y справа.

следовательностью из Y длины $p^{(m)}$. Коэффициент 2 – вес этой свертки. Поэтому задачу 1 можно трактовать как сумму разностей взвешенных сверток.

Наконец, задачу 1 можно интерпретировать как поиск оптимального разбиения целочисленного отрезка $[1, N]$ на $M + 1$ интервалов, границы которых определяются неравенствами $1 \leq n_1 < \dots < n_m < \dots < n_M \leq N$.

Источником задачи 1 является задача аппроксимации последовательности Y последовательностью $X \in \mathcal{X}$, где \mathcal{X} – множество допустимых аппроксимирующих последовательностей, по критерию минимума суммы квадратов расстояний между элементами Y и X . Предполагается, что каждый элемент множества \mathcal{X} порождается допустимыми квазипериодическими (флуктуационными) повторами $m = 1, \dots, M$, последовательности U . При этом каждый квазипериодический повтор порождается допустимыми преобразованиями последовательности U . Этими преобразованиями являются: 1) сдвиг U на переменную величину, которая между соседними повторами не превышает $T_{\max} \leq N$, 2) переменное растягивающее отображение U в последовательность переменной длины, которое определяется в виде повторов элементов из U , причем кратность каждого из этих повторов – переменная величина.

Рассмотрим следующую модель порождения последовательностью U последовательности X и покажем, что эта модель и квадратичный критерий аппроксимации индуцируют задачу 1. Иными словами, покажем, что задачу 1 можно интерпретировать как задачу приближения по упомянутому критерию.

Обозначим через $k_t^{(m)}$, $t = 1, \dots, q$, $m = 1, \dots, M$, кратность повтора t -го элемента из U в ее m -м растянутом повторе

$$\underbrace{(u_1, \dots, u_1)}_{k_1^{(m)}}, \underbrace{(u_2, \dots, u_2)}_{k_2^{(m)}}, \dots, \underbrace{(u_q, \dots, u_q)}_{k_q^{(m)}} \tag{2.5}$$

длины

$$p^{(m)} = k_1^{(m)} + \dots + k_q^{(m)}, \tag{2.6}$$

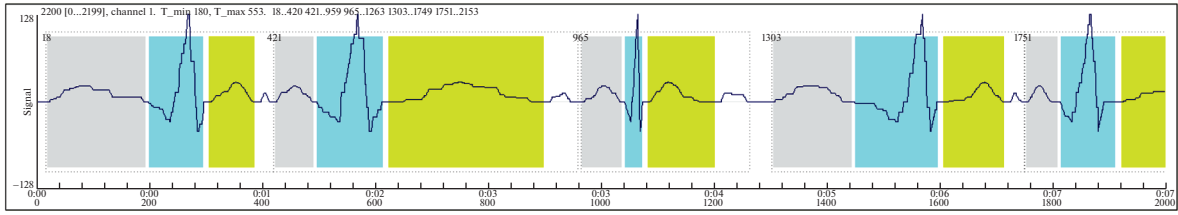
где $p^{(m)} \in \mathcal{P}$, положим

$$h_i^{(m)} = \begin{cases} u_1, & \text{если } i = 1, \dots, k_1^{(m)}, \\ u_2, & \text{если } i = k_1^{(m)} + 1, \dots, k_1^{(m)} + k_2^{(m)}, \\ \dots & \\ u_q, & \text{если } i = k_1^{(m)} + \dots + k_{q-1}^{(m)} + 1, \dots, k_1^{(m)} + \dots + k_q^{(m)}, \\ 0, & \text{если } i < 1, \quad i > p^{(m)}, \end{cases} \tag{2.7}$$

и определим n -й элемент последовательности $X = (x_1, \dots, x_N)$ по следующей формуле:

$$x_n = \sum_{m=1}^M h_{n-n_m+1}^{(m)}, \quad n = 1, \dots, N. \tag{2.8}$$

Формула (2.8) есть сумма M растянутых последовательностей вида (2.7). В этой формуле $n = n_m$, $n_m \in \mathcal{M}$, позиционирует начальный номер m -го растянутого повтора в последовательности X .



Фиг. 2. Пример последовательности X , порожденной последовательностью U и ее допустимыми преобразованиями.

Легко видеть, что при каждом m образ множества $\{1, \dots, p^{(m)}\}$ при отображении $J^{(m)}$ можно представить схематично в виде

$$\underbrace{1, \dots, 1}_{k_1^{(m)}} \underbrace{2, \dots, 2}_{k_2^{(m)}} \dots \underbrace{q, \dots, q}_{k_q^{(m)}}.$$

Поэтому для кратности $k_i^{(m)}$ в формуле (2.7) справедливо равенство

$$k_i^{(m)} = \left| \left\{ i \mid J^{(m)}(i) = t, i \in \{1, \dots, p^{(m)}\} \right\} \right|, \quad t = 1, \dots, q, \tag{2.9}$$

которое устанавливает связь между индексами растянутой последовательности $h_i^{(m)}$ и индексами последовательности U . При этом формулу (2.7) можно записать в эквивалентном кратком виде

$$h_i^{(m)} = \begin{cases} u_{J^{(m)}(i)}, & \text{если } i = 1, \dots, p^{(m)}, \\ 0, & \text{если } i < 1, \quad i > p^{(m)}, \end{cases} \tag{2.10}$$

$m = 1, \dots, M,$

с помощью сжимающего отображения $J^{(m)}$, которое определяется формулой (2.4).

