УДК (004.352.242)

ИССЛЕДОВАНИЕ МЕТОДА НЕПАРАМЕТРИЧЕСКИХ ОЦЕНОК ДЛИНЫ ПЕРИОДА И ПЕРИОДИЧЕСКОЙ СОСТАВЛЯЮЩЕЙ СИГНАЛА

В.В. Марченко, О.Г. Берестнева[1], Д.В. Девятых[1], Е.Ф. Суханова[1]

ЗАО «Элекард Девайсез», г. Томск
[1]Томский политехнический университет
E-mail: ogb6@yandex.ru

Рассматриваются методы выделения скрытой периодической составляющей сигнала. В качестве основного и наиболее подходящего для использования в данной работе был выбран метод непараметрических оценок длины периода и периодических составляющих, так как именно этот метод позволяет выделять скрытые составляющие из любых периодических сигналов, а не только гармонических. Исследование метода проводилось на модельных сигналах, показана зависимость относительной погрешности метода от количества скрытых периодов и шага дискретизации.

Ключевые слова:
Статистические оценки, показатель разброса, показатель размаха, скрытые периодические составляющие, погрешность.

Key words:
Statistical estimates, spread index, range index, hidden periodic components, inaccuracy.

Скрытые периодические процессы протекают практически во всех областях жизни и затрагивают множество сфер человеческой деятельности. Так, например, задачи выделения скрытых периодических составляющих возникают в геологии, сейсмологии, экономике, медицине. На сегодняшний день изучение методов поиска периодичностей, скрытых в наблюдаемых процессах, является по-прежнему актуальной задачей. Статья является продолжением цикла работ по данной тематике [1–3].

Основными параметрами ритмов являются амплитуда, период и фаза. Выход параметров ритмов за пределы нормы или появление их там, где они раньше не обнаруживались, как правило, связано с изменением свойств и возможной неисправностью системы. Существуют различные методы выделения частотных ритмов, такие как метод Стокса, спектральное представление сигнала с помощью ряда Фурье, корреляционный анализ и метод непараметрических оценок длины периода и периодических составляющих [4].

В работе рассмотрен метод непараметрических оценок длины периода и периодических составляющих, так как он позволяет выделять скрытые составляющие из сигналов любой периодической природы, а не только гармонической.

Целью исследования является изучение свойств и ограничений метода статистической оценки скрытой периодической составляющей на модельных сигналах. Объект исследования – сигналы, предполагающие наличие скрытой периодической составляющей, предмет исследования – периодические свойства сигналов, выделенные методом непараметрических оценок.

Кратко опишем суть выбранного метода. Рассмотрим достаточно широкий класс практически полезных непараметрических оценок длины периода и периодической составляющей во временных рядах. Во многих прикладных задачах рассматривают временной ряд (или случайный процесс)

$$y(t)=x(t)+e(t),$$

где $x(t)$ – детерминированная периодическая функция от времени t, т. е. $x(t)=x(t+T)$ при некотором T, где T – длина периода, а $e(t)$ – «шумы», случайные погрешности, искажающие периодический сигнал. Требуется оценить (минимальную) длину периода $T=T_0$ и периодическую составляющую $x(t)$. При этом не предполагается, что функция $x(t)$ входит в какое-либо параметрическое семейство, например конечных сумм синусов и косинусов, т. е. рассматривается задача непараметрического оценивания (минимальной) длины периода и периодической составляющей сигнала [5].

Пусть рассматриваемые функции $y(t)$, $x(t)$, $e(t)$ определены на отрезке $[0; A]$. При фиксированном T рассмотрим «куски» сигнала $y(t)$ на последовательных отрезках длины T, т. е. на отрезках $[0;T]$, $[T;2T]$, $[2T;3T]$,... Удобно ввести последовательность функций на отрезке $[0; T]$, полученную сдвигами этих кусков к началу координат:

$$y_1(t)=y(t),\ y_2(t)=y(t+T),\ y_3(t)=y(t+2T),...$$

Все они определены на отрезке $[0;T]$. Число этих функций равно числу полных периодов длины T, укладывающихся на отрезке $[0; A]$, т. е. равно целой части числа A/T. В дальнейшем из всех периодов будем рассматривать и оценивать только наименьший [5].

