

Программа повышения конкурентоспособности

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение

высшего образования

«Уральский федеральный университет имени

первого Президента России Б. Н. Ельцина»

МЕТОДЫ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

Лекция № 6

Адаптивный анализ временных рядов

Екатеринбург

2017

Содержание

Цель изучения материала	3
Перечень компетенций, формирующихся или получающих приращение в результате прослушивания лекции.....	3
План лекции.....	4
Список литературы	4
Часть 1. Адаптивный анализ нестационарных временных рядов	5
Часть 2. Метод «Гусеница»-SSA. Сингулярный спектральный анализ.....	7
2.1. Этап разложения. Шаг вложения и шаг сингулярного разложения.....	8
2.2. Этап восстановления. Шаг группировки и шаг усреднения	9
2.3. Выбор способа группировки компонент в методе «Гусеница»	12
2.4. Выбор длины окна L компонент в методе «Гусеница»	15
Часть 3. Метод вейвлет-разложения временных рядов	17
3.1. Схема вейвлет-декомпозиции	18
3.2. Вейвлетная пакетная декомпозиция WPD.....	22
Часть 4. Эмпирическая модовая декомпозиция	24
Часть 5. Частотно-временные характеристики временных рядов.....	30
5.1. Спектрограмма	30
5.2. Мгновенная частота. Преобразование Гильберта	32
5.3. Метод прямой квадратуры	34

Цель изучения материала

Целью данной лекции является изучение методов анализа временных рядов, не требующих от них определенных свойств и характеристик, а также априорной информации о процессе, породившем данную выборку. В представленном материале подробно описаны методики использования адаптивных методов, включая самые последние разработки в данной области. Упор при описании методов делается на их практическую применимость, нежели на их теоретические свойства и параметры.

Перечень компетенций, формирующихся или получающих приращение в результате прослушивания лекции

Способность анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.

Умение разрабатывать стратегии проектирования, определением целей проектирования, критериев эффективности, ограничений применимости.

Умение определить и обосновать научные методы и соответствующие им технологии при решении исследовательской задачи.

Способность осуществлять сбор, анализ научно-технической информации, отечественного и зарубежного опыта по тематике исследования.

Способность разрабатывать методы решения нестандартных задач и новые методы решения традиционных задач

План лекции

1. Определение понятия адаптивных методов анализа ВР
2. Отличия в методике анализа нестационарных временных рядов
3. Теоретическое описание метода сингулярного спектрального анализа
4. Практические аспекты метода «Гусеница»-SSA
5. Декомпозиция на основе вейвлетов, ее самая современная версия
6. Алгоритм эмпирической модовой декомпозиции
7. Построение спектрограмм ряда
8. Методы и алгоритмы поиска частотно-временных характеристик компонент

Список литературы

1. Golyandina, N., and A. Zhigljavsky Singular Spectrum Analysis for time series. — Springer Briefs in Statistics. — Springer. — 2013. — 120 с. ISBN 978-3-642-34912-6.
2. Голяндина Н.Э. Метод «Гусеница»-SSA: анализ временных рядов: Учеб. пособие. — СПб. — 2004. — 76 с.
3. Huang N. E. The Hilbert-Huang transform and its applications / Ed. By S. S. Shen. Interdisciplinary mathematical sciences. 5 Toh Tuck Link, Singapore 596224: World Scientific Publishing Company Co. Pte. Ltd. — 2005. — P. 311. ISBN: 9812563768.
4. Добеши И. Десять лекций по вейвлетам. — Ижевск: РХД. — 2001. — 464 с.
5. Yeh J.R., Shieh J.S., Huang N.E. Complementary Ensemble Empirical Mode Decomposition: A Novel Noise Enhanced Data Analysis Method. — Advances in Adaptive Data Analysis. 2010. — Vol. 2, no. 2. — P. 135–156.
6. Вайнштейн Л. А., Вакман Д. Е. Разделение частот в теории колебаний и волн. М: Наука, 1983. — С. 287.

Часть 1. Адаптивный анализ нестационарных временных рядов

Все методы анализа и декомпозиции ВР, рассмотренные до сих пор, опирались только на их важное свойство – свойство стационарности. Напомним, что в широком смысле стационарные временные ряды имеют мат. ожидание и дисперсию (да и другие статистические показатели) в качестве функции, не зависящей от времени. Если это не так, то ВР должен считаться нестационарным.

Что вообще означает зависимость по времени для выборки случайного процесса? Это значит, что в каждой точке временной сетки, в которой был наблюдаем отсчет ВР, значение какой-либо его характеристики будет зависеть от этого момента наблюдения. На практике, большинство ВР имеют компоненты, у которых параметры и характеристики (в том числе мат. ожидание и дисперсия) зависят от времени и являются функциями от времени. Для их описания придется использовать не средние значения, а некоторые **мгновенные значения**, то есть такие, которые однозначно определяются в заданный отсчет временной сетки. Поэтому и вместо оценки средних значений для анализа нестационарных временных рядов будет необходимо строить оценки **частотно-временных характеристик** составляющих их компонент.

Вообще говоря, класс нестационарных процессов и класс процессов с изменяющимися частотно-временными характеристиками, не совпадают, их эквивалентность теоретически до сих пор не доказана. Но достоверно известно, что все нестационарные ВР имеют изменяющиеся частотно-временные характеристики, обратное не верно. Как же тогда можно анализировать и, главное, декомпозировать подобные нестационарные ряды?

Анализом нестационарных временных рядов занимается особый класс методов, называемых **методами адаптивного анализа временных рядов**. Этот вид методов **не требует** от исходных ВР предположений о **стационарности**, как, впрочем, и каких-либо априорных знаний и предположений о его свойствах *вообще*.

Откуда взялось прилагательное «*адаптивный*» в названии этих методов? Впервые этот термин для обозначения целого класса методов, занимающихся анализом НВР, был предложен Н. Хуангом в 1998 году. За этим названием лежит следующее объяснение. Адаптивные методы призваны декомпозировать и анализировать свойства ВР с изменяющимися во времени параметрами. Для оперативного отслеживания этих изменений необходимо, чтобы и сами методы могли без вмешательства человека **адаптировать** свои параметры, методику и построение алгоритма прямо по ходу анализа ВР. Для достижения этих условий, адаптивные методы, как мы увидим, имеют очень малое число начальных параметров (не более 2), а все необходимые критерии и оценки строят на основе характера самого анализируемого ВР и его компонент.

В связи с этим *адаптивные методы анализа* весьма эффективно применимы, в том числе, и для анализа стационарных временных рядов. Более того, так как у них заведомо меньше параметров, чем у подобных стационарных методов, их можно намного быстрее использовать для исходной оценки параметров, делимости, числа компонент и их состава внутри ряда. По этой причине все представленные здесь методы получили очень широкое распространение, как универсальные средства анализа и декомпозиции временных рядов.

Часть 2. Метод «Гусеница»-SSA. Сингулярный спектральный анализ

Одним из первых и ключевых подобных методов является так называемый метод **Сингулярного Спектрального Анализа** (ССА или SSA). Благодаря Голяндиной Н.Э. в русскоязычной литературе метод получил название «Гусеница». Мы будем использовать скорее аббревиатуру SSA, по которой проще искать дополнительные материалы на английском.

Метод SSA основан на преобразовании одномерного временного ряда в многомерный с помощью параметрической сдвиговой процедуры (отсюда и название «Гусеница») с последующим применением к многомерному временному ряду метода главных компонент PCA (*Principal component analysis*). При этом происходит исследование полученной многомерной траектории с помощью сингулярного разложения и восстановления (аппроксимации) ряда по выбранным главным компонентам. Целью метода является разложение временного ряда на интерпретируемые аддитивные составляющие. При этом метод SSA не требует стационарности анализируемого ВР, знания модели тренда, а также сведений о наличии в ряде периодических составляющих и их периодах, что очень удобно при анализе сложных ВР. Важно только отметить, что, тем не менее, исходная используемая в методе SSA модель является все-таки *аддитивной*.

Пусть есть исследуемый ненулевой временной ряд $F = f(t) = \{f(t_0), \dots, f(t_{N-1})\}$ с числом отсчетов N , $t_i = i\Delta t$, где Δt – период выборки. Определим еще заранее параметр метода – длину окна L , $1 < L < N$. Тогда алгоритм метода SSA в целом состоит из 2 этапов и 4 шагов (по 2 шага на этап): этап **разложения** и этап **восстановления**.