В соответствии с (2.7) и (2.8) участок последовательности X с номерами элементов от $n = n_{m-1}$ до $n = n_m$, соответствующий $(m - 1)$ -му растянутому повтору, можно представить схематично в следующем виде (в нижней строке – номера элементов последовательности X):

$$\begin{array}{ccccccc} \dots, & \underbrace{0, u_1, \dots, u_1, \dots, u_1}_{k_1^{(m-1)}}, & \underbrace{u_2, \dots, u_2, \dots, u_2}_{k_2^{(m-1)}}, & \dots, & \underbrace{u_q, \dots, u_q}_{k_q^{(m-1)}}, & 0, \dots, 0, & u_1 \dots \\ \dots, & n_{m-1}, & \dots, & \dots, & \dots, & n_m \dots \end{array}$$

Пример последовательности X приведен на фиг. 2 в виде графика. На этой фигуре представлены 5 растянутых повторов, которые порождены последовательностью U , изображенной на фиг. 1. Кратности повторений элементов из U и интервалы между двумя растянутыми повторами генерировались датчиком случайных чисел. Чередующимися раскрашенным участкам последовательности U на фиг. 1 соответствуют чередующиеся растянутые (по отношению к участкам из U) раскрашенные участки в последовательности X на фиг. 2. Ступенчатость сигнала на этой фигуре обусловлена кратными повторениями элементов из U .

Поясним смысл ограничений (2.2) и (2.3) в формулировке задачи 1. В этих формулах T_{\max} обозначает верхнюю границу на интервал между двумя соседними растянутыми повторами, ℓ обозначает верхнюю границу на длину $p^{(m)}$ растяжения внутри каждого из этих интервалов, q устанавливает нижнюю границу на длину допустимых растяжений. Нижняя граница $p^{(m-1)}$ в формуле (2.3) означает, что каждый интервал между двумя последовательными растянутыми повторами не может быть меньше, чем длина растянутой последовательности из этого интервала. В целом формулы (2.2) и (2.3) устанавливают соотношение верхних и нижних границ для допустимых порождающих преобразований в виде сдвигов и растяжений последовательности U .

Для суммы квадратов расстояний между элементами Y и X с учетом (2.8) и (2.10) путем несложных преобразований получим равенство

$$\sum_{n=1}^N (x_n - y_n)^2 = \sum_{n=1}^N y_n^2 + \sum_{m=1}^M \sum_{i=1}^{p^{(m)}} \{u_{J^{(m)}(i)}^2 - 2y_{n_m+i-1}u_{J^{(m)}(i)}\}.$$

Первая сумма в правой части этого равенства не зависит от каких-либо переменных задачи 1, а вторая – совпадает с (2.1). Поэтому минимизация суммы квадратов расстояний (аппроксимационная задача) эквивалентна сформулированной задаче 1.

3. БЛИЗКИЕ ПО ПОСТАНОВКЕ ЗАДАЧИ

Приложениями задачи 1 являются проблемы обработки, анализа и интерпретации дискретно-временных сигналов, типичных для мониторинга состояний наземных, подземных, подводных, аэрокосмических, медицинских и других объектов. Эта задача, очевидно, является подходящей математической моделью для решения прикладных проблем из отмеченных областей в тех случаях, когда сигналы имеют квазипериодическую структуру, в виде флуктуирующих повторов участков сигнала (например, импульсов). Поэтому задачу 1 можно трактовать в терминах каждого из отмеченных приложений.

Задача 1 является обобщением ранее исследованной задачи дискретной оптимизации. Частный случай задачи 1, в котором $p^{(m)} = q$, $J^{(m)}(i) = i$, при каждом $m = 1, \dots, M$, исследовался в [1]–[3]. В [2] и [3] обоснован точный алгоритм с временем работы $\mathcal{O}(T_{\max}N)$.

В [1] исследовался вариант указанного частного случая, в котором M – часть входа (не является оптимизируемой переменной). В цитируемой работе построен алгоритм, гарантирующий отыскание оптимального решения за время $\mathcal{O}(T_{\max}MN)$. Этот вариант частного случая задачи 1 послужил подходящей моделью, в частности, для решения прикладных проблем геофизического и аэрокосмического мониторинга подземных и космических объектов. Алгоритм с несущественными модификациями программно-инженерного плана был дважды переизложен в [4] и [5], а затем использован в центре космических исследований NASA. Он, в частности, позволил обнаружить не одну сотню новых экзопланет по сильно зашумленным астрофизическим данным [5], [6]. Авторы цитируемых работ отмечают, что известные подходы к решению такого рода прикладных мониторинговых проблем оказались существенно менее результативными в плане помехоустойчивости. Другим результативным применением алгоритма является решение обратных геофизических задач, связанных с проблемами выявления сейсмических источников, а также прогнозированием техногенных катастроф [7].

Что касается приложений задачи 1, то для нее в качестве иллюстрации в разд. 7 приведены лишь два примера к обработке медицинских сигналов, имеющих вид квазипериодической последовательности импульсов. Как известно, в биомедицинских импульсных последовательностях интервалы между повторами импульсов и форма самих импульсов в повторях сопровождаются возможными флуктуациями, обусловленными естественными для природы квазипериодическими процессами.

4. РАЗМЕР МНОЖЕСТВА ДОПУСТИМЫХ РЕШЕНИЙ

Оценим размер (мощность) множества \mathcal{X} , т.е. число допустимых решений задачи 1.

Заметим, что значение M , очевидно, ограничено сверху, т.е. $M \leq M_{\max} \leq N$, где $M_{\max} = \left\lfloor \frac{N}{q} \right\rfloor$ – максимально возможное число повторов последовательности длины не менее q в последовательности длины $N \geq q$. Из комбинаторных соображений имеем оценку

$$|\mathcal{X}| \leq (T_{\max} - q + 1)q^{T_{\max}-q} M_{\max} (N - q + 1) (T_{\max} - q + 1)^{M_{\max}-1}. \tag{4.11}$$

В правой части этого неравенства произведение первых двух множителей соответствует мощности множества вариантов выбора растяжения последовательности U . Эти два множителя, умноженные на M_{\max} , определяют мощность совокупности допустимых вариантов растяжений последовательности длины не менее q в последовательности длины N . Четвертый множитель соответствует мощности множества вариантов выбора расположения в последовательности

длины N первой из M подпоследовательности длины q . Наконец, пятый множитель соответствует мощности множества вариантов расположения в последовательности длины N оставшихся $(M_{\max} - 1)$ подпоследовательностей длины не менее q .

