Информационные процессы, Том 13, № 3, 2013, стр. 198–205.

Очистка сигнала от шумов с использованием вейвлет преобразования и фильтра Калмана

М.В.Обидин*, **А.П.Серебровский****

**Московский физико-технический институт (государственный университет), Москва , Россия*
***Институт проблем передачи информации, Российская академия наук, Москва, Россия*
Поступила в редколлегию 13.06.2013

Аннотация—Рассмотрен метод очистки сигналов от шумов сочетающий в себе вейвлет преобразование и фильтр Калмана. Приводится сравнение с трешолдингом. Приводятся примеры очистки от шумов для реальных экономических данных.

КЛЮЧЕВЫЕ СЛОВА: вейвлеты, трешолдинг, фильтрация, биортогональные вейвлеты, фильтр Калмана.

1. ВВЕДЕНИЕ

Один из методов анализа зашумленных сигналов - это трешолдинг. По сути он представляет из себя разложение рассматриваемого сигнала в виде его вейвлет-спектра с последующей его обработкой. Следует заметить, что вейвлет-спектр это функция от двух аргументов - времени и масштаба. Однако в случае дискретного вейвлет преобразования мы получаем набор из N последовательностей, где номер последовательности соответствует координате масштаба, а номер элемента в последовательности - временной координате. Следует так же заметить, что даже для очень длинных исходных сигналов N - величина достаточно маленькая, она ограничена величиной $log_2(M)$, где M - количество отсчетов в рассматриваемом сигнале. В тоже время длинны последовательностей в дискретном вейвлет-спектре могут быть достаточно большими (длинна одной из последовательностей всегда будет близка к $\frac{M}{2}$). Все это позволяет обрабатывать последовательности вейвлет спектра независимо друг от друга. Так, жесткий трешолдинг устанавливает определенный пороговый уровень для каждой последовательности ДВС и заменяет на 0 все компоненты последовательности меньше этого порога. Такой подход, совмещенный с адаптивным алгоритмом выбора порога, позволяет трешолдингу очень успешно удалять шум без какой-либо дополнительной информации о сигнале.

К сожалению подобный метод не лишен некоторых недостатков, так например попытка очистить от шумов график индекса NASDAQ для компании Apple Inc. с использование трешолдинга (и критерия Штейна (SURE) для выбора порога) приводит к полностью неудовлетворительным результатам. В тоже время, если попытаться применить трешолдинг не к индексу, а к логарифму от него, то результат будет значительно более успешным (см рисунок 1). Данный пример показывает, что простой пороговый трешолдинг далеко не всегда является достаточно успешным приемом в очистке сигнала от шумов.

Метод описываемый в данной статье заключается в применении фильтра Калмана [1] для обработки последовательностей вейвлет спектра вместо трешолдинга. Используемый фильтр строится на достаточно общих предпосылках, что позволяет применять его к широкому классу сигналов, оценки этого фильтра, полученного Липцером и Хасьминским, обладают (вне пограничного слоя) наилучшей скоростью сходимости к нулю при $N \longrightarrow \infty$ в смысле среднеквад-

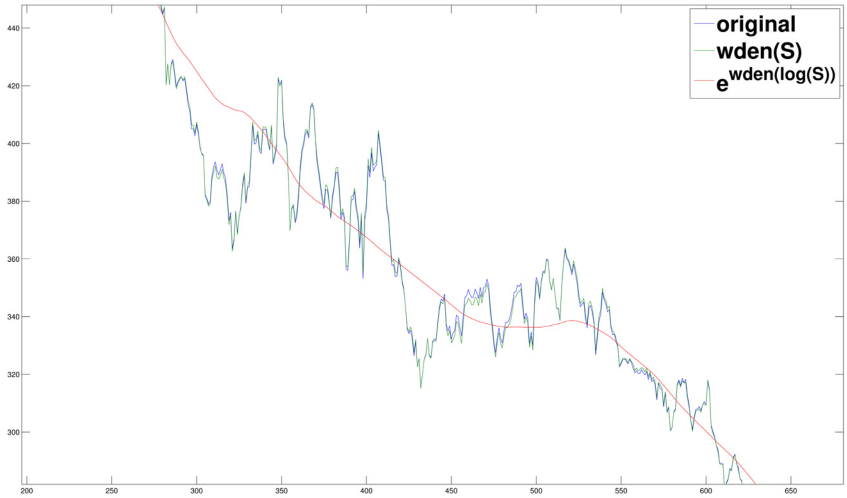

Рис. 1. Пример недостатка трешолдинга. Первый график - исходный сигнал S. Второй - $wden(S)$, где $wden$ - функция выполняющая трешолдинг. Третий - $e^{wden(log_2(S))}$.