Существуют разные ситуации поведения функции и периодической составляющей:

1. Если $T=T_0$ – истинный период (или кратный ему), и погрешности $e(t)$ отсутствуют, то все введенные в предыдущем абзаце функции совпадают между собой и с периодической составляющей:

$$x(t)=y_1(t)=y_2(t)=y_3(t)=...$$

при всех t из $[0;T]$. При наличии погрешностей полного совпадения не будет. Однако отклонения определяются лишь шумами в различные моменты времени. При этом в качестве оценки периодической составляющей $x(t)$ естественно взять среднее арифметическое $y_{ср}(t)$ функций $y_1(t)$, $y_2(t)$, $y_3(t)$,... [5].

2. Если же T отличается от истинного периода T_0 (и кратных ему величин), то различия функций $y_1(t)$, $y_2(t)$, $y_3(t)$,... между собой определяются также и различием значений $x(t)$ в точках, отстоящих друг от друга на интервалы, длина которых кратна T [5].
3. Если число T/T_0 иррационально, то можно показать, что значения $t+mT(\mathrm{mod}\,T_0)$, где m – натуральные числа такие, что $t+mT<A$, асимптотически (при росте A) равномерно заполняют отрезок $[0;T_0]$, а потому при выполнении соответствующих условий регулярности, например непрерывности периодической составляющей сигнала, функция $y_{ср}(t)$ приближается к константе – среднему значению периодического сигнала $x(t)$ [5].
4. Если же число T/T_0 рационально, то наблюдаем промежуточный случай, в котором $y_{ср}(t)$, как можно показать, приближается к периодической функции с периодом $T=T_0/n$ при некотором натуральном n. Эта функция получена усреднением n последовательных участков длины T_0/n периодического сигнала $x(t)$. Она не является константой, хотя разброс ее значений меньше, чем для исходного периодического сигнала, поскольку T_0 – минимальная длина периода [5].

Для оценивания T целесообразно ввести два показателя: показатель разброса $F(T;Y)$ $F(T;y_1(t),y_2(t),y_3(t),...)$ множества функций $y_1(t)$, $y_2(t)$, $y_3(t)$,... на отрезке $[0;T]$ и показатель размаха $G(T;Y)=G(T;y_{ср}(t))$ функции $y_{ср}(t)$ на отрезке $[0;T]$ [5].

Символ Y означает здесь, что показатели разброса и размаха строятся по функции $y(t)$. При этом показатель разброса нацелен на оценку различий в значениях семейства функций при одном и том же значении аргумента. А показатель размаха – на различие значений одной и той же функции при разных значениях аргумента [5].

Для дискретного времени их можно адаптировать двумя способами: либо заменив *sup* на *max*, а интеграл на сумму; либо расширив область определения используемых функций на весь отрезок, например, соединив соседние точки отрезками или использовав для заполнения пропусков сплайны более высокого порядка [5].

Ввести показатели разброса $F(T;Y)=F(T;y_1(t), y_2(t),y_3(t),...)$ можно разными способами. Пусть $k=[A/T]$. Можно использовать различные функционалы супремумного типа (здесь и далее число слагаемых k не будем указывать в обозначении функционалов). Первым рассмотрим максимальный разброс непосредственно между значениями функций:

$$F_1(T,Y)=$$
$$=\sup\{|\,y_i(t)-y_j(t)\,|, i,j=1,2,...,0\le t\le T\}\ \ [5].$$

Второй функционал супремумного типа будет учитывать не произвольные отклонения, а только отклонения от «средней функции», т. е. иметь вид

$$F_2(T,Y)=\sup\{|\,y_i(t)-y_{cp}(t)\,|, i=1,2,...,k, 0\le t\le T\}.$$

Третий функционал показывает, какую зону «заметают» значения функций:

$$F_3(T,Y)=\sup\{|\,y_i(t), i=1,2,...,k, 0\le t\le T\}-$$
$$-\inf\{y_i(t), i-1,2,...,k, 0\le t\le T\}.$$