Из анализа (4.11) ясно, что при фиксированном q имеется

$$|\mathcal{X}| = \mathcal{O}(q^{T_{\max}} N^2 (T_{\max})^{O(N)})$$

допустимых решений задачи 1, среди которых требуется найти наилучшее. Ясно также, что даже при $q = 1$ и $T_{\max} = 2$ перебрать напрямую элементы этого экспоненциального по мощности множества \mathcal{X} допустимых решений за приемлемое время вряд ли удастся, т.к. N — часть входа задачи.

5. ОСНОВЫ АЛГОРИТМА

Ниже мы формулируем две вспомогательные экстремальные задачи и доказываем утверждения, гарантирующие отыскание их точных алгоритмических решений. Эти задачи и утверждения необходимы для обоснования алгоритма решения задачи 1.

Нам потребуется следующая вспомогательная

Задача 2. Дано: таблица $\{w_{i,j}, i = 1, \dots, p, j = 1, \dots, q\}$ вещественных чисел. Найти: индексы суммирования, доставляющие минимум функции

$$W = \sum_{i=1}^p w_{i,J(i)}$$

при ограничениях

$$\begin{aligned} J(1) &= 1, & J(p) &= q, \\ 0 &\leq J(i) - J(i-1) \leq 1, & i &= 2, \dots, p, \end{aligned} \tag{5.12}$$

на искомые индексы.

Справедлива следующая

Лемма 1. Минимум целевой функции W в задаче 2 находится по формуле

$$W^* = W_{p,q}, \tag{5.13}$$

а значение $W_{p,q}$ находится в результате вычислений по рекуррентной формуле

$$W_{s,t} = \min \{W_{s-1,t}, W_{s-1,t-1}\} + w_{s,t}, \quad s = 1, \dots, p, \quad t = 1, \dots, q, \tag{5.14}$$

при начальном и граничных условиях

$$W_{s,t} = \begin{cases} 0, & s = 0, \quad t = 0, \\ +\infty, & s = 0, \quad t = 1, \dots, q, \\ +\infty, & s = 1, \dots, p, \quad t = 0. \end{cases} \tag{5.15}$$

Оптимальные значения индексов суммирования находятся по следующему правилу:

$$\begin{aligned} J^*(p) &= q, \\ J^*(i-1) &= \begin{cases} J^*(i), & \text{если } W_{i-1,J^*(i)} \leq W_{i-1,J^*(i)-1}, \\ J^*(i) - 1, & \text{если } W_{i-1,J^*(i)} > W_{i-1,J^*(i)-1}, \end{cases} \\ i &= p, p-1, \dots, 2. \end{aligned} \tag{5.16}$$

Доказательство. Легко видеть, что ограничения (5.12) определяют множество

$$\mathcal{J}_q(p) = \{J : \{1, \dots, p\} \rightarrow \{1, \dots, q\} | J(1) = 1, J(p) = q, 0 \leq J(i) - J(i-1) \leq 1, i = 2, \dots, p\} \tag{5.17}$$

допустимых соответствий индексов суммирования.

Обозначим задачу 2 через $\langle p, q \rangle$ и погрузим эту задачу в семейство подзадач $\langle s, t \rangle$ поиска минимальных значений

$$W_{s,t} = \min_{J \in \mathcal{J}_t(s)} \sum_{i=1}^s w_{i,J(i)}, \quad s = 1, \dots, p, \quad t = 1, \dots, q,$$

на множестве сжимающих отображений

$$\mathcal{F}_t(s) = \{J : \{1, \dots, s\} \rightarrow \{1, \dots, t\} | J(1) = 1, J(s) = t, 0 \leq J(i) - J(i-1) \leq 1, i = 2, \dots, s\}.$$

Из принципа оптимальности Беллмана и (5.17) следует, что значение $W_{s,t}$ может быть выбрано только из двух возможных вариантов:

$$\{W_{s-1,t} + w_{s,t}, W_{s-1,t-1} + w_{s,t}\}, \tag{5.18}$$

где

$$W_{s-1,t} = \min_{J \in \mathcal{F}_t(s-1)} \sum_{i=1}^{s-1} w_{i,J(i)},$$

$$W_{s-1,t-1} = \min_{J \in \mathcal{F}_{t-1}(s-1)} \sum_{i=1}^{s-1} w_{i,J(i)}. \tag{5.19}$$

Из (5.18) и (5.19) следует справедливость формул (5.14) и (5.13). Справедливость начального и граничных условий в формуле (5.15) следует из (5.17). Правило (5.16) является очевидным следствием формул (5.13), (5.14) и (5.15). Лемма 1 доказана.

Кроме того, нам потребуется следующая вспомогательная

Задача 3. Дано: совокупность $\{g_p(n), p = q, \dots, \ell, n = 1, \dots, N - p + 1\}$ числовых последовательностей и натуральное число $T_{\max} \geq \ell$. Найти: набор $\{(n_1, p_1), \dots, (n_M, p_M)\}$ пар натуральных чисел, в котором $n_m \in \{1, \dots, N - p_m + 1\}$, $p_m \in \{q, \dots, \ell\}$, $m = 1, \dots, M$, а также размерность M этого набора, доставляющие минимум целевой функции

$$G((n_1, p_1), \dots, (n_M, p_M)) = \sum_{m=1}^M g_{p_m}(n_m), \tag{5.20}$$

при ограничениях (2.2) и (2.3).