2.1. Этап разложения. Шаг вложения и шаг сингулярного разложения

Итак, начнем с этапа разложения. Первый шаг в методе SSA называется **вложение**. Процедура вложения переводит исходный временной ряд в последовательность многомерных векторов. Процедура вложения образует $K = N - L + 1$ векторов вложения, имеющих размерность L .

$$X_i = (f_{i-1}, \dots, f_{i+L-2})^T, 1 \leq i \leq K \quad (6.1)$$

X – **траекторная матрица** ряда F и состоит из векторов вложения в качестве столбцов $X = [X_1 : \dots : X_K]$.

По-другому траекторную матрицу можно представить в более точном виде выражения (6.2):

$$X = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} f_0 & f_1 & f_2 & \cdots & f_{K-1} \\ f_1 & f_2 & f_3 & \cdots & f_K \\ f_2 & f_3 & f_4 & \cdots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \cdots & f_{N-1} \end{pmatrix} \quad (6.2)$$

Очевидно, что $x_{ij} = f_{i+j-2}$ и поэтому матрица X имеет одинаковые элементы на «диагоналях» $i + j = \text{const}$. Таким образом, траекторная матрица является ганкелевой. Существует взаимно-однозначное соответствие между ганкелевыми матрицами размерности $L \times K$ и рядами длины $N = L + K - 1$.

Как видно, на первом шаге мы просто строим траекторную многомерную матрицу на основе одномерного ВР. На втором шаге мы производим **сингулярное разложение** (Singular Value Decomposition = SVD) траекторной матрицы ряда.

Пусть $S = XX^T$. Обозначим $\lambda_1, \dots, \lambda_L$ собственные числа матрицы S , взятые в неубывающем порядке ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$), и U_1, \dots, U_L – ортонормированную систему собственных векторов матрицы S , соответствующих собственным числам.

Пусть $d = \max\{i: \lambda_i > 0\}$. Если обозначить $V_i = X^T U_i / \sqrt{\lambda_i}, i=1, \dots, d$, то сингулярное разложение матрицы X может быть записано как

$$X = X_1 + \dots + X_d, \quad (6.3)$$

где $X_i = \sqrt{\lambda_i} U_i V_i^T$. Каждая из этих матриц X_i имеет ранг 1. Поэтому их можно назвать элементарными матрицами. Набор $(\sqrt{\lambda_i}, U_i, V_i)$ называют i -й **собственной тройкой** сингулярного разложения (6.3).

На этом заканчивается первый этап разложения исходного временного ряда на собственные тройки. После первого этапа используются всевозможные алгоритмы и методики, позволяющие выбрать сумму (6.3) таким образом, чтобы она наиболее близко отражала искомую модель ВР в виде тренда, сезонов и компонент. Когда такая комбинация собственных троек будет найдена (об этом мы будем говорить отдельно), начинается второй этап метода SSA – **восстановление**.

2.2. Этап восстановления. Шаг группировки и шаг усреднения

На этом этапе третий шаг называется **группировкой** и отражает как раз процесс приведения (6.3) в комбинацию искомых компонент модели ВР. На основе разложения (6.3) процедура группировки, по сути, делит все множество индексов $\{1, \dots, d\}$ на m непересекающихся подмножеств I_1, \dots, I_m .

Пусть $I = \{i_1, \dots, i_p\}$. Тогда результирующая матрица X_I соответствующая группе I , определяется как $X_I = X_{i_1} + \dots + X_{i_p}$. Такие матрицы вычисляются для $I = I_1, \dots, I_m$, тем самым разложение (6.3) может быть записано в сгруппированном виде:

$$X = X_{I_1} + \dots + X_{I_m} \quad (6.4)$$

Процедура выбора множеств $I = I_1, \dots, I_m$ и называется **группировкой** собственных троек. Это особый этап, на котором чаще всего возникают ошибки.

На последнем четвертом шаге алгоритма SSA, называемом **диагональным усреднением**, каждая матрица сгруппированного разложения (6.4) переводится в новый ряд длины N .

Пусть Y – некоторая $L \times K$ матрица с элементами y_{ij} , где $1 \leq i \leq L, 1 \leq j \leq K$. Положим $L^* = \min(L, K), K^* = \max(L, K)$ и $N = L + K - 1$. Пусть $y_{ij}^* = y_{ij}$, если $L \leq K$ и $y_{ij}^* = y_{ji}$ – иначе. **Диагональное усреднение** переводит матрицу Y во временной ряд g_0, \dots, g_{N-1} по формуле:

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m, k-m+2}^* & 0 \leq k < L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+2}^* & L^* - 1 \leq k < K^*, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m, k-m+2}^* & K^* \leq k < N. \end{cases} \quad (6.5)$$

Выражение (6.5) соответствует усреднению элементов матрицы вдоль диагоналей $i + j = k + 2$: выбор $k = 0$ дает $g_0 = y_{11}$, для $k = 1$ получаем $g_1 = (y_{12} + y_{21}) / 2$ и т.д. Заметим, что если матрица Y является траекторной матрицей некоторого ряда (h_0, \dots, h_{N-1}) (то есть если эта матрица является ганкелевой), то $g_i = h_i$ для всех i .

Применяя диагональное усреднение (6.5) к результирующим матрицам X_{I_k} получаем ряды $F^{(k)} = (f_0^{(k)}, \dots, f_{N-1}^{(k)})$ и, следовательно, исходный ряд $F = (f_0, \dots, f_{N-1})$ раскладывается в сумму m рядов $F = \sum_{k=1}^m F^{(k)}$.

Если компоненты ряда разделимы и собственные тройки сгруппированы правильно, то все матрицы в разложении (6.4) являются ганкелевыми, и декомпозиция исходного ряда в аддитивную сумму пройдет без ошибок. На практике какие-либо из этих условий будут нарушаться, а потому и

используется процедура диагонального усреднения, которая позволяет «сгладить» выраженные особенности траекторных троек.

Шаг группировки очень тесно связан с понятием **разделимости** ВР, которое мы рассматривали ранее. С учетом определений видов разделимости, приведенных выше, было математически установлено, что при **точной разделимости** метод SSA способен декомпозировать ВР только на совокупность константы (среднего уровня) и строго гармонических компонент. При **асимптотической разделимости** ситуация существенно улучшается – только долгосрочные тенденции не могут быть разбиты на составные части. А самое главное, что при асимптотической разделимости, становится возможным разделение детерминированной и случайной составляющих. При **приближенной разделимости** можно декомпозировать ВР на самые мельчайшие детали, но, тем не менее, остается одно важное ограничение метода SSA. Дело в том, что, как и в автокорреляционных методах, разделимость близких частот оказывается зависящей от параметра метода – его длины окна L . В целом, при разделимости сезонных и циклических компонент на основе метода SSA действует правило, что компоненты с частотами ω_1 и ω_2 разделимы, если частотно друг от друга они удалены на более чем $\omega = 1 / \min(L, K)$, то есть выбор длины окна L влияет на возможность разделения компонент.

Для метода SSA есть один очевидный параметр – длина окна L , и не очень очевидный параметр – способ группировки собственных троек и соответствующих им компонент согласно (6.4). В силу адаптивности метода, оба этих параметра должны быть выбраны на основе характера самого изучаемого ряда, нежели чем на основе каких-либо статистических критериев и оценок.

2.3. Выбор способа группировки компонент в методе «Гусеница»

Сначала найдем способ группировки компонент. Пусть длина окна фиксирована и равна L . В ходе алгоритма было выполнено уже два шага и на руках у нас имеются тогда собственные тройки исходного ряда. Как нам выбрать и сгруппировать их между собой?