В качестве показателя разброса представляется полезным рассмотреть то или иное отклонение совокупности функций $Y_i(q)$, $i=1,2,...,k$ друг относительно друга. Можно сказать, что эти функции заполняют некую «трубку», которая тоньше всего при истинном значении периода T, а внутри нее проходит периодическая составляющая $X(q)=x(t)=x(qT)$. Естественно рассмотреть различные функционалы интегрального типа. Например, можно проинтегрировать максимум модулей попарных разностей:

$$F_4(T,Y)=\int_0^1\max\{|\,Y_i(q)-Y_j(q)\,|, i,j=1,2,...,k\}dq\ [5].$$

Вместо максимума можно проинтегрировать сумму:

$$F_5(T,Y)=\int_0^1\sum_{i,j=1}^{k}|\,Y_i(q)-Y_j(q)\,|dq.$$

Как и для функционалов супремумного типа, естественно рассмотреть показатели разброса относительно «средней функции»:

$$F_6(T,Y)=\int_0^1\max\{|\,Y_i(q)-Y_{cp}(q)\,|, i=1,2,...,k\}dq,$$

$$F_7(T,Y)=\int_0^1\sum_{i,j=1}^{k}|\,Y_i(q)-Y_{cp}(q)\,|dq\ \ [5].$$

Список показателей разброса можно существенно расширить.

Показатели размаха также можно ввести самыми различными способами. Например, можно рассмотреть такой показатель:

$$G_1(T,Y)=\sup\{|\,y_{cp}(t)\,|, 0\le t\le T\}-$$
$$-\inf\{|\,y_{cp}(t)\,|, 0\le t\le T\}.$$

При практическом использовании рассматриваемых алгоритмов целесообразно учитывать дополнительные особенности реальных временных рядов. В частности, обратим внимание на неустойчивость супремумов по отношению к выбросам (резко выделяющимся наблюдениям) сравнительно с функционалами интегрального типа. Во многих ситуациях временные ряды дают резкие выбросы (всплески), которые затем, как правило, сглаживаются. Поэтому целесообразно в качестве показателей разброса и размаха использовать функционалы интегрального типа [5].

Выберем для реализации показатели разброса и размаха F_1 и G_1.

Приступим к исследованию выбранного метода на модельных сигналах и, прежде всего, убедимся в его способности находить ритмические составляющие в различных периодических сигналах.

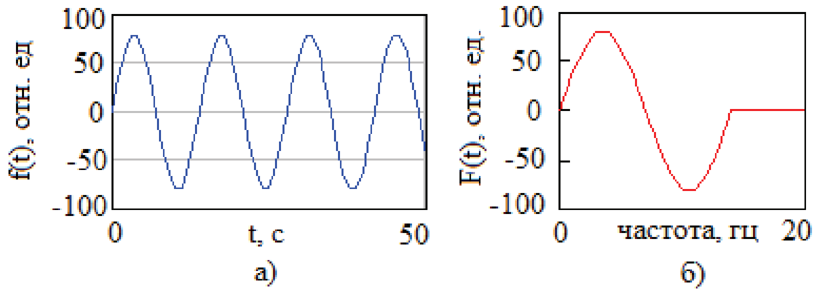

Рис. 1. *а) сигнал синуса; б) оценка периодической составляющей сигнала синуса*

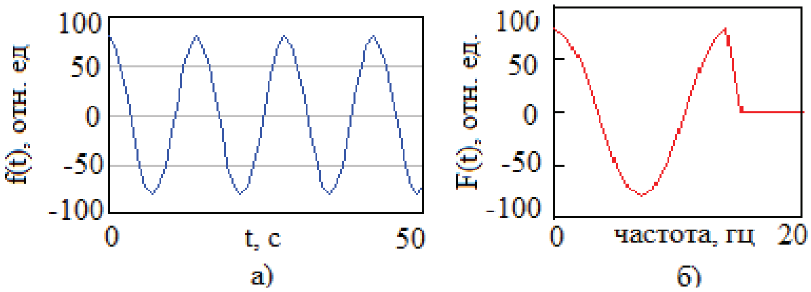

Рис. 2. *а) сигнал косинуса; б) оценка периодической составляющей сигнала косинуса*

Для этого возьмем гармонические сигналы синуса и косинуса, а также негармонические в виде функций «ступенька», «пила» и тангенса.