Справедлива следующая

Лемма 2. Минимум целевой функции G в задаче 3 находится по формуле

$$G^* = \min_{p \in \{q, \dots, \ell\}} \min_{n \in \{1, \dots, N-p+1\}} G_p(n), \tag{5.21}$$

а значения $G_p(n)$ условных оптимумов находятся в результате вычислений по рекуррентным формулам

$$G_p(n) = \begin{cases} g_p(n), & p = q, \dots, \ell, \quad n = 1, \dots, q, \\ \min \left\{ 0, \min_{i \in \{q, \dots, \min\{n-1, \ell\}\}} \min_{j \in \gamma_i(n)} G_i(j) \right\} + g_p(n), & p = q, \dots, \ell, \quad n = q + 1, \dots, N - p + 1, \end{cases} \tag{5.22}$$

где

$$\gamma_i(n) = \{k | \max\{n - T_{\max}, 1\} \leq k \leq n - i\}, \quad n = i + 1, \dots, N, \quad i = q, \dots, \ell. \tag{5.23}$$

Доказательство. Опираясь на ограничения (2.2) и (2.3), определяем множество

$$\Phi = \{ \{(n_1, p_1), \dots, (n_M, p_M)\} | p_m \in \{q, \dots, \ell\}, n_m \in \{1, \dots, N - p_m + 1\}, m = 1, \dots, M; p_{m-1} \leq n_m - n_{m-1} \leq T_{\max}, m = 2, \dots, M \} \tag{5.24}$$

наборов пар натуральных чисел, допустимых в задаче 3.

Для упрощения записи доказательства положим

$$\varphi = \{(n_1, p_1), \dots, (n_M, p_M)\}. \tag{5.25}$$

В соответствии с (5.20), (5.24) и (5.25), для оптимального значения целевой функции G имеем

$$G^* = \min_{\varphi \in \Phi} G(\varphi). \tag{5.26}$$

Определим подмножество

$$\Phi_p(n) = \{ \varphi \in \Phi | (n_M, p_M) = (n, p) \}, \quad p = q, \dots, \ell, \quad n = 1, \dots, N - p + 1, \tag{5.27}$$

наборов пар из множества Φ , у которых компоненты последней пары в наборе имеют фиксированные значения n и p .

Легко видеть, что

$$\Phi = \bigcup_{p=q}^{\ell} \bigcup_{n=1}^{N-p+1} \Phi_p(n). \tag{5.28}$$

Заметим, что для искомой размерности M наборов возможны два варианта: $M = 1$ и $M \geq 1$. Поэтому при каждом $p = q, \dots, \ell$ множество (5.27) в зависимости от M может быть представлено в виде

$$\Phi_p(n) = \begin{cases} \{\varphi \in \Phi_p(n) | M = 1\}, & n = 1, \dots, p - 1, \\ \{\varphi \in \Phi_p(n) | M = 1\} \cup \{\varphi \in \Phi_p(n) | M > 1\}, & n = p, \dots, N - p + 1. \end{cases} \tag{5.29}$$

Для индексов $p = q, \dots, \ell, n = 1, \dots, N - p + 1$ справедливо очевидное равенство

$$\{\varphi \in \Phi_p(n) | M = 1\} = \{(n, p)\}. \tag{5.30}$$

Рассмотрим множество $\{\varphi \in \Phi_p(n) | M > 1\}, p = q, \dots, \ell, n = p, \dots, N - p + 1$. Из ограничений (2.2), (2.3) и определения (5.23) следует, что если значения n_m и p_m фиксированы, то при каждом $m = 2, \dots, M$ для значений компонент $(m - 1)$ -й пары справедливо

$$p_{m-1} \in \{q, \dots, \min\{\ell, n_m - 1\}\}, \quad n_{m-1} \in \gamma_{p_{m-1}}(n_m).$$

Отсюда следует, что для индексов $p = q, \dots, \ell, n = p, \dots, N - p + 1$ справедливо равенство

$$\begin{aligned} \{\varphi \in \Phi_p(n) | M > 1\} &= \{(n_1, p_1), \dots, (n_M, p_M) | n_M = n, p_M = p, \\ &p_{m-1} \in \{q, \dots, \min\{\ell, n_m - 1\}\}, n_{m-1} \in \gamma_{p_{m-1}}(n_m), m = 2, \dots, M\} = \\ &= \{(n_1, p_1), \dots, (n_{M-1}, p_{M-1}), (n, p)\} \{(n_1, p_1), \dots, (n_{M-1}, p_{M-1}) \in \Phi_i(j), \\ &i \in \{q, \dots, \min\{\ell, n - 1\}\}, j \in \gamma_i(n)\}. \end{aligned} \tag{5.31}$$

Докажем формулу (5.22), используя полученные равенства (5.30) и (5.31). Рассмотрим семейство задач $\langle p, n \rangle$ вычисления оптимальных значений

$$G_p(n) = \min_{\varphi \in \Phi_p(n)} G(\varphi), \quad p = q, \dots, \ell, \quad n = 1, \dots, N - p + 1. \tag{5.32}$$

Два варианта вычислений, представленных в (5.22), следуют из двух вариантов в формуле (5.29). Для первого варианта формулы (5.22), когда $p = q, \dots, \ell, n = 1, \dots, p - 1$, из (5.32), (5.29) и (5.30) имеем

$$G_p(n) = \min_{\varphi \in \Phi_p(n), M=1} G(\varphi) = \min_{\varphi \in \{(n,p)\}} G(\varphi) = g_p(n). \tag{5.33}$$