Есть несколько свойств группировки, которые нам помогут:

- 1) Если сингулярные вектора имеют некоторую форму, то и восстановленные компоненты будут иметь схожую форму. То есть, если собственный вектор имеет линейную тенденцию, то и компонента имеет линейный тренд. Самое главное, **синусоидальные собственные вектора порождают тригонометрические составляющие ряда.**
- 2) Если восстановить ряд с помощью нескольких собственных троек с одним и тем же диапазоном частот, то и в периодограмме восстановленного ряда пик энергии будет находиться в том же диапазоне частот.
- 3) **Чем больше сингулярное значение, тем больше вклад соответствующей компоненты ряда.**

Опираясь на эти правила можно описать следующую методику группировки компонент:

- 1) Для выделения **тренда** из ряда нужно собрать все собственные тройки с медленно меняющимися сингулярными векторами. В простейшем случае, тренд выделяется с помощью **первой** собственной тройки, которая, по построению алгоритма, будет иметь **наибольшее сингулярное значение.**

- 2) **Первые** сингулярные значения соответствуют **низкочастотным** колебаниям, с ростом номера собственной тройки растет их частота.
- 3) Если ряд содержит несколько гармоник, то каждая из них может быть идентифицирована на основе графиков их сингулярных векторов. Более того, **каждой гармонике** соответствует **два собственных числа**, довольно близких по своему значению, и, соответственно, **две собственные тройки**. То есть для группировки **тригонометрических компонент** нужно использовать **не менее двух** собственных троек.
- 4) **Периодические компоненты** общего вида группируются на основе всех собственных троек, которые соответствуют среднему периоду этой компоненты, определенной по периодограмме ряда. Это связано с тем, что при условии строгой разделимости компонент, близкие периоды неразделимы и должны быть отнесены к одной компоненте.
- 5) **Модулированные циклы** и сезоны имеют периоды, сконцентрированные возле пиков периодограммы, тем самым, их восстановление аналогично группировке тригонометрических компонент, за исключением того, что диаграммы их собственных векторов будут отличаться от правильных и симметричных фигур.
- 6) В целом группировка компонент происходит согласно аддитивной модели ВР из лекции 1, выражение (1.3).
- 7) Что касается **шума**, то по построению алгоритма декомпозиции, **все не сгруппированные собственные тройки** будут относиться к шуму. То есть шум можно получить двумя способами – вычитанием из исходного ВР восстановленного на 4 шаге после группировки ряда, либо на основе группировки собственных троек, которые были отвергнуты перечисленной выше методикой. Оба этих метода полностью эквивалентны.

В целом, методика группировки выглядит достаточно громоздко. В упрощенной модификации метода SSA, называемой Auto-SSA, вся эта методика полностью заменяется автоматизированными средствами анализа периодограммы ряда, на основе которой определяются наиболее значимые периоды компонент, которые затем соотносятся с периодами собственных векторов, и на этой основе происходит группировка.

Есть и другие методики группировки компонент: на основе сингулярных чисел, на основе w -корреляций, и т.д. Мы возьмем из них только несколько рекомендаций, которые нам могут пригодиться.

Во-первых, так как **сингулярные числа** всегда располагаются в **убывающем порядке** (от наибольших к меньшим), то на основе их расположения можно производить группировку. **Одиночные** большие отдельные сингулярные числа относить к тренду. **Парные** близкие, почти равные сингулярные числа относить к периодикам. Периодики объединять в модулированные циклы только для **близко расположенных** сингулярных чисел.

Во-вторых, с **ростом номера** сингулярной тройки, то есть с уменьшением модуля значения сингулярного числа, растет вероятность того, что эта собственная тройка относится **к шуму**. То есть собственные тройки с большими номерами можно вообще не исследовать, а сразу относить к шуму.

2.4. Выбор длины окна L компонент в методе «Гусеница»

Длина окна L является основным параметром алгоритма SSA. Выбор длины окна должен явно зависеть от самого изучаемого ряда. Рассмотрим сначала общие правила выбора длины окна L :

- 1) Для анализа структуры ряда **не имеет смысла брать длину окна больше половины** длины ряда. Это связано с тем, что на 4 этапе усреднения мы все равно выбираем минимальное значение из двух длин окна.
- 2) **Чем больше длина окна, тем более детальное разложение** исходного ряда, и **тем больше собственных троек** мы получим.
- 3) **Маленькая** длина окна приводит к **смешиванию** компонент.
- 4) Для выявления конкретной **периодики** ряда лучше брать длину окна, кратную периоду этой компоненты.

Наилучшей методикой выбора длины окна является схема построения разности последовательных собственных чисел от длины окна L , предложенная Ф. Рабайя:

- 1) Для заданного ВР выбираются длины окна L от 10-20 отсчетов до половины общего числа отсчетов ряда. $L = \overline{10..N/2}$
- 2) Для каждой длины окна вычисляются разности пар соседних собственных сингулярных чисел. $|\lambda_{i+1} - \lambda_i|^{(L)}$
- 3) Строятся зависимости разности последовательных собственных сингулярных чисел от длины окна L .

$$\begin{aligned} f_1(L) &= |\lambda_2 - \lambda_1|(L) \\ f_2(L) &= |\lambda_3 - \lambda_2|(L) \\ &\dots \\ f_k(L) &= |\lambda_{k+1} - \lambda_k|(L) \end{aligned}$$

- 4) На основе анализа этих графиков происходит выбор длины окна L и выбор группировки компонент.

Как из таких графиков разности получить длину окна и способ группировки? Все оказывается достаточно просто. Из этих графиков можно получить множество оптимальных длин окна, просто выбирая те точки абсцисс, которые имеют разность собственных сингулярных чисел близкую к нулю (или минимальную), для графиков, обладающих некоторой явной периодикой. Пересекая между собой эти множества точек, можно найти такую **длину окна**, которая будет наилучшей, по принципу **минимизации разности** пар собственных сингулярных чисел.

Во-вторых, из тех же графиков разности можно выявить и способ группировки. Оказывается, **периодические компоненты** будут иметь периодические функции разности пар собственных чисел **с тем же самым периодом**. Такие пары собственных сингулярных чисел надо объединять для получения одной компоненты с этим периодом.

В-третьих, пары сингулярных чисел, которые **нельзя группировать** друг с другом, будут иметь **монотонные возрастающие графики функций** разности от длины окна.

Таким образом, для метода SSA разработана эффективная методика, которая позволяет применять его в качестве адаптивного средства анализа нестационарных и стационарных временных рядов. Как мы увидим в дальнейшем, этот метод можно использовать, в том числе, и для прогнозирования временных рядов.

Часть 3. Метод вейвлет-разложения временных рядов

Вейвлет-преобразование появилось уже достаточно давно, но стало активно применяться для анализа ВР, как стационарных, так и нестационарных, в конце 20-го века. Оно уже успело зарекомендовать себя и в других областях, как эффективное средства для решения целого класса задач, связанных с подавлением шумов, сжатием данных, обработкой и синтезом сигналов, анализом изображений и т.д.

Вейвлет-преобразование очень похоже по своему характеру на преобразование Фурье, поэтому и методика его использования напоминает алгоритмы анализа ВР спектральными методами. Главное отличие этих двух преобразований заключается в том, что преобразование Фурье использует в качестве базиса гармонические функции фиксированной частоты, а вот вейвлет-преобразование – целый класс различных базисных функций различного масштаба и локализации по времени. Наличие этих двух параметров позволяют «развернуть» свойства процесса одновременно и в частотной, и во временной областях. Поэтому подобный тип методов называют средством *частотно-временной локализации* особенностей ряда.

Теоретическую основу любого вейвлет-анализа определяет выражение **непрерывного вейвлет-преобразования** (Continuous Wavelet Transform - CWT):

$$W_{\Psi}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \Psi^* \left(\frac{t-b}{a} \right) dt \quad (6.6)$$

где a – задает масштабирование и называется параметром растяжения, b – соответствует временному сдвигу и называется параметром положения, Ψ – материнский вейвлет или базис вейвлет-преобразования. Вариации a и b дают как раз разную частотно-временную локализацию.

Для выражения (6.6) существует и обратное преобразование

$$x(t) = C_{\Psi}^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_{\Psi}(a, b) \psi_{a, b}(t) \frac{da db}{a^2}, \quad (6.7)$$

где C_{Ψ} – постоянная величина и зависит только от вида Ψ :

$$C_{\Psi} = 2\pi \int_{-\infty}^{\infty} |\Psi(\xi)|^2 \xi^{-1} d\xi. \text{ Формула (6.7) может трактоваться двумя способами}$$

– как способ восстановления исходного ряда по его вейвлет-преобразованию, либо как способ записи ВР в виде суперпозиции вейвлетов.

