

## КУРС ЛЕКЦИЙ

### Тема 1. ВВЕДЕНИЕ В ЦИФРОВУЮ ОБРАБОТКУ СИГНАЛОВ

Цифровая обработка сигналов (ЦОС или DSP - digital signal processing) является одной из новейших и самых мощных технологий, которая активно внедрилась в широкий круг областей науки и техники: коммуникации, метеорология, радиолокация и гидролокация, медицинская визуализация изображений, цифровое аудио- и телевизионное вещание, разведка нефтяных и газовых месторождений, и многих других. Можно сказать, что происходит повсеместное и глубокое проникновение технологий цифровой обработки сигналов во все сферы деятельности человечества. Сегодня технология ЦОС относится к числу базовых знаний, которые необходимы ученым и инженерам всех отраслей без исключения.

Изучение принципов цифровой обработки сигналов не требует глубоких и всесторонних знаний по математике. Необходимо лишь немного опыта в элементарной алгебре, в способах декомпозиции (разложения) произвольных сигналов на составляющие и их обратной реконструкции. Почему же предмет считается сложным для понимания? Ответ в том, как материал преподносится для изучения. Одно дело писать уравнения, но другое - объяснить их значение с практической точки зрения. Именно в последнем и состоит цель данного курса.

Физические величины макромира, как основного объекта наших измерений и источника информационных сигналов, как правило, имеют непрерывную природу и отображаются непрерывными (аналоговыми) сигналами. Цифровая обработка сигналов оперирует с дискретными величинами, причем с квантованием как по координатам динамики своих изменений (во времени, в пространстве, и по любым другим изменяемым аргументам), так и по значениям физических величин. Математика дискретных преобразований зародилась в недрах аналоговой математики еще в 18 веке в рамках теории рядов и их применения для интерполяции и аппроксимации функций, однако ускоренное развитие она получила в 20 веке после появления первых вычислительных машин. В своих основных положениях математический аппарат дискретных преобразований подобен преобразованиям аналоговых сигналов и систем. Однако дискретность данных требует учета этого фактора, а его игнорирование может приводить к ошибкам. Кроме того, ряд методов дискретной математики не имеет аналогов в аналитической математике.

Стимулом развития дискретной математики является и то, что стоимость цифровой обработки данных меньше аналоговой и продолжает снижаться, а производительность вычислительных операций непрерывно возрастает. Немаловажным является и то, что системы ЦОС отличаются высокой гибкостью. Их можно дополнять новыми программами и перепрограммировать на выполнение различных операций без изменения оборудования. В последние годы ЦОС оказывает возрастающее влияние на все отрасли современной промышленности: телекоммуникации, средства информации, цифровое телевидение и пр. Интерес к научным и к прикладным вопросам цифровой обработки сигналов возрастает во всех отраслях науки и техники.

#### 1.1. ПРЕДИСЛОВИЕ К ЦИФРОВОЙ ОБРАБОТКЕ СИГНАЛОВ [52].

Цифровые сигналы формируются из аналоговых операций дискретизации – последовательным квантованием (измерением) амплитудных значений сигнала через определенные интервалы времени  $\Delta t$  или любой другой независимой переменной  $\Delta x$ . В результате равномерной дискретизации непрерывный по аргументу сигнал переводится в упорядоченную по независимой переменной последовательность чисел. В принципе разработаны методы ЦОС для неравномерной дискретизации данных, однако области их применения достаточно специфичны и ограничены. Условия, при которых возможно