Пусть сигнал задан в виде $f(t)=A\cdot\sin(2\pi\omega_x t)$, где A – амплитуда сигнала, ω_x – частота, заданная равноотстоящими по аргументу t значениями (рис. 1, *а*). Выберем следующие значения: A=80, ω_x=9 и проверим метод на данном сигнале. Оценка частоты сигнала равна 9,6 Гц (рис. 1, *б*).

Пусть исследуемая функция $f(t)=A\cdot\cos(2\pi\omega_x t)$, где A – амплитуда сигнала, ω_x – частота, задана равноотстоящими по аргументу t значениями. Выберем следующие значения: A=80, ω_x=9 (рис. 2, *а*). Оценка частоты сигнала равна 9,6 Гц (рис. 2, *б*).

Пусть исследуемая функция задана следующим образом

$$f(t)=\begin{cases} A, & \text{если } \operatorname{mod}(t,T/2)>0,5; \\ A/2, & \text{если } \operatorname{mod}(t,T/2)\le 0,5, \end{cases}$$

где A – амплитуда сигнала, mod $(t,T/2)$ – остаток от деления, заданный равноотстоящими по аргументу t значениями (рис. 3, *а*). Выберем значения A=10, T=1/9. Оценка частоты сигнала равна 9,9 Гц (рис. 3, *б*).

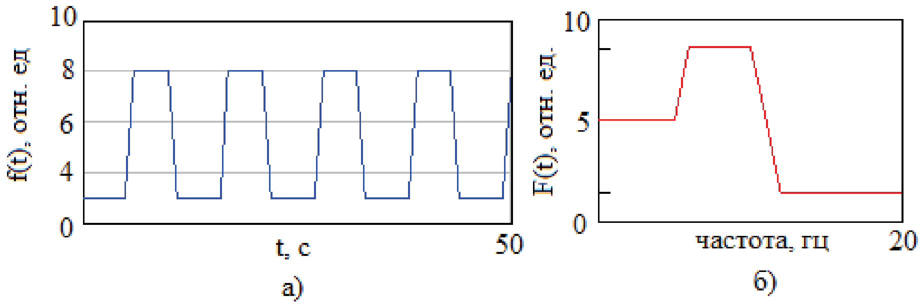

Рис. 3. *а) сигнал ступенчатой функции; б) оценка периодической составляющей сигнала ступенчатой функции*

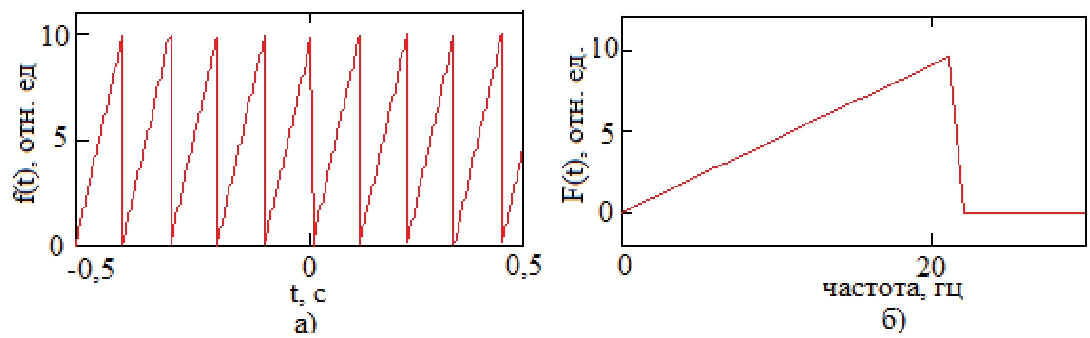

Рис. 4. *а) сигнал функции «пила»; б) оценка периодической составляющей сигнала функции «пила»*