ратической ошибки фильтрации, а его свойства позволяют избежать некоторых недостатков, присущих другим методам.

2. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ И ОБОЗНАЧЕНИЯ

Для краткости в данной статье приняты следующие обозначения:

ДВП Дискретное вейвлет преобразование

2.1. ВЕЙВЛЕТ РАЗЛОЖЕНИЕ

Основа ДВП - это 2 функции: масштабирующая и вейвлет. На основе каждой из этих функция строится свой фильтр. В литературе, посвященной данной тематике, их принято называть низкочастотный (low-pass) фильтр g и высокочастотный фильтр h.

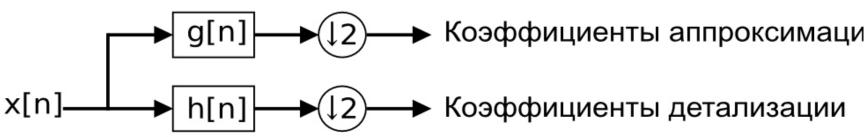

Рис. 2. Дискретное вейвлет преобразование.

Вейвлет преобразование выполняется следующим образом: сигнал независимо пропускается через оба фильтра и из получившихся сигналов выбираются только четные элементы. Таким образом на выходе у нас получаются две последовательности, каждая из которых в два раза короче исходной. Последовательность прошедшую h фильтр называют коэффициентами детализации. Последовательность прошедшую g фильтр называют коэффициентами аппроксимации. Эта процедура проиллюстрирована в рисунке 2

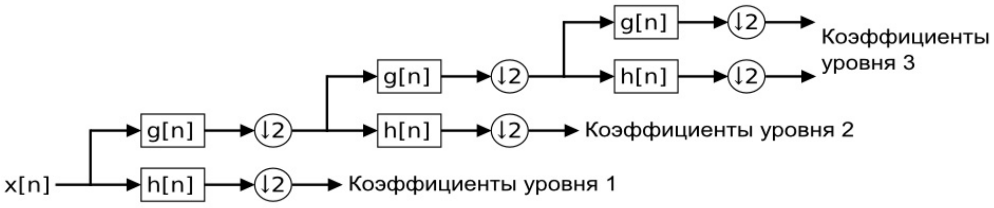

Рис. 3. Дискретное вейвлет преобразование глубины 3.

Если же мы проделаем ту же операцию над коэффициентами аппроксимации, то получим еще 2 набора коэффициентов называемых коэффициентами аппроксимации и детализации глубины 2, а саму процедуру будем называть ДВП глубины 2. Данную процедуру можно повторять до тех пор пока последовательность коэффициентов аппроксимации не станет слишком короткой. Таким образом мы можем получить ДВП глубины N, результатом работы которой являются N наборов коэффициентов детализации (каждая из которых в 2 раза короче предыдущей) и один набор коэффициентов аппроксимации. Иллюстрация этого процесса представлена на рисунке 3

Более подробное рассмотрение процесса вейвлет преобразования можно найти в статье [2], тем же кто заинтересован в глубоком понимании данного вопроса стоит обратиться к книге [3].

Важным свойством такого разложения являются обратимость и линейность. Это позволяет нам выполнить ДВП, как-то обработать коэффициенты, а затем выполнить обратное ДВП и получить измененный сигнал. Именно по такой схеме работает трешолдинг.

Однако это не единственная возможно использовать фильтрацию совместно с вейвлет преобразованием. В данной работе так же рассматривается следующий алгоритм:

Алгоритм 1.

1 Выполняем ДВП глубины N над исходным сигналом.

2а Заменяем все коэффициенты детализации и аппроксимации кроме коэффициентов детализации глубины 1

2б Применяем обратное ДВП к получившимся коэффициентам.

3в Применяем шаги 2а и 2б по очереди к каждому набору коэффициентов детализации и к набору коэффициентов аппроксимации.

В результате шага 2 мы получили $N+1$ сигнал, если их сложить, то мы получим исходный сигнал. Другими словами, с помощью вейвлет преобразования мы разложили наш сигнал на суммы из $N+1$ сигналов, каждый из которых мы можем обрабатывать независимо.

