
**ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ
СИСТЕМЫ**

УДК 519.218.82

© Т. Н. Князева, Л. В. Новиков, Н. И. Орешко

УДАЛЕНИЕ НЕСТАЦИОНАРНОГО ШУМА ИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

В статье приводится двухступенчатый метод очистки данных от шума, дисперсия которого изменяется в случайные моменты времени. На первом этапе на основе предложенного метода кластеризации вейвлет-коэффициентов, вычисленных по максимально накладываемому дискретному вейвлет-преобразованию, определяются участки шума с кусочно-постоянной дисперсией. На втором этапе производится расчет модифицированных порогов для каждого участка с последующим пороговым удалением шума. Предлагаемый подход по сравнению с традиционными методами очистки от шума, в которых порог рассчитывается по всей выборке данных, позволяет значительно эффективнее удалять локальный шум.

ВВЕДЕНИЕ

Результаты измерений многих физических процессов представляют собой аддитивную смесь квазидетерминированного сигнала и нестационарного шума, дисперсия которого изменяется в случайные моменты времени (гетероскедастический шум). В этом случае наиболее эффективным средством борьбы с шумами является вейвлет-технология, изначально предназначенная для исследования нестационарных процессов [1, 2].

В частности, вейвлетные подходы широко используются в задачах непараметрической регрессии для оценивания функции f на основе наблюдений $y(t_i)$ в дискретные моменты времени t_i :

$$y(t_i) = f(t_i) + \sigma_i \xi_i, \quad i = 0, 1, 2, \dots, N-1, \quad (1)$$

где $f(t_i)$ — истинное значение сигнала; N — объем выборки; ξ_i — шумовая составляющая, распределенная по нормальному закону с параметрами $\xi_i \sim N(0,1)$; σ_i — среднее квадратическое отклонение (СКО) шума. (В дальнейшем для простоты полагаем $t_i = i$). При этом стандартные гетероскедастические модели описания дисперсии погрешности измерений (шума) представляют собой достаточно гладкую и, кроме того, известную кривую [3].

При использовании вейвлетов пороговое значение в алгоритмах удаления шума формируется на основании оценки СКО, вычисленной по всей выборке [1]. Однако это приводит к заметному снижению качества очистки на участках, где шум значительно отличается по СКО от глобально вычисленного значения.

В настоящей работе решается задача удаления шума, дисперсия которого изменяется скачком

в случайные моменты времени, т. е. σ_i — кусочно-постоянная функция: $\sigma_i = \text{const}$ при значениях i из некоторого замкнутого интервала, образующего k -й кластер, $k = 1, 2, \dots, M$, M — число кластеров, $M \ll N$. Для этого в работе предлагается двухступенчатый способ очистки от шума, где на первом этапе на основе кластеризации (разбиения отсчетов на группы с близкими свойствами) вейвлет-коэффициенты (ВК) каждого уровня декомпозиции определяют группы данных с постоянной дисперсией, а затем на втором этапе производится расчет порогов для каждого кластера и пороговое удаление шума. При этом используется предложенное в [4] максимально накладываемое дискретное вейвлет-преобразование (МНДВП, "maximal overlap discrete wavelet transform").

МАКСИМАЛЬНО НАКЛАДЫВАЮЩЕЕСЯ ДИСКРЕТНОЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ

МНДВП является избыточным неортогональным преобразованием. Несмотря на это оно обеспечивает более качественную очистку от шума и обладает рядом преимуществ, в частности инвариантностью относительно сдвига и способностью работать с выборками произвольного объема.

На первом уровне вейвлет-преобразования ВК $W_1(i)$ и аппроксимирующие коэффициенты $V_1(i)$ рассчитываются по формулам:

$$\begin{aligned} W_1(i) &= \sum_{l=0}^{L-1} \tilde{g}_l y((i-l) \bmod N), \\ V_1(i) &= \sum_{l=0}^{L-1} \tilde{h}_l y((i-l) \bmod N), \end{aligned} \quad (2)$$

где

$$(i-l) \bmod N = \begin{cases} i-l, & \text{если } i-l \geq 0, \\ i-l+N, & \text{в остальных случаях;} \end{cases}$$

$\tilde{g}_l, \tilde{h}_l$ — коэффициенты высокочастотного и низкочастотного фильтров соответственно. Они могут быть получены из фильтров Добеши g_l, h_l , например, путем нормирования на $\sqrt{2}$, т. е. $\tilde{g}_l = g_l/\sqrt{2}$ и $\tilde{h}_l = h_l/\sqrt{2}$, таким образом что $\sum_{l=0}^{L-1} \tilde{g}_l = 0$ и $\sum_{l=0}^{L-1} \tilde{h}_l = 1$, где L — длина фильтра.