Для второго варианта формулы (5.22), когда $p = q, \dots, \ell, n = p, \dots, N - p + 1$, из (5.32) и (5.29) имеем

$$G_p(n) = \min \left\{ \min_{\varphi \in \Phi_p(n), M=1} G(\varphi), \min_{\varphi \in \Phi_p(n), M>1} G(\varphi) \right\}. \tag{5.34}$$

Далее, в формуле (5.34) для первого минимума справедлива цепочка равенств (5.33). Второй минимум в формуле (5.34) в соответствии с (5.31) и принципом оптимальности Беллмана выбирается из совокупности

$$\{G_i(j) + g_p(n), i \in \{q, \dots, \min\{\ell, n - 1\}\}, j \in \gamma_i(n)\}$$

возможных вариантов, где

$$G_i(j) = \min_{\varphi \in \Phi_i(j)} G(\varphi).$$

Иными словами, имеем

$$\min_{\varphi \in \Phi_p(n), M>1} G_p(n) = \min_{i \in \{q, \dots, \min\{n-1, \ell\}\}} \min_{j \in \gamma_i(n)} G_i(j) + g_p(n). \tag{5.35}$$

Наконец, справедливость формулы (5.22) следует из (5.33), (5.34) и (5.35), а справедливость (5.21) следует из (5.26), (5.28) и (5.32). Лемма 2 доказана.

Для отыскания компонент оптимальных наборов определим две функции:

$$\pi(n) = \begin{cases} 0, & \text{если } n = 1, \dots, q, \\ 0, & \text{если } \min_{i \in \{q, \dots, \min\{n-1, \ell\}\}} \min_{j \in \gamma_i(n)} G_i(j) \geq 0, \quad n = q + 1, \dots, N - q + 1, \\ \arg \min_{i \in \{q, \dots, \min\{n-1, \ell\}\}} \left\{ \min_{j \in \gamma_i(n)} G_i(j) \right\}, & \text{если} \\ \min_{i \in \{q, \dots, \min\{n-1, \ell\}\}} \min_{j \in \gamma_i(n)} G_i(j) < 0, & n = q + 1, \dots, N - q + 1, \end{cases} \quad (5.36)$$

и

$$I(n) = \begin{cases} -1, & \text{если } \pi(n) = 0, \\ \arg \min_{j \in \gamma_{\pi(n)}(n)} G_{\pi(n)}(j), & \text{если } \pi(n) > 0, \end{cases} \quad (5.37)$$

$$n = 1, \dots, N - q + 1.$$

Справедливо следующее очевидное

Следствие 1. Пусть выполнены условия леммы 2. Допустим, кроме того, что

$$\pi_1 = \arg \min_{p \in \{q, \dots, \ell\}} \left\{ \min_{n \in \{1, \dots, N-p+1\}} G_p(n) \right\}, \quad (5.38)$$

$$v_1 = \arg \min_{n \in \{1, \dots, N-\pi_1+1\}} G_{\pi_1}(n), \quad (5.39)$$

$$\pi_m = \pi(v_{m-1}), \quad v_m = I(v_{m-1}), \quad m = 2, \dots, M, \quad (5.40)$$

где M – наименьшее значение m , при котором $\pi(v_m) = 0$. Тогда набор

$$\Phi = \{(v_M, \pi_M), \dots, (v_1, \pi_1)\}$$

является оптимальным решением задачи 3.

6. АЛГОРИТМ

Опираясь на вспомогательные задачи и утверждения из разд. 5, формулируем алгоритм решения задачи 1.

Алгоритм \mathcal{A}

Вход: Y, U, T_{\max} и ℓ .

Прямой ход алгоритма.

Шаг 1 (построение семейства решений задач 2). Для каждого $n = 1, \dots, N - q + 1$ выполним:

Для каждого $p = q, \dots, \min\{\ell, N - n + 1\}$ выполним (Решение задачи 2):

(1) Вычислим

$$w_{s,t} = u_t^2 - 2y_{n+s-1}u_t, \quad s = 1, \dots, p, \quad t = 1, \dots, q. \quad (6.41)$$

(2) Найдем значение W^* по формулам (5.13), (5.14), (5.15) и последовательность $J^*(1), \dots, J^*(p)$ по формулам (5.16).

(3) Положим $W^*(n, p) = W^*$; $J^*(n, p) = \{J^*(i), i = 1, \dots, p\}$.

Конец построения семейства решений задач 2.

Шаг 2 (определение входа задачи 3). Для каждого $n = 1, \dots, N - p + 1$ положим $g_p(n) = W^*(n, p)$, $p = q, \dots, \ell$.

Шаг 3 (решение задачи 3). Вычислим совокупность значений $G_p(n)$, $p = q, \dots, \ell$, $n = 1, \dots, N - p + 1$, по формулам (5.22), (5.23), а так же значение G^* по формуле (5.21). Положим $F_A = G^*$.

Обратный ход алгоритма.

Шаг 4. Вычислим $\pi(n)$ и $I(n)$, $n = 1, \dots, N - q + 1$, по формулам (5.36) и (5.37).

Шаг 5. Найдем компоненты наборов (π_1, \dots, π_M) и (v_1, \dots, v_M) и их размерность M по формулам (5.38), (5.39) и (5.40).

Шаг 6. Положим $M_A = M$, $\mathcal{M}_A = \{v_M, \dots, v_1\}$, $\mathcal{P}_A = \{\pi_M, \dots, \pi_1\}$, $\mathcal{J}_A = \{J^*(v_M, \pi_M), \dots, J^*(v_1, \pi_1)\}$.

Выход: M_A , \mathcal{M}_A , \mathcal{P}_A , \mathcal{J}_A и F_A .

Справедлива следующая

Теорема 1. Алгоритм \mathcal{A} находит точное решение задачи 1 за время $\mathcal{O}(T_{\max}^3 N)$.