2.2. ФИЛЬТР КАЛМАНА

В данной статье мы использовали фильтр Калмана описываемый в работах [1] и [4]. Данный фильтр применим к сигналам из множества

$$\sum(\beta, L) = \left\{ f: \begin{array}{l} \text{существуют} k \text{производных}, f^{(0)}, f^{(1)}, ..., f^{(k)}; \\ |f^{(k)}(t_2) - f^{(k)}(t_1)| \le L|t_2 - t_1|^{\alpha}, \forall t_1, t_2, \alpha \in (0,1]; \\ \beta = k + \alpha \end{array} \right\}.$$

Он имеет вид

$$\widehat{F}^n(t_i) = \widehat{F}^n(t_{i-1}) + \frac{1}{n} a \widehat{F}^n(t_{i-1}) + q_n(X_i - A\widehat{F}(t_{i-1})),$$

где

$$\widehat{F}^n(t_i) = \begin{pmatrix} \widehat{f_n^{(0)}}(t_i) \\ \widehat{f_n^{(1)}}(t_i) \\ \vdots \\ \widehat{f_n^{(k)}}(t_i) \end{pmatrix}, \qquad q_n = \begin{pmatrix} q_0 n^{-\frac{2\beta}{2\beta+1}} \\ q_1 n^{-\frac{2\beta-1}{2\beta+1}} \\ \vdots \\ q_k n^{-\frac{2\beta-k}{2\beta+1}} \end{pmatrix},$$

$$A = \begin{pmatrix} 1 \; 0 \dots 0 \end{pmatrix}_{1\times(k+1)}, \qquad a = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}_{(k+1)\times(k+1)},$$

$$\begin{aligned} q_0(\gamma) &= U_{00}\left(\frac{\gamma}{\sigma}\right)^{\frac{1}{k+1}} \\ q_1(\gamma) &= U_{01}\left(\frac{\gamma}{\sigma}\right)^{\frac{2}{k+1}} \\ &\dots \\ q_k(\gamma) &= U_{0k}\left(\frac{\gamma}{\sigma}\right). \end{aligned}$$

Параметр γ подбирается [4] из условия минимизации функции

$$C(q) = \sigma^2 \left(\int_0^\infty q^* e^{(a-qA)^* t} e^{(a-qA)t} q dt + \left(\frac{L}{\sigma}\right)^2 \left[\left(\frac{1}{q_k}\right)^2 + \sum_{j=0}^{k-1} \left(\frac{q_j}{q_k}\right)^2 \right] \right).$$

Коэффициенты $U_{00}, \dots, U_{0k}$ берутся из таблицы (для $k \le 3$ полученной ещё Калачёвым [5]).

k	U_{00}	U_{01}	U_{02}	U_{03}	U_{04}
0	1	NA	NA	NA	NA
1	$\sqrt{2}$	1	NA	NA	NA
2	2	2	1	NA	NA
3	$\sqrt{4+\sqrt{8}}$	$2+\sqrt{2}$	$\sqrt{4+\sqrt{8}}$	1	NA
4	$1+\sqrt{5}$	$3+\sqrt{5}$	$3+\sqrt{5}$	$1+\sqrt{5}$	1

Данный фильтр на выходе дает не только сам сигнал, но и его производные. Однако производные в данной работе не используются, что означает, что мы можем выбрать любое $k \ge 0$.

К сожалению фильтр не сходится достаточно быстро, чтобы его можно было применять к коротким сигналам. Однако данный недостаток можно обойти следующим образом:

Алгоритм 2.

1 Применяем фильтр к сигналу используя нулевые начальные условия

2 Применяем фильтр к тому же сигналу, но с обратным временем. В качестве начальных условий берем последнюю точку полученную в шаге 2. Данный прием позволяет обойти пограничный слой около начальных условий.

3 Снова применяем фильтр к исходному сигналу (с прямым временем), но в качестве начальных условий берем последнюю точку с шага 2.

Результатом фильтрации является результат третьего прохода. Такой способ нахождения начальных условий значительно ускоряет сходимость фильтра и позволяет его применять к коротким сигналам.