ВК и аппроксимирующие коэффициенты j -го уровня вычисляются по формулам

$$\begin{aligned} W_j(i) &= \sum_{l=0}^{L-1} \tilde{g}_l V_{j-1}((i-2^{j-1}l) \bmod N), \\ V_j(i) &= \sum_{l=0}^{L-1} \tilde{h}_l V_{j-1}((i-2^{j-1}l) \bmod N); \end{aligned} \quad (3)$$

при этом $V_0(i) = X(i)$. Эти два соотношения составляют пирамидальный алгоритм МНДВП. Обратное МНДВП вычисляется с помощью обратного пирамидального алгоритма по следующей формуле:

$$\begin{aligned} V_{j-1}(i) &= \sum_{l=0}^{L-1} \tilde{g}_l W_j((i+2^{j-1}l) \bmod N) + \\ &+ \sum_{l=0}^{L-1} \tilde{h}_l V_j((i+2^{j-1}l) \bmod N). \end{aligned} \quad (4)$$

ПРОВЕРКА ГИПОТЕЗЫ О ГЕТЕРОСКЕДАСТИЧНОСТИ ШУМА

Проверка данных на наличие в них шума с кусочно-постоянной дисперсией осуществляется по ВК первого уровня декомпозиции $W_1(i)$ (определенным по формуле (2)), соответствующим высокочастотной шумовой составляющей. Для этого ВК $W_1(i)$ разбиваются на p одинаковых интервалов с помощью окна выбранной длины, соблюдая компромисс между точностью оценивания дисперсии и возможностью отслеживания локальных изменений дисперсии. По каждому интервалу определяется выборочная дисперсия $D_k^*, k=1, \dots, p$.

Затем проверяется гипотеза H_0 об однородности полученной последовательности из выборочных дисперсий D_k^* . Известны способы проверки этой гипотезы: с использованием критериев Бартлетта [5], Кохрана или на основе вычисления нормализованной суммы квадратов дисперсий ВК первого уровня декомпозиции [6]. В настоящей

работе используется критерий Бартлетта, суть которого заключается в следующем. Для p выборок $X_1, X_2, \dots, X_p$ вычисляются выборочные дисперсии $D_1^*, D_2^*, \dots, D_p^*$. При заданном уровне значимости q принимают гипотезу H_0 (о равенстве дисперсий $D_1^* = D_2^* = \dots = D_p^*$) в случае выполнения следующего неравенства: $A/B \leq \chi^2_{1-q}(p-1)$, где $\chi^2_{1-q}(p-1)$ — χ^2 -распределение Пирсона с $p-1$ степенями свободы. Величины A и B вычисляются по формулам, приведенным в [5].

УДАЛЕНИЕ ШУМА

Традиционная схема удаления шума включает этапы ДВП, пороговую обработку ВК с вычислением порогов по глобальной дисперсии и обратного ДВП. Предлагается модифицированный алгоритм удаления шума, который применяется в случае выполнения гипотезы о его гетероскедастичности, отличительной чертой которого является введение процедуры кластеризации ВК каждого уровня декомпозиции и расчет порогов по кластерам.

Прямое ДВП. Сначала выполняется прямое ДВП по формулам (3). Для снижения граничных эффектов на каждом уровне декомпозиции необходимо увеличивать выборку на длину фильтра слева и справа. Для этого можно использовать, например, полиномиальное или симметрическое продолжение в зависимости от вида сигнала [7].

Кластеризация ВК. На разных уровнях декомпозиции вследствие особенностей МНДВП границы кластеров сдвигаются, поэтому для корректной очистки сигнала от шума необходимо проводить кластеризацию на каждом уровне декомпозиции, вычисляя границы кластеров с одинаковой дисперсией. В работе предлагается использовать алгоритм, основанный на сравнении дисперсий с применением F-распределения Фишера [5] и дальнейшим уточнении границ. Преимущества этого алгоритма состоят в том, что, во-первых, он не требует априорного знания числа кластеров, которое автоматически определяется в процессе кластеризации, во-вторых, алгоритм обеспечивает высокую точность разделения выборки на кластеры за счет разработанной процедуры уточнения границ.