полное восстановление аналогового сигнала по его цифровому эквиваленту с сохранением всей исходно содержащейся в сигнале информации, выражаются теоремами Найквиста, Котельникова, Шеннона, сущность которых практически одинакова. Для дискретизации аналогового сигнала с полным сохранением информации в его цифровом эквиваленте максимальные частоты в аналоговом сигнале должны быть не менее чем вдвое меньше, чем частота дискретизации, то есть  $f_{\max} \leq (1/2)f_d$ , т.е. на одном периоде максимальной частоты должно быть минимум два отсчета. Если это условие нарушается, в цифровом сигнале возникает эффект маскирования (подмены) действительных частот более низкими частотами. При этом в цифровом сигнале вместо фактической регистрируется "кажущаяся" частота, а, следовательно, восстановление фактической частоты в аналоговом сигнале становится невозможным. Восстановленный сигнал будет выглядеть так, как если бы частоты, лежащие выше половины частоты дискретизации, отразились от частоты  $(1/2)f_d$  в нижнюю часть спектра и наложились на частоты, уже присутствующие в этой части спектра. Этот эффект называется наложением спектров или алиасингом (aliasing). Наглядным примером алиасинга может служить иллюзия, довольно частая в кино – колесо автомобиля начинает вращаться против его движения, если между последовательными кадрами (аналог частоты дискретизации) колесо совершает более чем пол-оборота.

**Преобразование сигнала в цифровую форму** выполняется аналого-цифровыми преобразователями (АЦП). Как правило, они используют двоичную систему счисления с определенным числом разрядов в равномерной шкале. Увеличение числа разрядов повышает точность измерений и расширяет динамический диапазон измеряемых сигналов. Потерянная из-за недостатка разрядов АЦП информация невозможна, и существуют лишь оценки возникающей погрешности «округления» отсчетов, например, через мощность шума, порождаемого ошибкой в последнем разряде АЦП. Для этого используется понятие отношения «сигнал/шум» - отношение мощности сигнала к мощности шума (в децибелах). Наиболее часто применяются 8-, 10-, 12-, 16-, 20- и 24-х разрядные АЦП. Каждый дополнительный разряд улучшает отношение сигнал/шум на 6 децибел. Однако увеличение количества разрядов снижает скорость дискретизации и увеличивает стоимость аппаратуры. Важным аспектом является также динамический диапазон, определяемый максимальным и минимальным значением сигнала.

**Обработка цифровых сигналов** выполняется либо специальными процессорами, либо на универсальных ЭВМ и компьютерах по специальным программам. Наиболее просты для рассмотрения линейные системы. Линейными называются системы, для которых имеет место суперпозиция (отклик на сумму входных сигналов равен сумме откликов на каждый сигнал в отдельности) и однородность или гомогенность (изменение амплитуды входного сигнала вызывает пропорциональное изменение выходного сигнала). Для реальных объектов свойства линейности могут выполняться приближенно и в определенном интервале входных сигналов.

Если входной сигнал  $x(t-t_0)$  порождает однозначный выходной сигнал  $y(t-t_0)$  при любом сдвиге  $t_0$ , то систему называют инвариантной во времени. Ее свойства можно исследовать в любые произвольные моменты времени. Для описания линейной системы вводится специальный входной сигнал - единичный импульс (импульсная функция). В силу свойства суперпозиции и однородности любой входной сигнал можно представить в виде суммы таких импульсов, подаваемых в разные моменты времени и умноженных на соответствующие коэффициенты. Выходной сигнал системы в этом случае представляет собой сумму откликов на эти импульсы. Отклик на единичный импульс (импульс с единичной амплитудой) называют импульсной характеристикой системы  $h(n)$ . Соответственно, отклик системы на произвольный входной сигнал  $s(k)$  можно выразить сверткой

$$g(k) = h(n) \otimes s(k-n).$$

Если  $h(n)=0$  при  $n<0$ , то систему называют каузальной (причинной). В такой

системе реакция на входной сигнал появляется только после поступления сигнала на ее вход. Некаузальные системы физически невозможно реализовать в реальном масштабе времени. Если требуется реализовать свертку сигналов с двусторонними операторами (при дифференцировании, преобразовании Гильберта, и т.п.), то это выполняется с задержкой (сдвигом) входного сигнала минимум на длину левосторонней части оператора свертки.