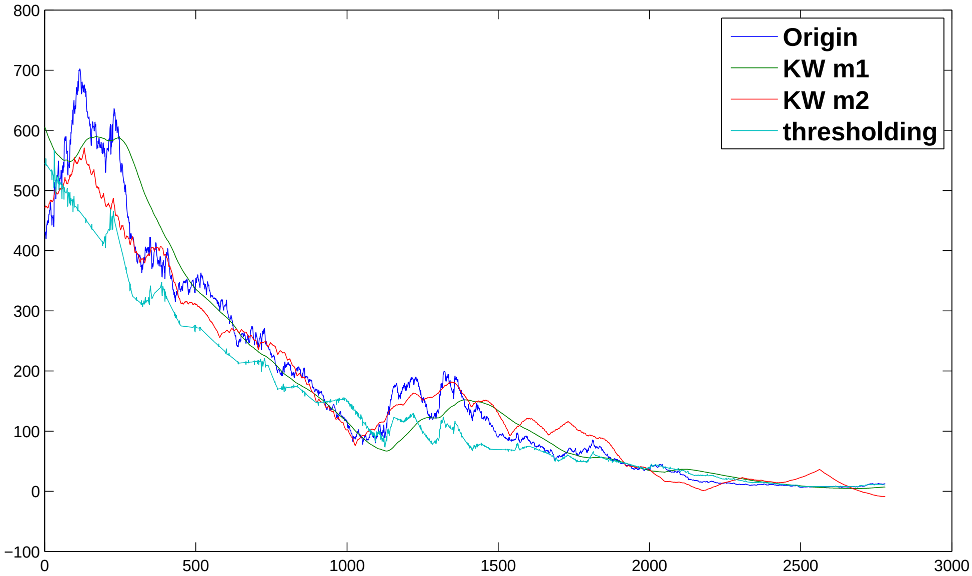

Рис. 4. Сравнение трешолдинга с KW фильтрами. Тип вейвлета - bior2.2 (подробнее см. [3] [6]), глубина разложения - 11, $\sigma = 4$, $k = 1$, $\gamma = 49$.

3. КОНЕЧНЫЙ ВИД ФИЛЬТРА И ЕГО ВАРИАНТЫ

Объединяя вейвлет преобразование и калмановский фильтр мы получаем два варианта фильтрации:

1)Мы используем вейвлет преобразование чтобы разделить исходный сигнал на сумму сигналов, а затем эти сигналы пропускаем через фильтр Калмана и затем суммируем их. Этот вариант фильтра значительно отличается от трешолдинга и в ряде случаев дает значительно превосходящие его результаты. Далее этот фильтр будет обозначаться $KWm1$.

2)Мы используем фильтр Калмана для обработки коэффициентов детализации сигнала (коэффициенты аппроксимации мы оставляем не тронутыми), а затем выполняем над ними обратное ДВП. Данный вариант похож на трешолдинг. Недостатком данного варианта является то, что фильтрация коэффициентов аппроксимации дает очень плохой результат, поэтому их приходится оставлять не тронутыми. В случае если глубина вейвлет разложения мала, мы получим длинный ряд коэффициентов аппроксимации, и фильтр будет иметь несколько незаконченный вид. Далее этот фильтр будет обозначаться $KWm2$.