Сначала проводится предварительная кластеризация. На каждом уровне j декомпозиции последовательность ВК разбивается на p интервалов заданной длины (например, по 10 коэффициентов). Далее на каждом интервале определяется вы-

борочная дисперсия $D_{j,k}^*$ ($k=1, \dots, p$), где k — номер интервала. Определяются все дисперсии, попадающие в доверительный интервал при заданном уровне значимости q :

$$F_{q/2}(r, r) \bar{D}_j^* \leq D_{j,k}^* \leq F_{1-q/2}(r, r) \bar{D}_j^*,$$

где F — распределение Фишера с числом степеней свободы r , $\bar{D}_j^*$ — медиана из p дисперсий. Из отобранных таким образом дисперсий вычисляется новая медиана и, в соответствии с приведенным неравенством, уточняются границы кластера. Удаляем из последовательности дисперсий те значения, которые вошли в сформированный кластер и затем процедура повторяется до тех пор, пока множество дисперсий не станет пустым. Соседние интервалы с дисперсиями, входящими в один кластер, объединяются и пересчитываются границы кластеров.

В результате выполнения алгоритма на разных уровнях декомпозиции может оказаться неодинаковое число p_j кластеров ($k=1, 2, \dots, p_j$), каждый из которых объединяет ВК с индексами, образующими неперекрывающиеся подмножества $K_{j,k}$. В этих обозначениях ВК, входящие в k -й кластер j -го уровня декомпозиции имеют индексы $i \in K_{j,k}$, т. е. $W_j(i)|_{i \in K_{j,k}}$. Это означает, в частности, что каждый кластер может включать в себя несколько участков коэффициентов $W_j(i)$ с индексами, например, [1–20], [45–80] и т. п.

Так как начальная кластеризация проводилась по выборочным дисперсиям, рассчитанным на определенных интервалах, то на границах этих интервалов некоторые ВК могли попасть не в свой кластер, поэтому необходимо уточнение границ кластеров. Уточнение границ участков кластеров осуществляется путем минимизации или максимизации локальных дисперсий двух соседних кластеров при перемещении границы, которая их разделяет.

Пороговая обработка. По полученным кластерам на каждом уровне декомпозиции пересчитываются дисперсии и вычисляются пороги. Формула расчета СКО для k -го кластера на j -м уровне декомпозиции имеет вид

$$\sigma_{j,k} = \sqrt{2 \left(\frac{1}{z_k - 1} \sum_{i \in K_{j,k}} (W_j(i) - \bar{W}_j^k)^2 \right)^{1/2}}, \quad (5)$$

где $\bar{W}_j^k$ — среднее значение ВК, входящих в кластер k ; z_k — число ВК в k -м кластере.

Пороговые значения рассчитываются в зависи-

мости от выбранного типа порога. Например, если порог многоуровневый универсальный, то пороговые значения определяются для каждого уровня вейвлет-преобразования и для каждого кластера по формуле

$$thr_{j,k} = (\sigma_{j,k} \sqrt{2 \log z_k}) / 2^j. \quad (6)$$

На следующем этапе, с учетом рассчитанных пороговых значений и выбранного правила усечения (сравнения с порогом) производится корректировка ВК. Например, если правило усечения жесткое [1], то вейвлет-коэффициенты, прошедшие пороговую обработку, $\tilde{W}_j(i)$ вычисляются по формуле

$$\tilde{W}_j(i) = \begin{cases} W_j(i), & \text{если } |W_j(i)| > T; \\ 0, & \text{если } |W_j(i)| \leq T, \end{cases} \quad (7)$$

где $T = thr_{j,k}$ при $i \in K_{j,k}$.

На заключительном этапе выполняется обратное МНДВП, обеспечивающее восстановление исходного сигнала, используя измененные ВК всех уровней разложения и аппроксимирующие коэффициенты последнего уровня по формуле (4).

МОДЕЛИРОВАНИЕ

Эффективность применения рассмотренных выше алгоритмов проверена на модели (1), состоящей из шести пиков гауссовой формы $f(t)$ (см. рис. 1, а), искаженных медленно линейно нарастающим шумом ξ_i и шумом со скачками дисперсии σ_i в случайные моменты времени и случайной продолжительности (см. рис. 1, б). Проверка шума $\sigma_i \xi_i$ на гетероскедастичность проводилась по ВК первого уровня декомпозиции, вычисленных по формуле (2), разбитым на $p=103$ интервала (рис. 2, а). В качестве фильтров h_n, g_n приняты коэффициенты вейвлетов Добеши 4-го порядка. Использовался критерий Бартлетта при уровне значимости $q=0.95$. Отношение величин A/B оказалось равным 1702.5, что больше квантиля χ^2 -распределения Пирсона, взятого из таблиц, — 126.5741 [5]. Следовательно, гипотеза H_0 об однородности дисперсии, как и следовало ожидать, в данном случае не выполняется. Этот факт иллюстрируется на рис. 2, б, где показаны величины выборочных дисперсий в каждом интервале. Затем вычислялись ВК по формулам (3) до уровня $j=5$; на каждом уровне проводилась кластеризация и уточнение границ кластеров в соответствии