3.1. Схема вейвлет-декомпозиции

Аналогично автокорреляционным и спектральным методам, для анализа ВР нам требуются не эти теоретические определения вейвлетов, а их дискретные формы записи с учетом конечного отрезка времени и числа отсчетов. Подобное вейвлет-преобразование называется **дискретным (DWT)** и определяется в виде:

$$W_{m, n} = a_0^{-m/2} \int x(t) \psi(a_0^{-m} t - nb_0) dt \quad (6.8)$$

$$x(t) = K_{\Psi} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} W_{m, n} \psi_{m, n}(t) \quad (6.9)$$

где m и n – целые числа, K_{Ψ} – постоянная нормировки.

Выражение (6.9) уже очень похоже на представление исходного ВР в виде аддитивной смеси компонент подобно аддитивной модели. Надо только правильно осуществить выбор параметров a_0, b_0 и функции ψ . На практике DWT дорабатывают до некоторой итерационной схемы, построенной на основе *ортонормированных базисных функций* $\psi_{m, n}$. В частности, для $a_0 = 2$ и $b_0 = 1$ существует целое семейство функций ψ , обладающих нужными для нас свойствами частотно-временной локализации (например, базис Хаара).

Для декомпозиции ВР на составные компоненты с помощью вейвлет преобразования DWT используют **схему фильтрации**. В самом деле, выражение (6.8) напоминает свертку преобразования Фурье с заданной оконной функцией. Поэтому в целом методика оказывается аналогичной.

Сначала исходный ВР $x(t)$ пропускается через НЧ фильтр с импульсным откликом g и через ВЧ фильтр с импульсным откликом h , в результате чего получаются свертки:

$$\begin{aligned} y[n] &= (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]g[n-k] \\ y[n] &= (x * h)[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] \end{aligned} \quad (6.10)$$

Эти фильтры связаны между собой и называются квадратурными зеркальными фильтрами **QMF**. В результате свертки получаются **детализирующие** коэффициенты cD_i (после ВЧ-фильтра) и коэффициенты **аппроксимации** cA_i (после НЧ-фильтра). Так как половина частотного диапазона сигнала была отфильтрована, то, согласно теореме Котельникова, отсчёты сигналов можно **проредить** в 2 раза (т.е. провести их «децимацию» *downsampling*):

$$\begin{aligned} y_{low}[n] &= \sum_{k=-\infty}^{\infty} x[k]g[2n-k] \\ y_{high}[n] &= \sum_{k=-\infty}^{\infty} x[k]h[2n-k] \end{aligned} \quad (6.11)$$

Первый шаг такой схемы (6.11) представлен на рисунке 6.1.

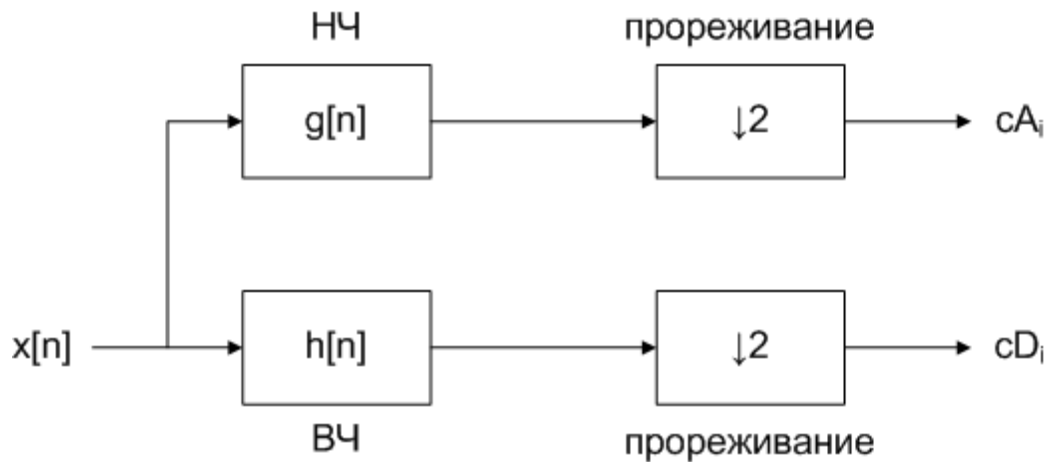


Рисунок 6.1 – Схема разложения ряда согласно (6.11)

На следующем шаге к детализирующим коэффициентам cD_i применяется свертка, в результате чего выстраивается двоичное дерево фильтров (рис. 6.2). Это дерево имеет несколько уровней, его листья и узлы соответствуют пространствам с различной частотно-временной локализацией. Варьируя число каскадов, можно получать разное число детализирующих коэффициентов cD_i и, меняя уровень декомпозиции исходного временного ряда, коэффициентов аппроксимации cA_i .

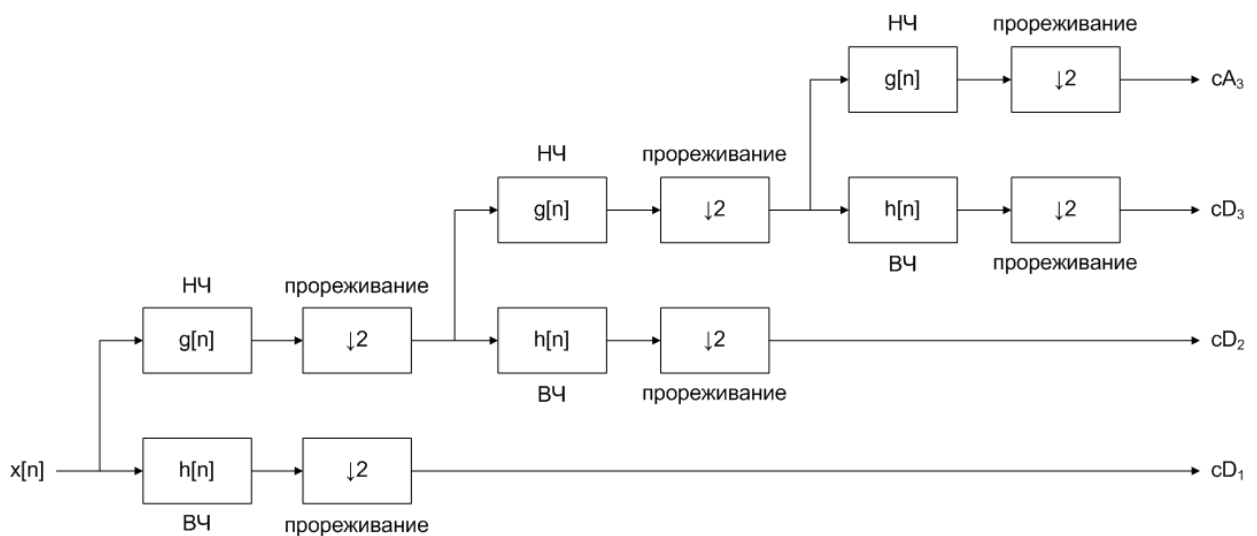


Рисунок 6.2 – Три уровня дерева вейвлет-декомпозиции

Несмотря на кажущуюся сложность такой схемы декомпозиции временных рядов, ее реализация оказывается достаточно простой и, самое главное, весьма быстрой. Тем не менее, у чистой вейвлет-декомпозиции на основе DWT есть несколько недостатков. Во-первых, это наличие только одного направления декомпозиции. То есть, чтобы мы не делали, без изменения базисных параметров вейвлета другой декомпозиции не получить. В отличие от аналогичных методов декомпозиции НБР, где есть этап группировки компонент, здесь мы не сможем адаптировать полученный результат к искомой модели. Во-вторых, найденные значения коэффициентов аппроксимации cA_i и деталей cD_i теоретически не всегда являются оптимальными. Все эти проблемы, как мы увидим в дальнейшем, позволяет решить *адаптивная* версия вейвлет-декомпозиции, называемая **вейвлетной пакетной декомпозицией** (Wavelet Packet Decomposition - **WPD**).