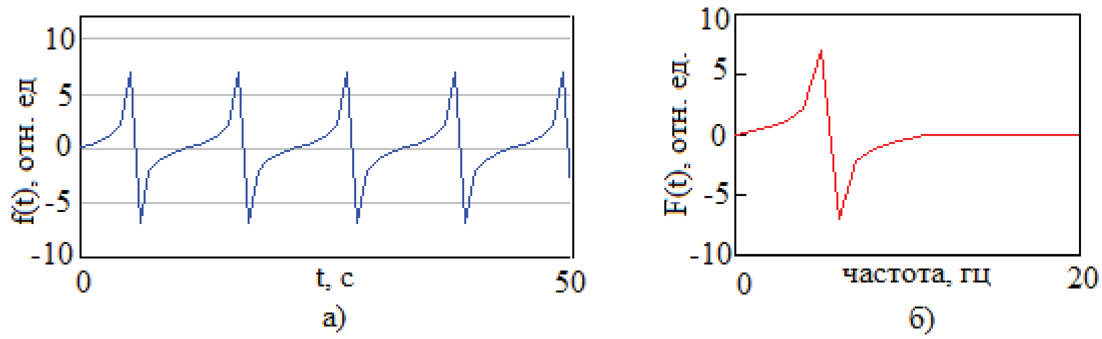

Рис. 5. *а) сигнал функции тангенс; б) оценка периодической составляющей сигнала функции тангенс*

Пусть исследуемая функция задана в виде «пилы» $f(t)=A\cdot\mathrm{mod}(t,T)$, где A – амплитуда сигнала, mod (t,T) – остаток от деления, заданный равноотстоящими по аргументу t значениями (рис. 4, *а*). Выберем следующие значения: $A=10$, $T=1/9$. Оценка частоты сигнала равна 9,4 Гц (рис. 4, *б*).

Пусть исследуемая функция $f(t)=A\mathrm{tg}(2\pi\omega_x t)$, где A – амплитуда сигнала, ω_x – частота, заданная равноотстоящими по аргументу t значениями (рис. 5, *а*). Выберем следующие значения: $A=10$, $\omega_x=9$. Оценка частоты сигнала равна 9,9 Гц (рис. 5, *б*).

Проведем исследование зависимости точности обнаружения скрытых параметров от количества периодов, содержащихся в реализации процесса, и шага дискретизации.

Так как гармонические сигналы является наиболее распространенными среди периодических сигналов, встречающихся в повседневной жизни, для дальнейших исследований возьмем функцию синуса.

Исследуем функцию $f(t)=A\cdot\sin(2\pi\omega_x t)$, где A – амплитуда сигнала, ω_x – частота, заданная равноотстоящими по аргументу t значениями. Выберем следующие значения: $A=80$, $\omega_x=9$.

Известно, что для дискретизации сигнала без потери информации согласно теореме Котельникова частота отсчетов должна быть в 2 раза выше верхней граничной частоты спектра сигнала. В качестве оптимального шага, позволяющего, не сохраняя лишних данных, восстановить сигнал с любой заданной точностью, возьмём значение $\Delta t_{опт}=0,5(f_{max})^{-1}$, где f_{max} – максимальная частота, содержащаяся в спектре сигнала, измеренная в герцах.

Из табл. 1 видно, что при уменьшении шага дискретизации величина относительных погрешностей по частоте также уменьшается. При шаге дискретизации в 5–6 раз меньше оптимального, метод достигает оптимальной точности ~90 %.

Очевидно, что длительность анализируемого сигнала влияет на количество периодов, содержащихся в реализации. Поэтому интересно выяснить, как зависит точность обнаружения скрытых параметров от количества периодов, содержащихся в исследуемой реализации процесса.

При найденном оптимальном шаге дискретизации определим минимальную длину реализации. Результаты исследования представлены в табл. 2.