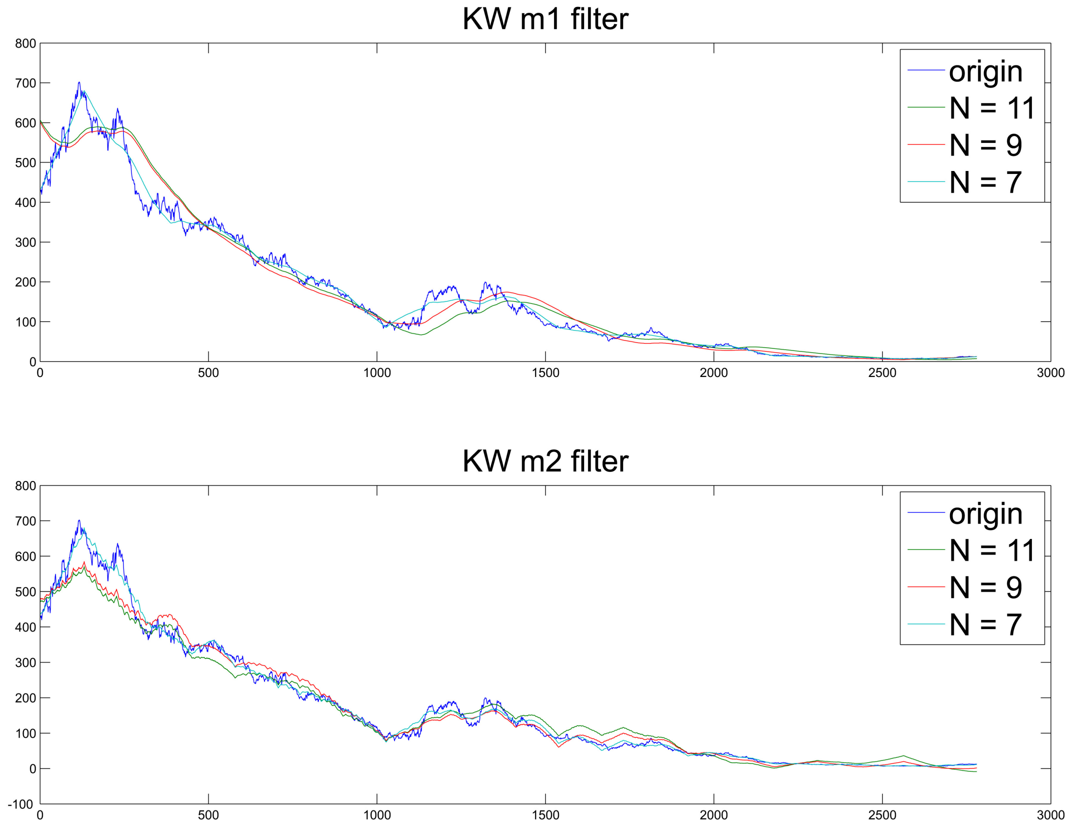

Рис. 5. Влияние глубины разложения. Тип вейвлета - bior2.2, $\sigma = 4$, $k = 1$, $\gamma = 49$.

Оба варианта фильтра можно видоизменять регулируя тип вейвлетов, глубину разложения, а так же параметры фильтра Калмана.

4. ПРИМЕРЫ ПРИМЕНЕНИЯ

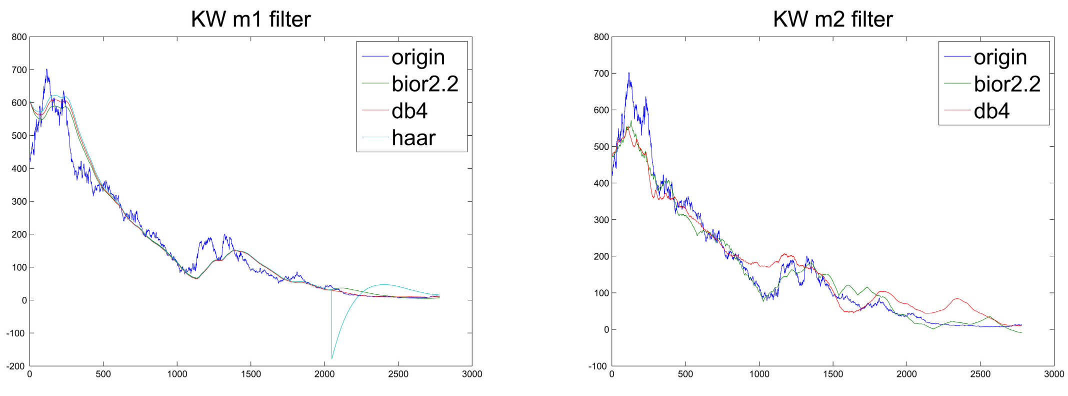

Рис. 6. Влияние типа вейвлета. Глубина разложения - 11, $\sigma = 4$, $k = 1$, $\gamma = 49$.

Фильтр применялся к данным цен на нефть и к индексу NASDAQ различных компаний за длительные периоды времени.

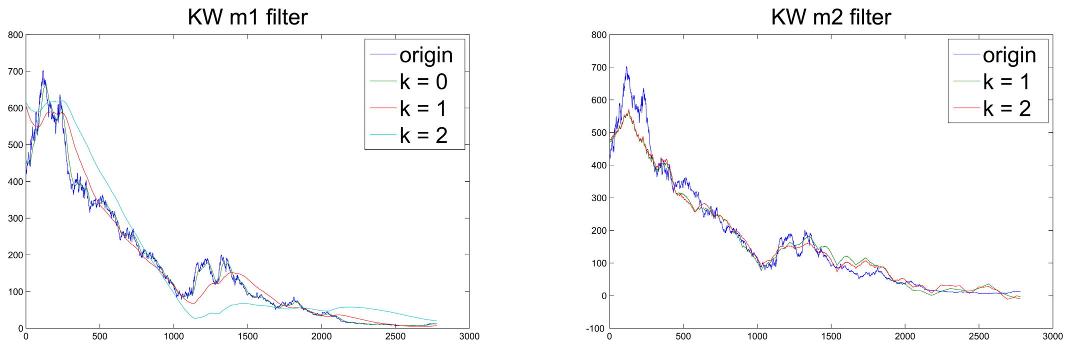

Рис. 7. Влияние параметров фильтрации . Глубина разложения - 11, $\sigma = 4$, $\gamma = 49$.