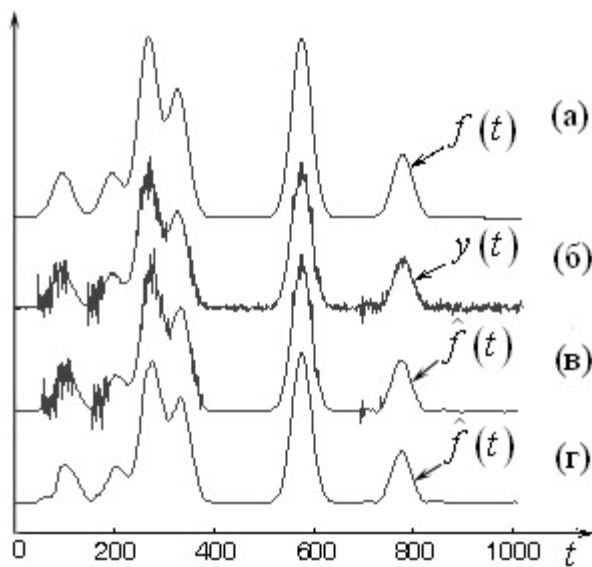


Рис. 1. Удаление нестационарного шума. Кривые: (а) — полезный сигнал; (б) — наблюдаемый сигнал; (в) — результат удаления шума классическим методом; (г) — результат удаления шума предлагаемым методом

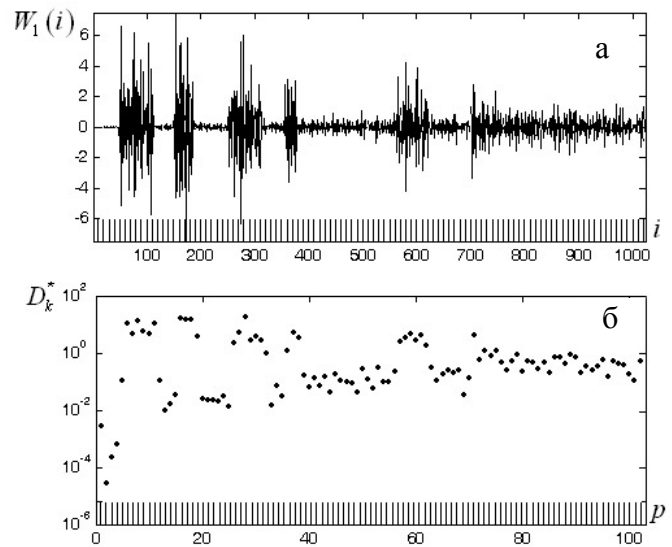


Рис. 2. Проверка гипотезы на гетероскедастичность шума. а — вейвлет-коэффициенты первого уровня декомпозиции; б — выборочные дисперсии D_k^*