3.2. Вейвлетная пакетная декомпозиция WPD

Вейвлетная пакетная декомпозиция WPD является модернизированной версией предыдущей схемы декомпозиции, поэтому она отличается только в нескольких существенных деталях:

- 1) Повторная свертка и фильтрация применяется не только к детализирующим коэффициентам cD_i , но и к коэффициентам аппроксимации cA_i , в результате чего строится полное двоичное дерево декомпозиции (см. рис. 6.3).
- 2) Из полного двоичного дерева выбирается только одна ветвь декомпозиции, но не по принципу возрастания/убывания, а таким образом, чтобы конечная ветвь имела наименьшее значение энтропии среди всех ветвей двоичного дерева.
- 3) Базисный вейвлет выбирается таким образом, чтобы его дочерние вейвлеты для фильтрации было возможным вычислить до этапа построения дерева, обратно пропорционально степени двойки, что позволяет снизить сложность задачи до $O(N \log_2 N)$.
- 4) Масштаб и смещение базисного вейвлета выбираются **адаптивно** на основе той ветви дерева, что обладает наименьшей энтропией.

Подобные улучшения, лежащие в основе методики WPD, позволили сделать метод вейвлет-анализа *адаптивным*, при этом существенно снизив вычислительные затраты на построение полного дерева декомпозиции. К достоинствам этого метода можно отнести высокую скорость, универсальность и возможность ручного выбора ветвей дерева декомпозиции и коэффициентов cD_i и cA_i . Недостатком метода является только **ручной выбор уровня** декомпозиции, так как с ростом этого параметра число ветвей растет, и выбор оптимальной ветви все больше усложняется. На практике, уровень ограничивают числом не более 9-10.

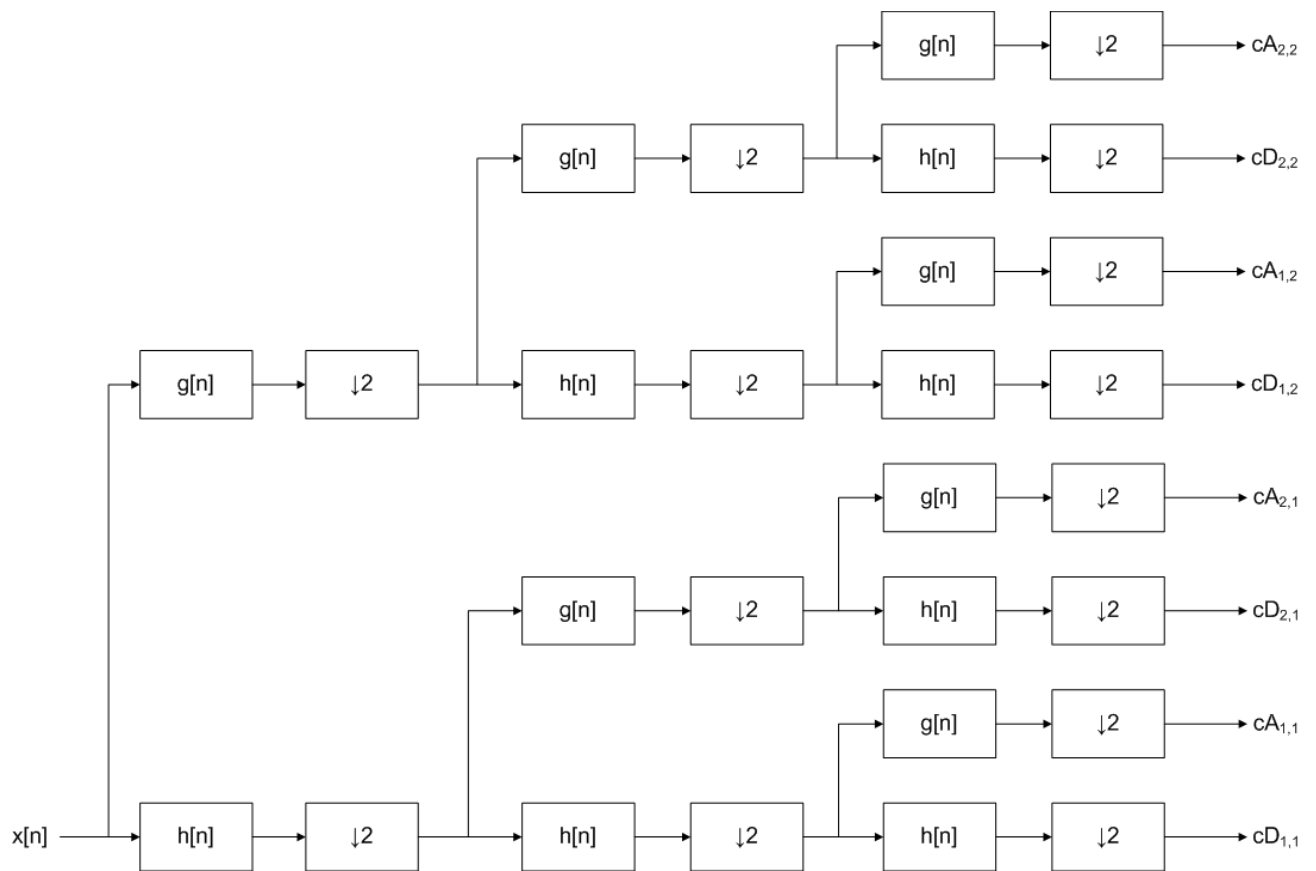


Рисунок 6.3 – Трехуровневое дерево WPD

Часть 4. Эмпирическая модовая декомпозиция

Одним из самых новых методов разбиения ВР, не требующих от него наличия свойства стационарности, является метод **Эмпирической Модовой Декомпозиции (Empirical Mode Decomposition, далее EMD)**, и как, мы увидим в дальнейшем, ее модификации.

Алгоритм EMD по своей базовой идее похож на сингулярный спектральный анализ: любой ВР $x(t)$ разделим на аддитивную смесь (6.12) компонент $c_j(t)$ (или **мод**, как принято в описании данного алгоритма) и остаточного ряда $r_n(t)$, описывающего случайную составляющую ВР.

$$x(t) = \sum_{i=1}^{n-1} c_i(t) + r_n(t) \quad (6.12)$$

Каждая из этих мод является некоторой базисной функцией и несет в себе информацию, соответствующую тому или иному реальному физическому процессу, участвующему в формировании анализируемого ВР. При этом набор базисных компонент, которые могут быть, вообще говоря, не ортогональными друг другу, не фиксирован, но является *адаптивным* и зависит только от вида исходных данных. В известной мере, эти компоненты можно сравнить с гармоническими функциями, используемыми при разложении в ряд Фурье. Однако, в отличие от них, компоненты ВР, выделенные с помощью ННТ, могут иметь амплитудную и частотную модуляцию, зависящую от времени, и быть не ортогональными друг другу.

Базовый алгоритм EMD может быть представлен в виде следующей последовательности шагов (с учетом всех изменений, что были внесены в методику построения разбиения от 2015 года):

- 1) В исходном ВР $x(t)$ находят все точки экстремума.
- 2) С помощью сплайнов (**кубических** или другой иной заданной формы) вычисляют две интерполяционные огибающие кривые,

проходящие, соответственно, через максимумы и минимумы анализируемого ряда. Верхнюю огибающую кривую, проходящую через максимумы, обозначим U , нижнюю огибающую кривую, проходящую через минимумы – L .

3) Вычислим среднюю кривую от этих двух огибающих: $m = (U + L) / 2$.

4) Вычтем это среднее из ВР $x(t)$: $r = x(t) - m$.

5) Повторить шаги 1-4 **десять раз**, либо до выполнения критерия отсеивания.

6) Полученный после отсеивания на шагах 1-5 остаток $r(t)$ принимается за найденную компоненту $c_i(t) = r$, где i – номер компоненты или число пройденных циклов алгоритма EMD.

7) Эта найденная компонента $c_i(t)$ вычитается из исходного ряда $x(t)$.

Полученный остаток принимается за новый ряд для декомпозиции:

$$\hat{x}(t) = x(t) - c_i(t)$$

8) Перейти к шагу 1. Повторять шаги 1-7 n раз, где n – число компонент, зависящее от длины N исходного ВР и определяемое как $n = \text{целое}[\log_2 N + 1]$.

Подробная блок-схема алгоритма EMD представлена на рисунке 6.4. В основе алгоритма EMD лежат только численные методы построения интерполяционных кривых, поэтому такой алгоритм как раз не требует от исходного ряда никакой стационарности. Метод EMD, в совокупности с другими средствами расчета частотно-временных характеристик найденных компонент, образует один из ключевых *адаптивных методов анализа* временных рядов.