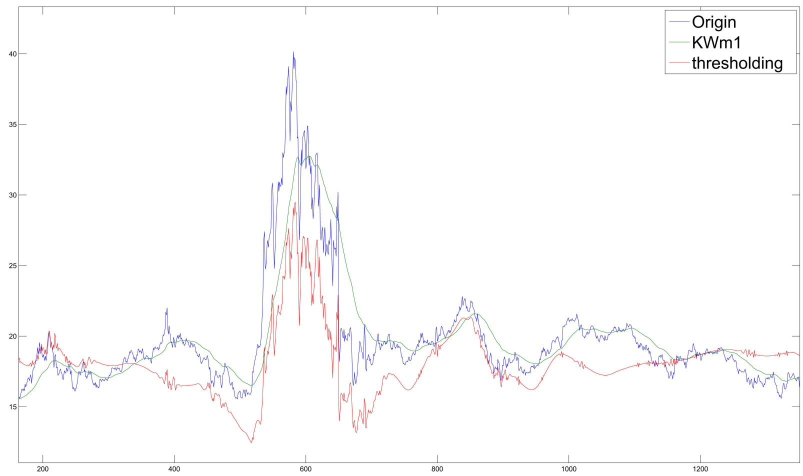

Рис. 8. Сравнение фильтра $KWm1$ при $k = 0$ с трешолдингом. Глубина разложения - 12, $\sigma = 4$, $\gamma = 49$.

На рисунке 4 можно увидеть сравнение двух вариантов фильтра с трешолдингом и с исходными данным индекса акций Apple Inc. Трешолдинг в данном примере применялся к натуральному логарифму от индекса акции, а затем бралась экспонента от полученного результата.

На рисунках 5 показана зависимость результатов фильтрации от глубины вейвлет разложения для первой и второй модификаций соответственно.

На рисунках 6 можно увидеть влияние типа вейвлета на результаты фильтрации. Типы вейвлетов обозначены согласно обозначениям принятым в MATLAB (подробнее о вейвлетах в MATLAB см. [6]). Для фильтра $KWm2$ не приведен пример с вейвлетом Хаара ввиду сильного несовпадения по масштабу с остальными графиками. Как видно тип вейвлета слабо влияет на $KWm1$ фильтр (исключение составляет вейвлет Хаара, он оказался полностью непригоден для обоих версий фильтра) и оказывает значительное влияние на $KWm2$ фильтр, что является достаточно неожиданным фактом.

На рисунках 7 отражено влияние параметров фильтра Калмана на результат фильтрации. Крайне интересный результат получается у фильтра $KWm1$ при $k = 0$.

На рисунке 8 можно увидеть более детальный пример сравнения этого варианта фильтра с тршеолдингом.

5. ВЫВОДЫ

В данной работе получен новый тип фильтра сочетающего в себе положительные качества как трешолдинга, так и калмановской фильтрации. Фильтр достаточно гибок в настройке и позволяет подбирать параметры фильтра под конкретные задачи.

СПИСОК ЛИТЕРАТУРЫ

1. Р. Ш. Липцер, Р. З. Хасьминский, Слежение за гладкой функцией регрессии. ТВП, 2002, том 47, выпуск 3, стр. 567-575.
2. Обидин М. В. Серебровский А. П. Вейвлеты и адаптивный трешолдинг. Информационные процессы, 2013, том 13, № 2, стр. 91-99.
3. Добеши И. Десять лекций по вейвлетам. Ижевск: НИЦ "Регулярная и хаотическая динамика"2001.
4. Goldentayer L, Liptser R, On-line Tracking of a Smooth Regression Function. Statistical Inference for Stochastic Processes 2006, 9(1), pp.17-30.
5. Калачев М. Г. Один метод многократного дифференцирования сигнала в системах автоматического управления. Автоматика и телемеханика, 1970, 6, стр. 29-36.
6. Смоленцев Н.К. Вейвлет-анализ в MATLAB. М.: ДМК Пресс. 2010

Method of denoising with using Kalman filter and wavelet transform

Obidin M.V., Serebroski A.P.

The paper deals with the method of denoising that combine wavelets transform and Kalman filter. Author compare this filter with thresholding. Examples of denoising real economics data are provided.

KEYWORDS: wavelets, thresholding, filtering, Kalman filter.