Доказательство. Для оптимального значения F^* целевой функции задачи 1 справедлива следующая цепочка равенств:

$$\begin{aligned}
 F^* &\stackrel{1}{=} \min_{\mathcal{M}, \mathcal{P}, \mathcal{J}} \sum_{m=1}^M \sum_{i=1}^{p_m} \{u_{J^{(m)}(i)}^2 - 2y_{n_m+i-1} u_{J^{(m)}(i)}\} \stackrel{2}{=} \min_{\mathcal{M}, \mathcal{P}} \min_{J^{(1)}, \dots, J^{(M)}} \sum_{m=1}^M \sum_{i=1}^{p_m} \{u_{J^{(m)}(i)}^2 - 2y_{n_m+i-1} u_{J^{(m)}(i)}\} = \\
 &\stackrel{3}{=} \min_{\mathcal{M}, \mathcal{P}} \sum_{m=1}^M \left\{ \min_{J^{(m)} \in \mathcal{J}_q(p_m)} \sum_{i=1}^{p_m} \{u_{J^{(m)}(i)}^2 - 2y_{n_m+i-1} u_{J^{(m)}(i)}\} \right\} = \\
 &\stackrel{4}{=} \min_{\mathcal{M}, \mathcal{P}} \sum_{m=1}^M \left\{ \min_{J \in \mathcal{J}_q(p_m)} \sum_{i=1}^{p_m} \{u_{J(i)}^2 - 2y_{n_m+i-1} u_{J(i)}\} \right\} \stackrel{5}{=} \min_{\mathcal{M}, \mathcal{P}} \sum_{m=1}^M W^*(n_m, p_m) \stackrel{6}{=} F_A.
 \end{aligned} \tag{6.42}$$

В цепочке равенств (6.42) равенство 1 – определение оптимального решения задачи 1. Равенство 2 очевидно. Справедливость равенства 3 следует из того, что каждый элемент внутренней суммы в левой части равенства 2 зависит только от $J^{(m)}$ при каждом $m = 1, \dots, M$. Равенство 4 очевидно. Равенство 5 следует из того, что выражение во внешних скобках в левой части равенства 5 есть оптимальное решение (в соответствии с леммой 1) задачи 2, полученное при выполнении шага 1 алгоритма. Равенство 6 справедливо, так как левая часть этого равенства, согласно лемме 2, есть оптимальное значение целевой функции задачи 3 с входными данными, заданными на шаге 2 алгоритма. Таким образом, оптимальность решения, найденного алгоритмом, следует из (6.42).

Оценим трудоемкость алгоритма, используя его пошаговую запись. На шаге 1 при каждом $n = 1, \dots, N - p + 1$ задача 2 решается, очевидно, за время $\mathcal{O}(qT_{\max})$, так как $q \leq p \leq \ell \leq T_{\max}$. Последовательность решений этих задач находится за время $\mathcal{O}(T_{\max}^2 N)$, так как при каждом $n = 1, \dots, N - p$ имеет место вложенность таблиц:

$$\begin{aligned}
 \{w_{s,t}, s = 1, \dots, p, t = 1, \dots, q\} &\subset \{w_{s,t}, s = 1, \dots, p + 1, t = 1, \dots, q\}, \\
 \{W_{s,t}, s = 1, \dots, p, t = 1, \dots, q\} &\subset \{W_{s,t}, s = 1, \dots, p + 1, t = 1, \dots, q\}.
 \end{aligned}$$

Шаг 2 выполняется за время $\mathcal{O}(T_{\max} N)$, так как $\ell \leq T_{\max}$. На шаге 3 при каждом $n = 1, \dots, N - p + 1$ и каждом $p = q, \dots, \ell$ вычисление $G_p(n)$ по формуле (5.22) выполняется за время $\mathcal{O}(T_{\max}^2)$, так как $q \leq p \leq \ell \leq T_{\max}$ и $\max_{i,n} |\gamma_i(n)| \leq T_{\max}$ в соответствии с (5.23). Следовательно, итоговая трудоемкость шага 3 оценивается величиной $\mathcal{O}(T_{\max}^3 N)$.

На шаге 4 вычисление по формулам (5.36) и (5.37), как и на шаге 1, выполняется за время $\mathcal{O}(T_{\max}^2 N)$. Шаги 5 и 6 выполняются за время $\mathcal{O}(N)$.

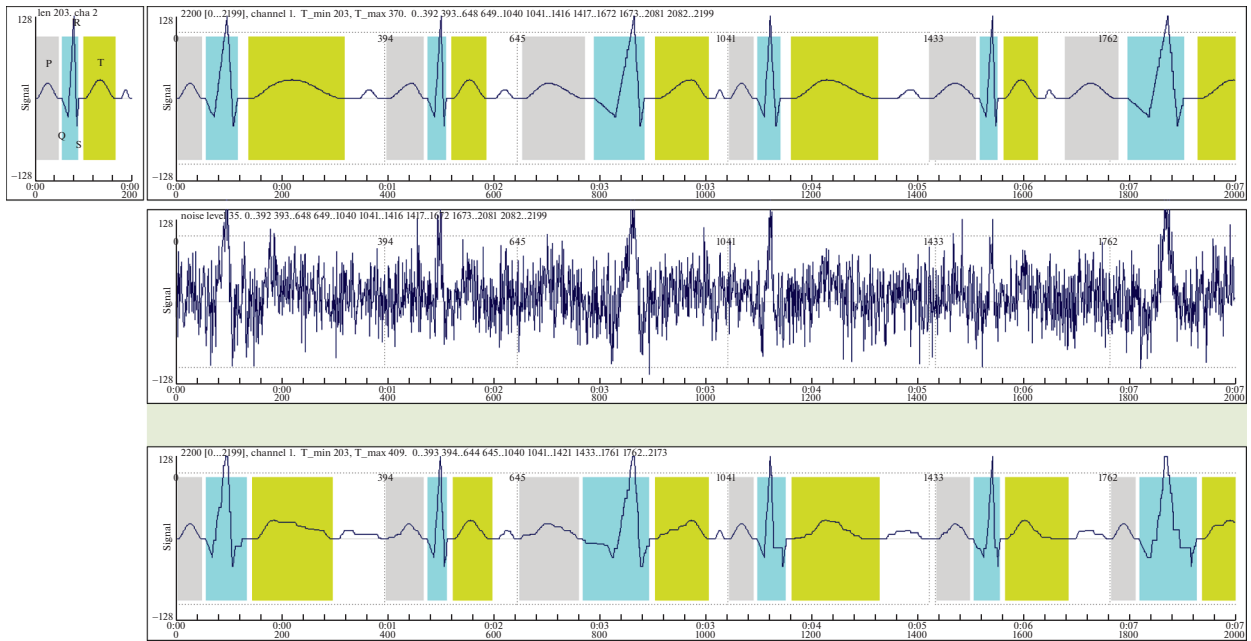
Суммируя временные затраты на всех шагах алгоритма, получаем итоговую трудоемкость алгоритма \mathcal{A} . Теорема 1 доказана.