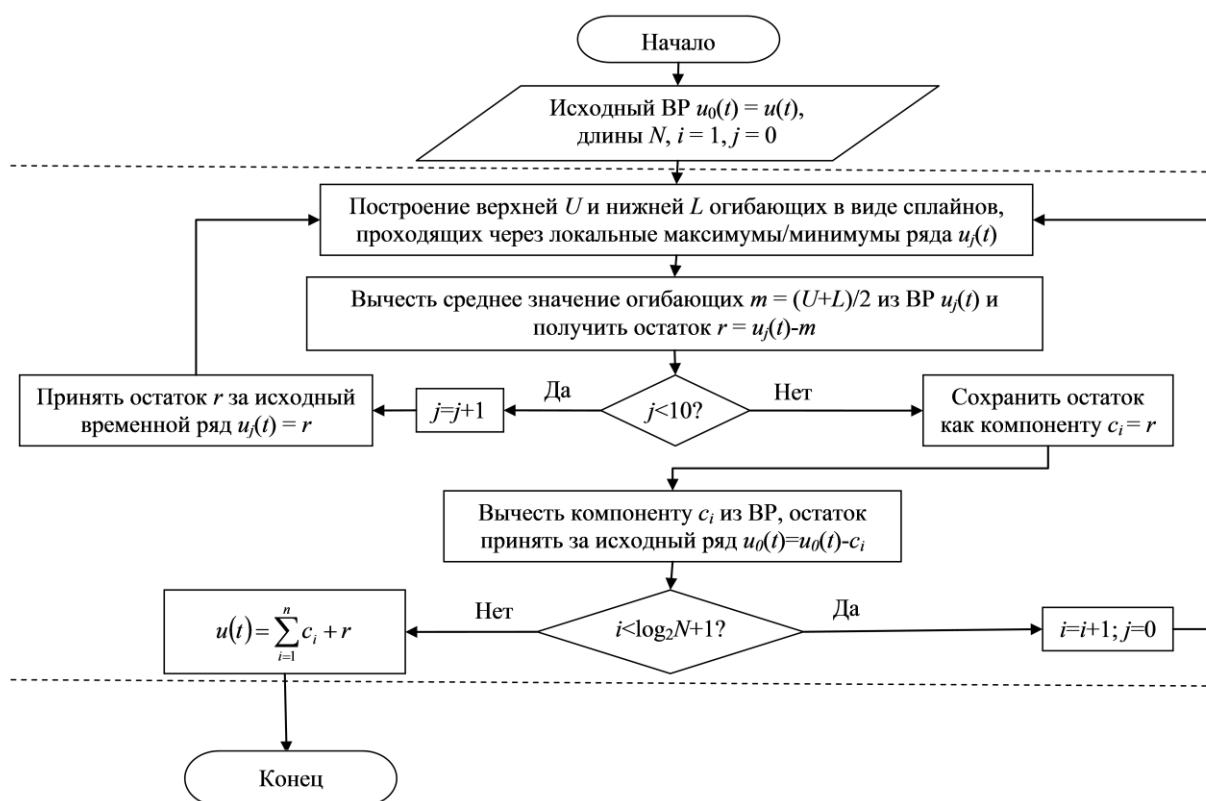


Рисунок 6.4 – Подробная схема алгоритма EMD

Рассмотрим простейший наглядный пример декомпозиции ряда с помощью EMD. Пусть он представляет собой отсчеты сигнала, состоящего из аддитивной суммы 4 гармонических компонент с разными частотами. В результате декомпозиции мы будем ожидать 4 компоненты и остаточный ряд, близкий к нулю. Так как число компонент зависит от длины исходного ряда как $n = \text{целое}[\log_2 N + 1]$, то компонент мы получим с излишком. В этом случае, подобно 3 шагу метода SSA, производится группировка этих компонент в соответствии с ожидаемой формой модели анализируемого ВР.

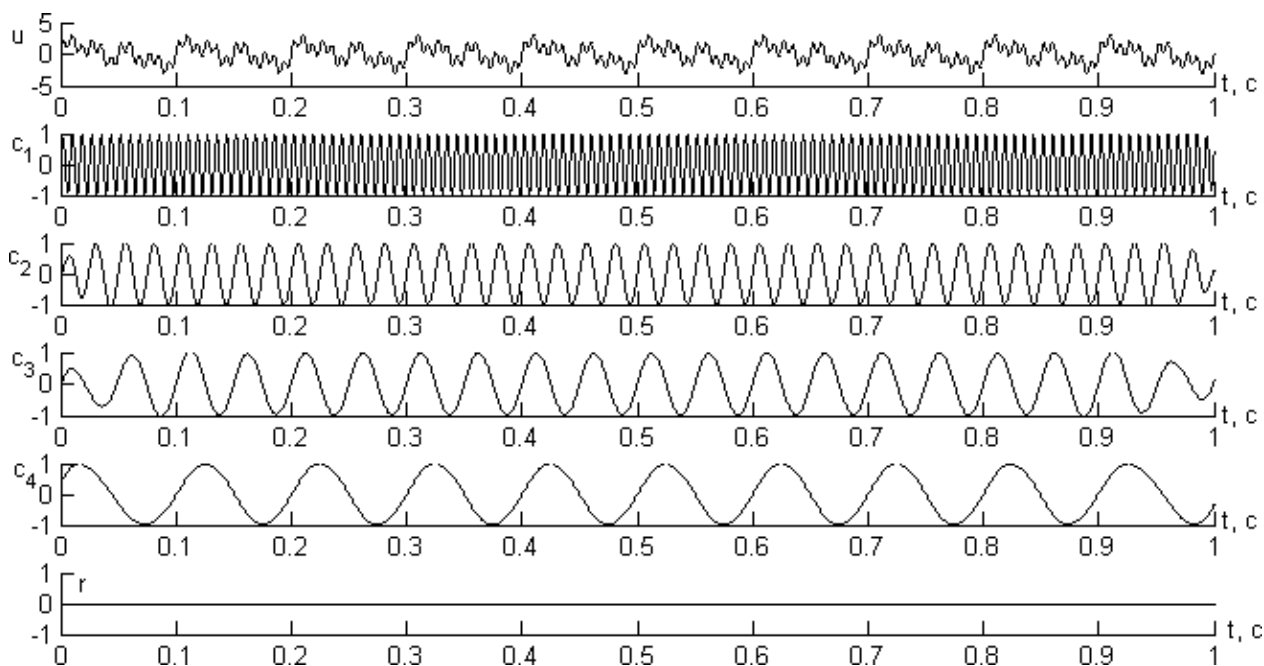


Рисунок 6.5 – Исходный ВР, его 4 компоненты и остаток

В нашем случае все получается достаточно просто: 4 гармонические компоненты выделяются за первые 4 шага декомпозиции, а все последующие моды, компенсируя друг друга, суммируются в остаточный ряд, близкий к нулю. Полученная декомпозиция представлена на рисунке 6.5. Как видно из него, результат оказывается положительным, но на практике такой идеальной картины получаться, конечно, не будет, так как наличие шума в исходном ряде всегда будет нарушать свойство точной разделимости компонент ряда.

Проблема метода EMD, связанная с искажением компонент (названная **эффектом смешивания мод**) при наличии в ВР шума, сильно ограничивала возможности по его применению для анализа реальных процессов, поэтому, со временем, появилась его новая модификация, названная Эмпирической Модовой Декомпозицией по Ансамблю (**Ensemble EMD = EEMD**). Оказалось, что если к анализируемому ВР *добавлять* белый шум с определенными характеристиками, а затем множество реализаций усреднять, то компоненты будут выделяться в соответствии с двоичной частотной фильтрацией, из-за чего смешивания мод можно будет избежать.