**Z-преобразование.** Для анализа дискретных сигналов и систем широко используется  $z$ -преобразование, которое является обобщением дискретного преобразования Фурье. Этим преобразованием произвольной непрерывной функции  $s(t)$ , равномерно дискретизированной и отображенной отсчетами  $s_k = s(k\Delta t)$ , ставится в соответствие степенной полином по  $z$  (или степенной полином по  $z^{-1} = 1/z$ ), последовательными коэффициентами которого являются отсчеты функции:

$$s_k = s(k\Delta t) \leftrightarrow TZ[s(k\Delta t)] = \sum_{k=-\infty}^{\infty} s_k z^k = S(z),$$

где  $z = \sigma + j\nu = r \cdot \exp(-j\varphi)$  - произвольная комплексная переменная. Это преобразование позволяет в дискретной математике использовать всю мощь дифференциального и интегрального исчисления, алгебры и прочих хорошо развитых разделов аналитической математики.

Дискретные системы обычно описывается линейными разностными уравнениями с постоянными коэффициентами:

$$y(k) = \sum b(n) x(k-n) - \sum a(m) y(k-m), \quad n=0, 1, \dots, N, \quad m=1, 2, \dots, M.$$

Этим уравнением устанавливается, что выходной сигнал  $y(k)$  системы в определенный момент  $k_i$  (например, в момент времени  $k_i\Delta t$ ) зависит от значений входного сигнала  $x(k)$  в данный ( $k_i$ ) и предыдущие моменты ( $k_i-n$ ) и значений сигнала  $y(k)$  в предыдущие моменты ( $k_i-m$ ).

$Z$ -преобразование этого уравнения, выраженное относительно передаточной функции системы

$$H(z) = Y(z)/X(z),$$

представляет собой рациональную функцию в виде отношения двух полиномов от  $z$ . Корни полинома в числителе называются нулями, а в знаменателе - полюсами функции  $H(z)$ . Значения нулей и полюсов позволяют определить свойства линейной системы. Так, если все полюсы  $X(z)$  по модулю больше единицы, то система является устойчивой (не пойдет “вразнос” ни при каких входных воздействиях). Нули функции  $Y(z)$  обращают в ноль  $H(z)$  и показывают, какие колебания вовсе не будут восприниматься системой (“антирезонанс”). Систему называют минимально-фазовой, если все полюсы и нули передаточной функции лежат вне единичной окружности  $|z|=1$  на комплексной  $z$ -плоскости. Попутно заметим, что применение  $z$ -преобразования с отрицательными степенями  $z^{-1}$  меняет положение полюсов и нулей относительно единичной окружности  $|z|=1$  (область вне окружности перемещается внутрь окружности, и наоборот).

**Природа сигналов.** По своей природе сигналы могут быть случайными или детерминированными.

К детерминированным относят сигналы, значения которых в любой момент времени или в произвольной точке пространства (а равно и в зависимости от любых других аргументов) являются априорно известными или могут быть определены (вычислены) по известной или предполагаемой функции, даже если мы не знаем ее явного вида.

Случайные сигналы непредсказуемы по своим значениям во времени или в пространстве. Для каждого конкретного отсчета случайного сигнала можно знать только вероятность того, что он примет какое-либо значение в определенной области возможных значений. Закон распределения случайных значений далеко не всегда известен. Одним из самых распространенных является нормальное распределение, плотность которого имеет вид симметричного колокола. Для его описания достаточно двух первых моментов

распределения случайных величин.

Наиболее простые характеристики законов распределения – среднее значение случайных величин (математическое ожидание) и дисперсия (математическое ожидание квадрата отклонения от среднего), характеризующая разброс значений случайных величин относительно среднего значения. Параметры динамики случайных сигналов во времени характеризуются функциями автокорреляции (количественная оценка взаимосвязи значений случайного сигнала на различных интервалах) или автоковариации (то же, при центрировании случайных сигналов). Аналогичной мерой взаимосвязи двух случайных процессов и степени их сходства по динамике развития является кросскорреляция или кроссковариация (взаимная корреляция или ковариация). Максимальное значение взаимной корреляции достигается при совпадении двух сигналов. При задержке одного из сигналов по отношению к другому положение максимума корреляционной функции дает возможность оценить величину этой задержки.