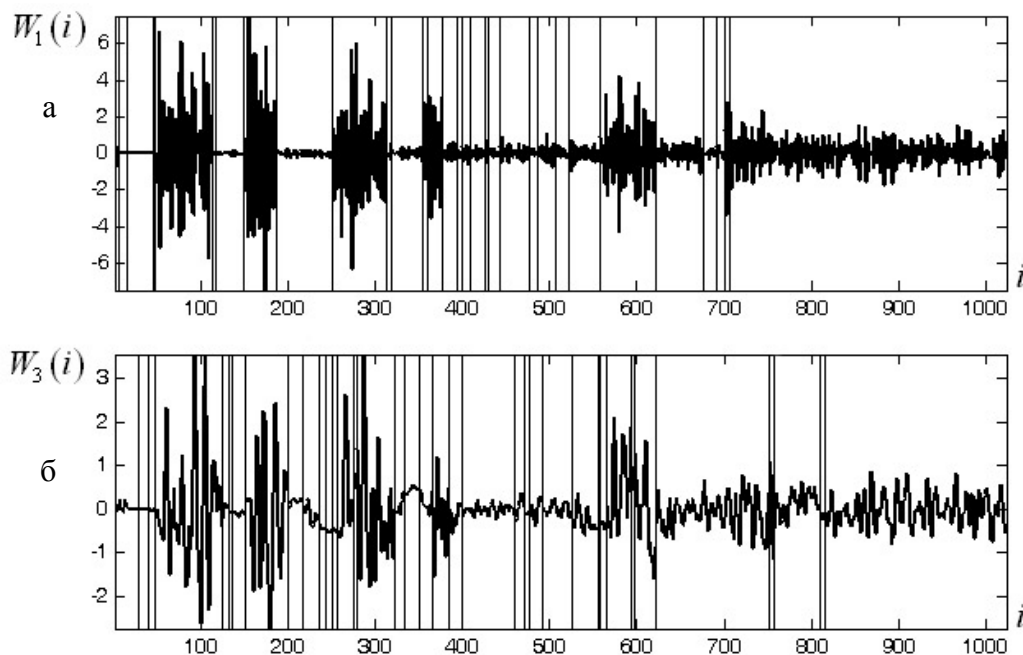


Рис. 3. Разбиение на кластеры вейвлет-коэффициентов: первого (а) и третьего (б) уровней декомпозиции после уточнения границ

с процедурой, изложенной выше, вычисление дисперсий и порогов по формулам (5) и (6) соответственно. Для примера на рис. 3 приведены ВК первого и третьего уровней декомпозиции с указанием уточненных границ кластеров. После сравне-

ния с порогами по формуле (7) выполнялось обратное МНДВП по формуле (4) для вычисления оценки полезного сигнала $\hat{f}(t_i)$. На рис. 1 приведены оценки функции $f(t_i)$, полученные двумя

способами: путем удаления шума классическим методом (график (в)) с помощью общего порога и предлагаемым методом (г) с помощью порога, вычисленного по кластерам. Из рис. 1 (в) видно, что участки сигнала, соответствующие самым зашумленным кластерам не очистились, поскольку общий порог оказался для данных кластеров ниже необходимого уровня. График (г) на рис. 1 показывает, что за счет вычисленного порога для каждого кластера удалось значительно подавить шум.

ЗАКЛЮЧЕНИЕ

Предложенный метод очистки от шума путем кластеризации ВК расширяет возможности порогового удаления шума и позволяет значительно эффективнее удалять локальный шум по сравнению с традиционными методами очистки от шума — практически полностью.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант 06-01-00457-а).

СПИСОК ЛИТЕРАТУРЫ

1. Donoho D., Johnstone I. Adapting to Unknown

- Smoothness via Wavelet Shrinkage // J. Amer. Statist. Assoc. 1995. V. 90. P. 1200–1224.
2. Vidakovic B. Statistical Modeling by Wavelets. New York: John Wiley & Sons, 1999. 382 p.
3. Rawlings J., Pantula S., Dickey D. Applied Regression Analysis. Springer, 1989. 658 p.
4. Percival D., Walden A. Wavelet Methods for Time Series Analysis. London: Cambridge University Press, 2000. 594 p.
5. Кобзарь А.И. Прикладная математическая статистика. М.: ФИЗМАТЛИТ, 2006. 816 с.
6. Inclan C., Tiao G. Use of Cumulative Sums of Squares for Retrospective Detection of Change of Variance // JASA. 1994. V. 427. P. 913–935.
7. Strang G., Nguyen T. Wavelets and Filter Banks. Boston: Wellesley-Cambridge Press, 1996. 512 p.

ОАО "Научно-инженерный центр Санкт-Петербургского государственного электротехнического университета" (Князева Т.Н., Оreshko Н.И.)

Институт аналитического приборостроения РАН, Санкт-Петербург (Новиков Л.В.)

Материал поступил в редакцию 18.04.2008.

REMOVAL OF NON-STATIONARY NOISE FROM EXPERIMENTAL DATA

T. N. Knayzeva¹, L. V. Novikov², N. I. Oreshko¹

¹Research & Engineering Center of Saint-Petersburg Electrotechnical University

²Institute for Analytical Instrumentation RAS, Saint-Petersburg

In this paper we present two-step method for denoising. The noise variance is changed in random moments of time. The first step of our method consists of clustering wavelet coefficients which are computed by maximal overlap discrete wavelet transform. Then one defines of noise parts with piecewise constant variance. The second step consists of calculation of the modified thresholds for each part and removing noise on basis of these thresholds. The approach practically completely allows to remove local noise in comparison with traditional methods of denoising in which the threshold calculate on all sample of data.