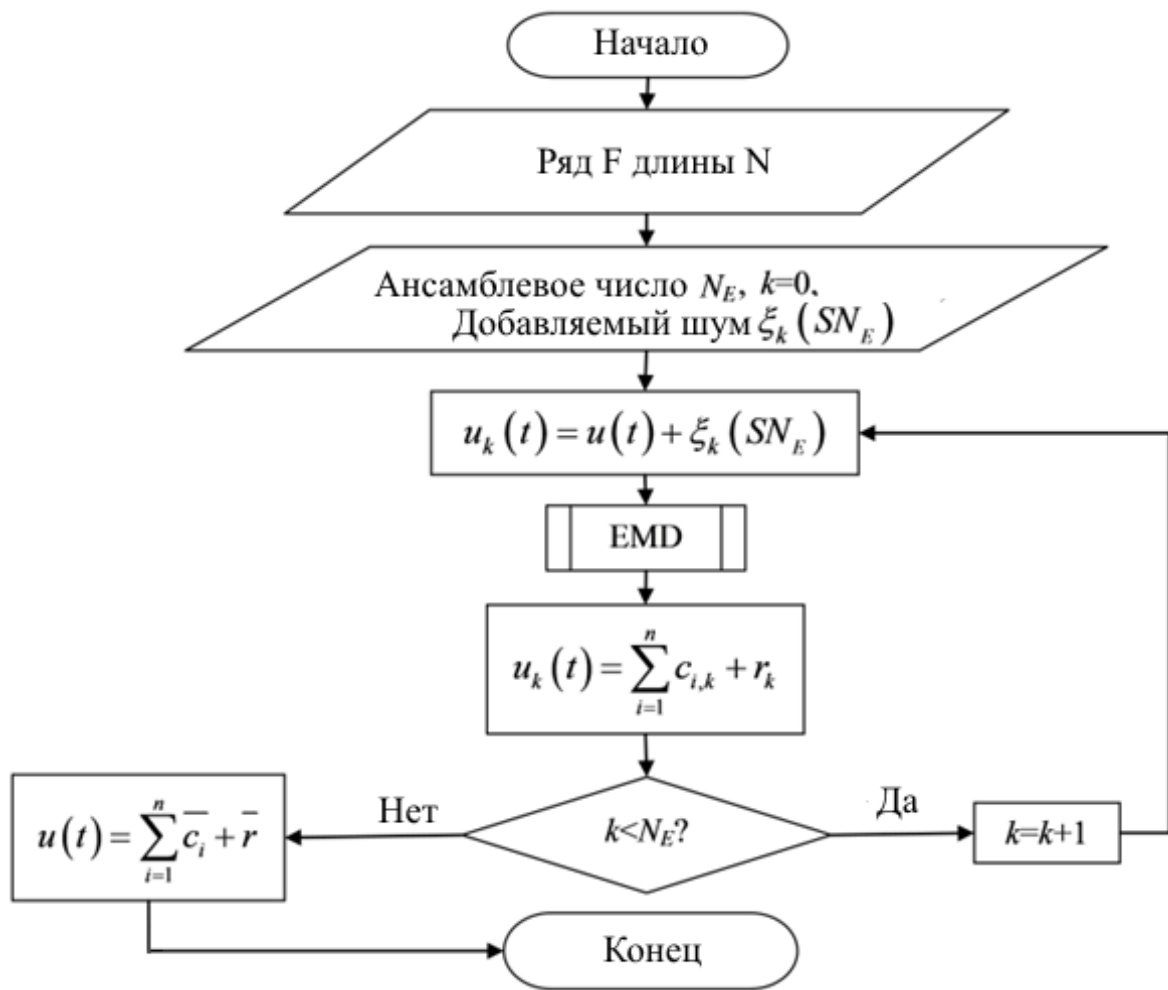


Рисунок 6.6 – Блок-схема алгоритма EEMD

Таким образом, был создан алгоритм декомпозиции по ансамблю EEMD, алгоритм которого может быть описан следующей последовательностью действий (при условии, что EMD уже реализован) (см. рис. 6.6):

1) Для ВР $u(t)$ длины N найти общее число компонент как $n = \text{целое}[\log_2 N + 1]$.

2) Нормировать исходный ВР $\tilde{u}(t) = \frac{u}{\sigma}$, где

$$\sigma = \left(\frac{1}{N-1} \sum_{i=1}^N (u_i - \bar{u})^2 \right)^{1/2}.$$

- 3) Определить два входных параметра – ансамблевое число N_E , и соотношение сигнал/добавленный шум SN_E . Не вдаваясь в подробности выбора этих параметров по разным методикам, можно определить по умолчанию $N_E = 100$, $SN_E = 20\text{дБ}$.
- 4) Добавить к нормированному ВР $\tilde{u}(t)$ выборку белого шума известного соотношения сигнал/шум SN_E .

$$u_\xi = \tilde{u}(t) + \xi_n$$

- 5) Декомпонировать полученный ряд u_ξ с помощью EMD.
- 6) Повторить шаги 4 и 5 заданное N_E раз. На каждой итерации такого цикла запоминается результат декомпозиции, а выборка шума добавляется каждый раз разная.
- 7) Усреднить по ансамблю выделенные моды/компоненты и денормировать. $c_i(t) = \tilde{c}_i(t) \cdot \sigma$.

Есть еще более новая и точная модификация метода, называемая **комплементарной EEMD (CEEMD)**, где на 4 шаге алгоритма создается два подмножества компонент – к которым добавляют выборку белого шума со знаком «плюс», и к которым добавляют ту же выборку белого шума со знаком «минус». На сегодняшний момент среди всего множества модификаций **CEEMD** является самой точной и эффективной методикой адаптивной декомпозиции временных рядов. Ее главным недостатком является высокая вычислительная сложность порядка $O(N^2)$. В настоящее время стали появляться работы по оптимизации этого метода с помощью распараллеливания самого процесса и его линеаризации, в результате чего удалось достичь ощутимого выигрыша в скорости – до $O(N \log_2 N)$. Так что в дальнейшем этот метод имеет неплохие шансы на развитие и применение в самых различных областях.

Часть 5. Частотно-временные характеристики временных рядов

Как можно было заметить по описанию выше перечисленных методов, все они производят адаптивную декомпозицию ВР. Но кроме декомпозиции ряда нас еще интересуют и их свойства, особенно, частотно-временные характеристики их компонент.

Лучшим средством анализа гармонических рядов является построение их спектра, то есть зависимости мощности от частот. Если сама частота при этом изменяется, то такой спектр не будет отражать ключевые свойства компоненты. Но что будет, если ввести еще одну ось – ось времени, тем самым изобразив частотно-временные характеристики в виде многомерного спектра? Эта идея и легла в основу первых реализаций по способу построения частотно-временных характеристик временных рядов.

5.1. Спектрограмма

Исходное преобразование Фурье, лежащее в основе спектральных методов, имеет фиксированное частотное разрешение, при этом происходит интегрирование по всем возможным значениям переменной времени. Поэтому есть два способа построения на основе его модификаций частотно-временных диаграмм. Первый способ порождает, так называемый, **мгновенный спектр мощности** сигнала, за счет расчета плотности распределения энергии сигнала в частотно-временных координатах:

$$\rho(t, \nu) = \int_{-\infty}^{\infty} X^* \left(\nu + \frac{\varphi}{2} \right) X \left(\nu - \frac{\varphi}{2} \right) e^{\frac{2\pi}{\varphi} t} d\varphi \quad (6.13)$$

Этот метод мы использовать не будем, поэтому подробно на нем останавливаться нет смысла. Второй способ предполагает использование оконного преобразования Фурье, с разбиением ВР на малые отрезки времени, на которых частоты считаются квази-постоянными. Строя спектр Фурье для каждого из этих малых окон, мы получим зависимость и от частоты, и от времени, то есть оценку частотно-временной характеристики ряда.

Подобный метод построения называется **спектрограммой** (*spectrogram*) и позволяет оценить частотно-временные характеристики ВР. На рисунке 6.5 приведен пример построения спектрограммы для ЛЧМ сигнала, изменяющего частоту линейно от 1 до 300 Гц на отрезке в 2 секунды.