**Функциональные преобразования сигналов.** Одним из основных методов частотного анализа и обработки сигналов является преобразование Фурье. Различают понятия “преобразование Фурье” и “ряд Фурье”. Преобразование Фурье предполагает непрерывное распределение частот, ряд Фурье задается на дискретном наборе частот. Сигналы также могут быть заданы в наборе временных отсчетов или как непрерывная функция времени. Это дает четыре варианта преобразований – преобразование Фурье с непрерывным или с дискретным временем, и ряд Фурье с непрерывным временем или с дискретным временем. Наиболее практична с точки зрения цифровой обработки сигналов дискретизация и во временной, и в частотной области, но не следует забывать, что она является аппроксимацией непрерывного преобразования. Непрерывное преобразование Фурье позволяет точно представлять любые явления. Сигнал, представленный рядом Фурье, может быть только периодичен. Сигналы произвольной формы могут быть представлены рядом Фурье только приближенно, т.к. при этом предполагается периодическое повторение рассматриваемого интервала сигнала за пределами его задания. На стыках периодов при этом могут возникать разрывы и изломы сигнала, и возникать ошибки обработки, вызванные явлением Гиббса, для минимизации которых применяют определенные методы (весовые окна, продление интервалов задания сигналов, и т.п.).

При дискретизации и во временной, и в частотной области, обычно говорят о дискретном преобразовании Фурье (ДПФ):

$$S(n) = \sum_k s(k) \exp(-j2\pi kn/N),$$

где  $N$ - количество отсчетов сигнала. Применяется оно для вычисления спектров мощности, оценивания передаточных функций и импульсных откликов, быстрого вычисления сверток при фильтрации, расчете корреляции, расчете преобразований Гильберта, и т.п. Расчет ДПФ по приведенной формуле требует вычисления  $n$  коэффициентов, каждый из которых зависит от  $k$  элементов исходного отрезка, так что число операций не может быть меньше  $nk$ . Существует целое семейство алгоритмов, известное, как “Быстрое Преобразование Фурье” - БПФ, сокращающее число операций для вычисления коэффициентов до  $n \log(k)$ . “Быстрое” не следует трактовать, как “упрощенное” или “неточное”. При точной арифметике результаты расчетов ДПФ и по алгоритмам БПФ совпадают.

Известное применение находят и варианты преобразования Фурье: косинусное для четных и синусное для нечетных сигналов, а также преобразование Хартли, где базисными функциями являются суммы синусов и косинусов, что позволяет повысить производительность вычислений и избавиться от комплексной арифметики. Вместо косинусных и синусных функций используются также меандровые функции Уолша, принимающие значения только  $+1$  и  $-1$ . И, наконец, в последнее время в задачах спектрально-временного анализа нестационарных сигналов, изучения нестационарностей и локальных особенностей сигналов "под микроскопом", очистки от



шумов и сжатия сигналов начинают получать в качестве базисов разложения вейвлеты ("короткие волны"), локализованные как во временной, так и в частотной области.

Традиционные методы анализа данных предназначены, как правило, для линейных и стационарных сигналов и систем, и только в последние десятилетия начали активно развиваться методы анализа нелинейных, но стационарных и детерминированных систем, и линейных, но нестационарных данных. Между тем, большинство естественных материальных процессов, реальных физических систем и соответствующих этим процессам и системам данных в той или иной мере являются нелинейными и нестационарными, и при анализе данных используются определенные упрощения, особенно в отношении априорно устанавливаемого базиса разложения данных.

Необходимое условие корректного представления нелинейных и нестационарных данных заключается в том, чтобы иметь возможность формирования адаптивного базиса, функционально зависящего от содержания самих данных. Такой подход реализуется в методе преобразования Гильберта-Хуанга, хотя на данный момент без соответствующих достаточно строгих математических обоснований /54/. Хорошие результаты применения метода для решения многих практических задач позволяют надеяться, что за разработкой строгой теории метода дело не станет.

## 1.2. КЛЮЧЕВЫЕ ОПЕРАЦИИ ЦИФРОВОЙ ОБРАБОТКИ

Существуют многочисленные алгоритмы ЦОС как общего типа для сигналов в их классической временной форме (телекоммуникации, связь, телевидение и пр.), так и специализированные в самых различных отраслях науки и техники (геоинформатике, геологии и геофизике, медицине, биологии, военном деле, и пр.). Все эти алгоритмы, как правило – блочного типа, построенные на сколь угодно сложных комбинациях достаточно небольшого набора типовых цифровых операций, к основным из которых относятся свертка (конволюция), корреляция, фильтрация, функциональные преобразования, модуляция. Эти операции уже рассматривались в "Теории сигналов и систем". Ниже приводятся только ключевые позиции по этим операциям ("повторенье – мать ученья").