Замечание 1. Если T_{\max} – фиксированный параметр, то время работы алгоритма равно $\mathcal{O}(N)$.

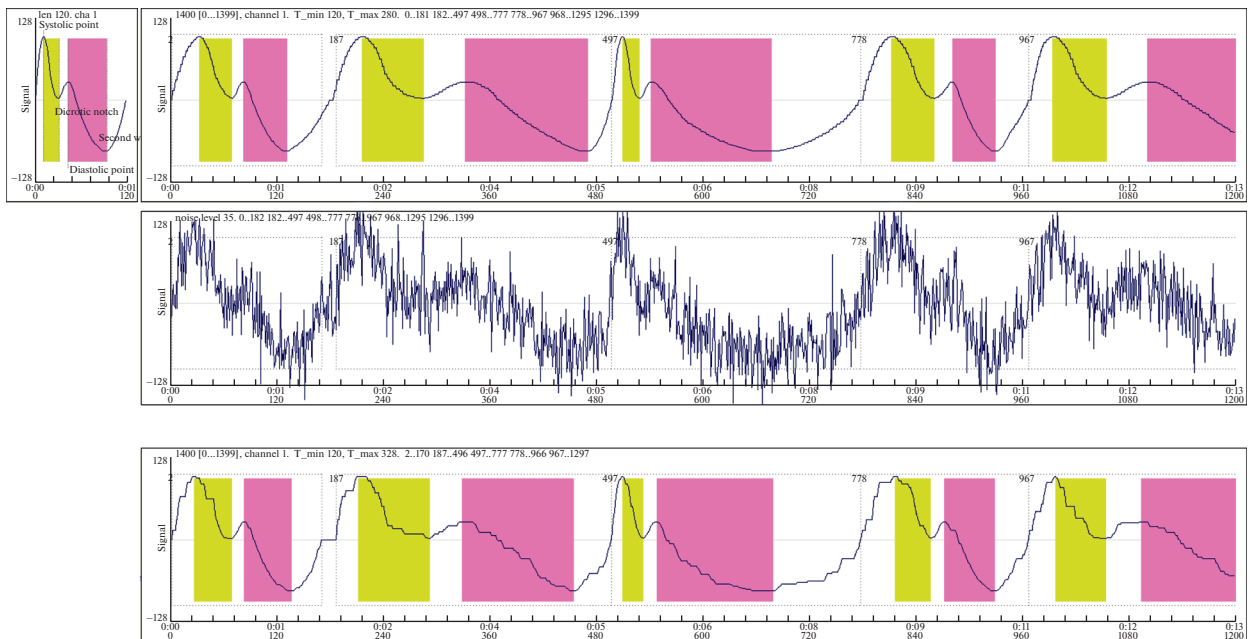
В следующем вспомогательном демонстрационном разделе мы приводим примеры применения полученного результата к решению двух модельных задач прикладного плана.

7. ПРИМЕРЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ

Очевидно, что с математической точки зрения не имеет значения, импульсу какой формы соответствует входная последовательность U . Тем не менее ниже для иллюстрации возможного



Фиг. 3. Пример помехоустойчивой обработки ECG-подобного сигнала.



Фиг. 4. Пример помехоустойчивой обработки PPG-подобного сигнала.

применения алгоритма в биомедицинских приложениях приведены два примера численного моделирования процесса обработки временных рядов (последовательностей), которые можно трактовать как квазипериодические последовательности флуктуирующих ECG-подобных (фиг. 3) и PPG-подобных импульсов (фиг. 4), искаженных аддитивной помехой.

На каждой фигуре в верхней части слева изображена последовательность U , соответствующая некоторому ECG- или PPG-импульсу. Описание формы этих импульсов, включая их характерные участки и значимые точки, широко представлено в литературе и интернете (см., например, [8]–[13]). На фиг. 3 этими участками являются P-Q-R-S-T-U, а на фиг. 4 – систолическая точка,

дикротическая точка, диастолическая точка. Характерные участки или интервалы между значимыми точками импульса выделены раскраской.

Справа от этого импульса изображена модельная программно сгенерированная квазипериодическая последовательность, соответствующая некоторой импульсной последовательности, порожденной квазипериодическими флуктуирующими повторами импульса U . В модельной последовательности каждый квазипериодический повтор последовательности U сопровождается ее флуктуирующими растяжениями, а интервалы между повторами не постоянны. Квазипериодичность этих повторов видна по растянутым характерным участкам импульса, которые выделены цветом.