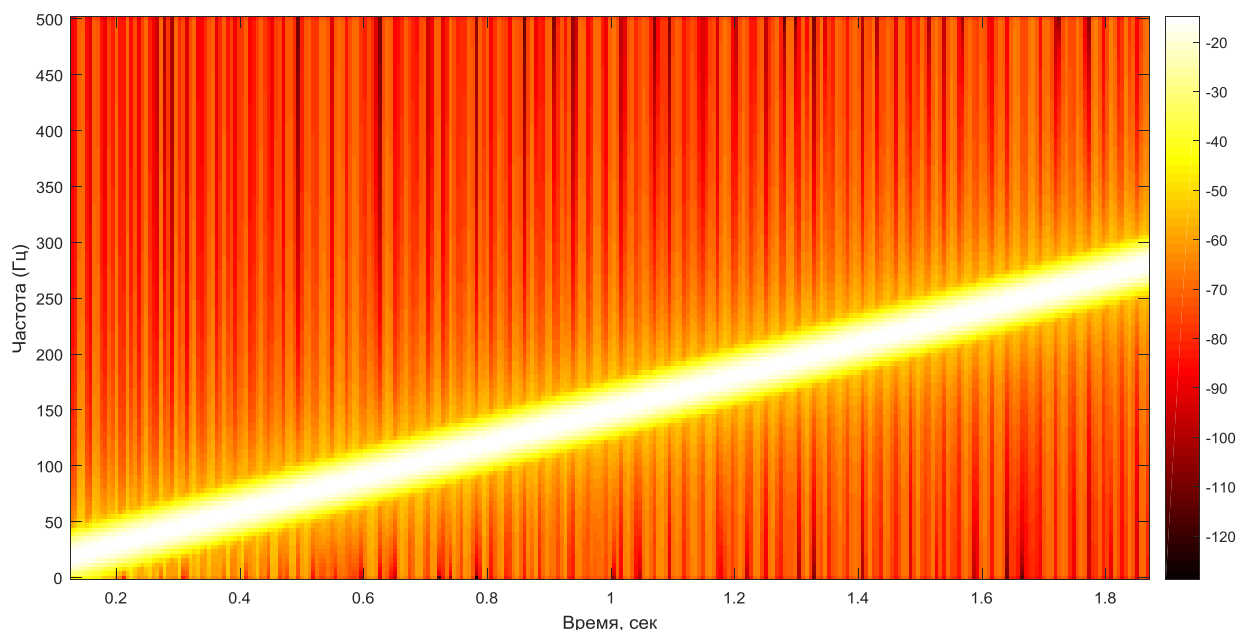


Рисунок 6.7 – Спектрограмма ЛЧМ сигнала

Из рисунка 6.7 видно, что частота (и соответственно период) ряда зависит от времени, можно определить характер этой зависимости (линейный), и даже оценить его параметры.

Благодаря тому, что спектрограмма опирается на хорошо известное, отработанное и быстрое (через FFT) преобразование Фурье, этот метод построения частотно-временных характеристик наиболее широко распространен. Тем не менее, у него есть ряд недостатков, который привел к появлению альтернативных методов оценки частотно-временных характеристик компонент. Во-первых, в каждый момент времени ширина спектра все равно остается фиксированной и имеет конечную длину. Во-вторых, частотно-временное разрешение спектрограммы остается постоянным и конечным. Поэтому кроме спектрограмм есть и другие эффективные методы расчета частотно-временных характеристик.

5.2. Мгновенная частота. Преобразование Гильберта

О том, что во временных рядах могут встречаться компоненты, у которых период и, соответственно, частота являются функциями от времени, было известно еще в начале 20-го века. Одним из способов описания характеристики таких рядов было бы представление понятия периода/частоты как функции от времени. Применительно к временным рядам, в каждой точке временной сетки, на которой определен ВР, нам требуется указать соответствующее значение периода/частоты. Поэтому такое представление периода/частоты в виде функции от времени называют **мгновенным периодом / мгновенной частотой**. Так как оба эти понятия взаимно однозначны, дальше будем использовать только понятие **мгновенной частоты**.

Проблема с мгновенной частотой состоит в том, что ее надо однозначно определять в текущий момент времени, что теоретически не совсем возможно на поле рациональных чисел. По этой причине в 1946 году Габором было предложено и введено понятие **аналитического сигнала**, позволяющего однозначно определить и описать понятия **мгновенной амплитуды, мгновенной частоты и мгновенной фазы** любого процесса.

Пусть у нас есть ВР $u(t)$, определенный на своем временном интервале. Тогда **аналитический сигнал (АС)** определяется для этого ВР, как

$$w(t) = u(t) + jv(t) = a(t)e^{j\phi(t)} \quad (6.14)$$

где $v(t)$ – ряд, сопряженный к исходному ряду **по Гильберту**:

$$v(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{u(\tau)}{t - \tau} d\tau = H(u) \quad (6.15)$$

Благодаря введению понятия АС через оператор **преобразования Гильберта** $H(u)$, мы можем для любого ряда $u(t)$ ввести такие понятия и функции от времени, как:

Огибающей сигнала (амплитудная модуляция ряда):

$$a(t) = |w(t)| = \sqrt{u^2(t) + v^2(t)}. \quad (6.16)$$

Фазы:

$$\varphi(t) = \arctg\left(\frac{v(t)}{u(t)}\right) = \arccos\left(\frac{u(t)}{a(t)}\right) = \arcsin\left(\frac{v(t)}{a(t)}\right). \quad (6.17)$$

И мгновенной частоты:

$$\omega(t) = \dot{\varphi}(t) = \frac{u(t)\dot{v}(t) - \dot{u}(t)v(t)}{a^2(t)}. \quad (6.18)$$

Стоит отметить, что, так как мгновенная частота рассчитывается как производная, то на точность ее расчета будут влиять методы дифференцирования, значения исходного ряда и методы построения огибающей кривой и сопряженного сигнала.

Получается, что после проведения декомпозиции ВР любым из адаптивных методов, представленных ранее, мы сможем получить функцию мгновенной частоты / мгновенного периода от времени в чистом виде, с максимально доступной точностью для заданной временной сетки ряда. Выражения (6.15) – (6.18) лежат в основе всех современных методов построения частотно-временных диаграмм характеристик временных рядов. Вариациям подвергаются только способы дифференцирования и метода дискретизации и расчета самого Преобразования Гильберта (нормированные методы, дискретные, частные, спектральные).

Несмотря на все преимущества преобразования Гильберта, есть у него и существенные недостатки. Во-первых, мгновенные функции всегда будут иметь краевые эффекты, то есть иметь искажения по краям временного интервала. Во-вторых, при дифференцировании могут возникать нули и бесконечности, что плохо. Но в настоящее время, среди всего множества методов оценки частотно-временных характеристик, преобразование Гильберта является самым оптимальным и эффективным средством.

5.3. Метод прямой квадратуры

Совсем недавно (2010 год) появился еще один новый способ построения функции мгновенной частоты от времени. Этот метод **прямой квадратуры** (*Direct Quadrature* - **DQ**) является полностью **эмпирическим**, поэтому **точность** его оказывается **низкой**, но он вполне может служить в качестве простой оценки частотно-временных характеристик временных рядов, особенно для **низкочастотных** периодик и циклов, для которых его точность даже превосходит точность преобразования Гильберта.

Методика этого подхода очень похожа на эмпирическую модовую декомпозицию:

- 1) Вычисляется новый ВР на основе исходного ряда, для чего все отсчеты берутся по модулю $|u(t)|$.
- 2) Строится эмпирическая огибающая кривая $e_1(t)$ с помощью интерполяции кубическими сплайнами, проходящая через максимумы ряда $|u(t)|$.
- 3) Исходный ВР $u(t)$ нормируется на эту огибающую: $u_1(t) = u(t) / e_1(t)$.
- 4) Шаги 2 и 3 повторяются, пока не будет выполняться условие $|u(t)| \leq 1$.

Полученный на последнем шаге ряд обозначим $u_n(t)$.

- 5) Вычисляем функцию АМ по формуле:

$$a(t) = \frac{u(t)}{u_n(t)} \quad (6.19)$$

- 6) Вычисляем квадратуру ряда: $\sin \varphi(t) = \sqrt{1 - u_n^2(t)}$.
- 7) Вычисляем **мгновенную частоту** по формулам (6.17) и (6.18).

Основой данного метода является получение аналитического сигнала не математическими выражениями, а в виде эмпирических оценок. Преимуществами метода являются отсутствие его параметров, простота реализации и скорость, платой за такую простоту и эмпиричность является низкая точность расчета функции мгновенной частоты.

Среди аналогичных методов оценки частотно-временных характеристик стоит еще отметить такие методы, как: обобщенный метод нулевых точек (*generalized zero crossing* - **GZC**), операторный метод расчета энергии сигнала (*teager energy operator* - **ТЕО**), и другие.