**Линейная свертка** – основная операция ЦОС, особенно в режиме реального времени. Для двух конечных причинных последовательностей  $h(n)$  и  $y(k)$  длиной соответственно  $N$  и  $K$  свертка определяется выражением:

$$s(k) = h(n) \otimes y(k) \equiv h(n) * y(k) = \sum_{n=0}^N h(n) y(k-n), \quad (1.2.1)$$

где:  $\otimes$  или  $*$  - символные обозначения операции свертки. Как правило, в системах обработки одна из последовательностей  $y(k)$  представляет собой обрабатываемые данные (сигнал на входе системы), вторая  $h(n)$  – оператор (импульсный отклик) системы, а функция  $s(k)$  – выходной сигнал системы. В компьютерных системах с памятью для входных данных оператор  $h(n)$  может быть двусторонним от  $-N_1$  до  $+N_2$ , например – симметричным  $h(-n) = h(n)$ , с соответствующим изменением пределов суммирования в (1.2.1), что позволяет получать выходные данные без сдвига относительно входных. При строго корректной свертке с обработкой всех отсчетов входных данных размер выходного массива равен  $K+N_1+N_2-1$ , и должны задаваться начальные условия по отсчетам  $y(k)$  для значений  $y(0-n)$  до  $n=N_2$ , и конечные для  $y(K+n)$  до  $n=N_1$ . Пример выполнения свертки приведен на рис. 1.2.1.

Преобразование свертки однозначно определяет выходной сигнал для установленных значений входного сигнала при известном импульсном отклике системы. Обратная задача деконволюции - определение функции  $y(k)$  по функциям  $s(k)$  и  $h(n)$ , имеет решение только при определенных условиях. Это объясняется тем, что свертка может существенно изменить частотный спектр сигнала  $s(k)$  и восстановление функции  $y(k)$  становится невозможным, если определенные частоты ее спектра в сигнале  $s(k)$  полностью утрачены.