Таблица 1. *Зависимость относительной погрешности метода непараметрических оценок от шага дискретизации*

Шаг дискретизации	Погрешность по частоте, %
$\Delta t=\Delta t_{опт}$	33,33
$\Delta t=\Delta t_{опт}/2$	33,33
$\Delta t=\Delta t_{опт}/3$	20
$\Delta t=\Delta t_{опт}/4$	14,28
$\Delta t=\Delta t_{опт}/5$	11,11
$\Delta t=\Delta t_{опт}/6$	9,08
$\Delta t=\Delta t_{опт}/7$	7,69
$\Delta t=\Delta t_{опт}/8$	6,67
$\Delta t=\Delta t_{опт}/9$	5,88
$\Delta t=\Delta t_{опт}/10$	5,27
$\Delta t=\Delta t_{опт}/11$	4,77

Анализируя полученные результаты можно сказать, что для точного выделения скрытой гармоники в исходной реализации должно содержаться, по крайней мере, два скрытых периода. Таким образом, следуя значениям табл. 2, можно сделать вывод, что погрешность по частоте метода непараметрических оценок составляет 9,09 %.

Таблица 2. *Зависимость относительной погрешности метода непараметрических оценок от количества скрытых периодов*

Количество периодов	Погрешность по частоте, %
1	200
2	9,09
3	9,09
4	9,09
5	9,09

Стоит также отметить, что следует рациональным образом выбирать длину анализируемой зависимости. Помимо общего увеличения объема вычислений, анализ излишне длинного участка невыгоден еще и потому, что при неустойчивом периоде смешиваются в кучу все участки, где период изменяется, что приводит к неубедительным результатам.

Основные результаты, полученные в работе.

1. Выделены основные свойства метода: универсальность в смысле применимости к различным сигналам, а также гибкость за счет использования различных показателей разброса и размаха.

2. Проведено исследование метода на модельном сигнале на зависимость точности обнаружения скрытых параметров:
   2.1) шаг дискретизации целесообразно брать в 5–6 раз меньше шага, определяемого по теореме Котельникова, так как дальнейшие уменьшения шага не приводят к существенному уменьшению погрешности;
   2.2) для точного выделения скрытой гармоники в исходной реализации должно содержаться, по крайней мере, два скрытых периода, при этом погрешность по частоте метода непараметрических оценок составляет 9,09 %.
3. Наиболее точную оценку периодических сигналов имеет функция «пила». Оценка частоты сигнала данной функции составляет 9,4 Гц. Сигналы синуса и косинуса также имеют достаточно точную оценку, составляющую 9,6 Гц. Оценки частот сигналов тангенса и ступенчатой функции оказались мене точными, их оценки составили 9,9 Гц.

Работа выполнена при финансовой поддержке РГНФ, проект № 12–06–12057в «Создание системы алгоритмических и программных средств обработки, представления и анализа экспериментальных данных в социальных и медицинских исследованиях».

СПИСОК ЛИТЕРАТУРЫ

1. Абдулкина Н.Г., Алайцева С.В., Константинова Л.И., Кохановская Ю.Г., Кочегуров В.А., Марченко В.В., Степаненко Н.П. Применение геометрического метода анализа фазового портрета для оценки биоэлектрической активности головного мозга у подростков с дисфункцией гипоталамо-гипофизарно-тиреоидной системы // Вестник новых медицинских технологий. – 2009. – Т. 16. – № 1. – С. 14–17.
2. Кочегуров В.А., Константинова Л.И., Марченко В.В. Использование фазового отображения ЭЭГ для контроля состояния здоровья пациентов с тиреопатологией // Известия Томского политехнического университета. – 2011 – Т. 319. – № 5. – С. 125–129.
3. Берестнева О.Г., Пеккер Я.С. Выявление скрытых закономерностей в сложных системах // Известия Томского политехнического университета. – 2009. – Т. 315. – № 5. – С. 138–143.
4. Магазинникова А.Л. Основы цифровой обработки сигналов. – Томск: Томский межвузовский центр дистанционного образования, 2002. – 129 с.
5. Орлов А.И. Прикладная статистика. – М.: Изд-во «Экзамен», 2004. – 190 с.