В средней части каждой фигуры (под модельной последовательностью) изображена последовательность Y . Эта последовательность генерировалась как результат сложения модельной последовательности с последовательностью независимых одинаково распределенных гауссовских случайных величин с нулевым математическим ожиданием. Высокий уровень помехи (дисперсии) моделировался специально для демонстрации способности алгоритма как инструмента для обработки сильно зашумленных данных (как известно, высокий уровень помехи нетипичен для биомедицинских приложений). Входными последовательностями для алгоритма являются U и Y .

Последовательность X_A , приведенная в нижней части фигуры, является оценкой для модельной (ненаблюдаемой) последовательности в верхней части фигуры. Элементы X_A легко находятся по входной последовательности U и трем выходным наборам алгоритма по формулам (2.9), (2.7), (2.8) разд. 2 при $M = M_A$, $\mathcal{M} = \mathcal{M}_A$, $\mathcal{P} = \mathcal{P}_A$, $\mathcal{J} = \mathcal{J}_A$. По этим же формулам можно найти оценки для характерных растянутых участков и точек. Данные на фиг. 3 получены при $q = 128$, $T_{\max} = 370$, $N = 2000$, а на фиг. 4 при $q = 120$, $T_{\max} = 280$, $N = 1200$.

Сравнивая визуально верхние и нижние графики (временные ряды), можно оценить качество получаемых решений как вполне приемлемое для практики помехоустойчивой обработки данных в виде квазипериодических последовательностей флуктуирующих импульсов. В настоящей работе мы ограничиваемся лишь приведенными примерами, так как прикладные задачи находятся вне рамок настоящей работы. Здесь нас интересовали лишь новая дискретная экстремальная задача и алгоритм ее решения как математический инструмент для робастного решения сходных прикладных проблем. Вместе с этим отметим, что из любопытства алгоритм был опробован на целом ряде модельных импульсов различной формы и показал такую же высокую результативность.

ЗАКЛЮЧЕНИЕ

В работе установлена полиномиальная разрешимость не изученной ранее экстремальной задачи. Построен точный алгоритм, который имеет линейное время выполнения при фиксации одного из входов задачи как параметра.

В численных экспериментах мы продемонстрировали, что предложенный алгоритм послужит подходящим инструментом для решения некоторых прикладных проблем помехоустойчивой обработки квазипериодических импульсных последовательностей, в частности проблем обработки, анализа и интерпретации ECG и PPG сигналов.

Практический интерес представляет применение предложенного алгоритма к решению задачи распознавания, в которой имеется не одна последовательность U , а алфавит последовательностей, содержащий последовательности, соответствующие эталонным импульсам, например, каких-либо заболеваний.

С математической точки зрения значительный интерес представляет исследование дискретной экстремальной задачи, в которой по входной последовательности Y , имеющей квазипериодическую структуру, требуется найти последовательность U фиксированной длины. Не меньший математический интерес представляет модифицированная задача 1, в которой число суммируемых разностей сверток является частью входа задачи. Исследование этих задач представляется делом ближайшей перспективы.

СПИСОК ЛИТЕРАТУРЫ

1. Кельманов А.В., Хамидуллин С.А. Апостериорное обнаружение заданного числа одинаковых подпоследовательностей в квазипериодической последовательности // Ж. вычисл. матем. и матем. физ. 2001.

- Т. 41. № 5. С. 807–820. Kel'manov A.V., Khamidullin S.A. Posterior detection of a given number of identical subsequences in a quasi-periodic sequence // *Comp. Math. and Math. Phys.* 2001. V. 41. № 5. P. 762–774. (перевод)
2. Kel'manov A.V., Khamidullin S.A., Okol'nishnikova L.V. A Posteriori Detection of Identical Subsequences in a Quasiperiodic Sequence // *Pattern Recognition and Image Analysis.* 2002. V. 12. № 4. P. 438–447.
 3. Кельманов А.В., Хамидуллин С.А., Окольнишникова Л.В. Апостериорное обнаружение одинаковых подпоследовательностей-фрагментов в квазипериодической последовательности // *Сиб. ж. индустр. математики.* 2002. Т. 5. № 2(10). С. 94–108.
 4. Kel'manov A.V., Jeon B. A posteriori joint detection and discrimination of pulses in a quasiperiodic pulse train // *IEEE Transactions on Signal Processing.* 2004. V. 52. № 3. P. 645–656.
 5. Carter J.A., Agol E., et al. Kepler-36: a pair of planets with neighboring orbits and dissimilar densities // *Science.* 2012. V. 337. № 6094. P. 556–559.
 6. Carter J.A., Agol E. The quasiperiodic automated transit search algorithm // *The Astrophysical Journal.* 2013. V. 765. № 2. P. 132.
 7. Voskoboynikova G., Khairtdinov M. Numerical modeling of posteriori algorithms for the geophysical monitoring // *Communications in Computer and Information Science.* 2015. V. 549. P. 190–200.
 8. https://www.stepwards.com/page_id=24147
 9. Zhang D., Zuo W., Wang P. *Computational Pulse Signal Analysis.* Springer Nature Singapore Pte Ltd, 2018. 28 p.
 10. Rajni R., Kaur I. Electrocardiogram signal analysis – an overview // *Int. J. Comput. Appl.* 2013. V. 84. № 7. P. 22–25.
 11. https://www.comm.utoronto.ca/~biometrics/PPG_Dataset/index.html
 12. Shelley K., Shelley S. Pulse Oximeter Waveform: Photoelectric Plethysmography. In: Carol Lake, Hines R., Blitt C. (eds). *Clinical Monitoring.* W.B. Saunders Company. 2001. P. 420–428.
 13. Elgendi M. On the analysis of fingertip photoplethysmogram signals // *Current Cardiology Reviews.* 2012. V. 8. № 1. P. 14–25.