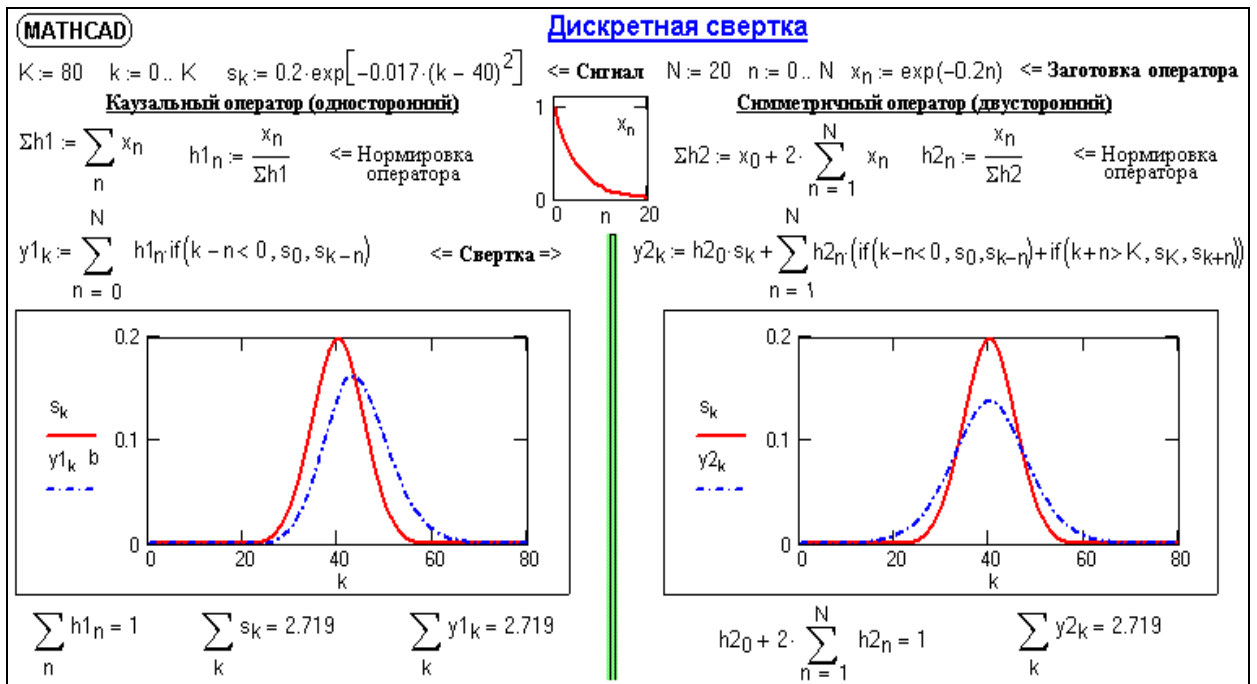


Рис. 1.2.1. Примеры дискретной свертки.

**Корреляция** существует в двух формах: автокорреляции и взаимной корреляции.

**Взаимно-корреляционная функция** (ВКФ, cross-correlation function - CCF), и ее частный случай для центрированных сигналов **функция взаимной ковариации** (ФВК) – это показатель степени сходства формы и свойств двух сигналов. Для двух последовательностей  $x(k)$  и  $y(k)$  длиной  $K$  с нулевыми средними значениями оценка взаимной ковариации выполняется по формулам:

$$K_{xy}(n) = (1/(K-n+1)) \sum_{k=0}^{K-n} x(k) y(k+n), \quad n = 0, 1, 2, \dots \quad (1.2.2)$$

$$K_{xy}(n) = (1/(K-n+1)) \sum_{k=0}^{K-n} x(k-n) y(k), \quad n = 0, -1, -2, \dots \quad (1.2.2')$$

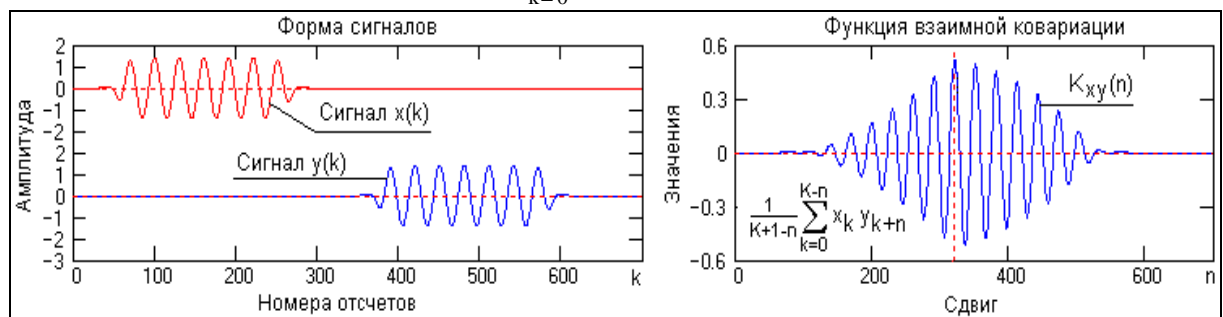


Рис. 1.2.2. Функция взаимной ковариации двух детерминированных сигналов.

Пример определения сдвига между двумя детерминированными сигналами, представленными радиоимпульсами, по максимуму ФВК приведен на рис. 1.2.2. По максимуму ФВК может определяться и сдвиг между сигналами, достаточно различными по форме.

На рис. 1.2.3 приведен аналогичный пример ФВК двух одинаковых по форме сигналов, на один из которых наложен шумовой сигнал. Мощность шума превышает мощность сигнала. Вычисление ФВК на рисунке выполнено в двух вариантах. Вариант 1 полностью соответствует формуле (1.2.2). Но в условиях присутствия в сигналах достаточно мощных шумов вычисление ФВК обычно выполняется по варианту 2 – с постоянным нормировочным множителем. Это определяется тем, что по мере увеличения сдвига  $n$  и уменьшения количества суммируемых членов в формуле (1.2.2) за счет

шумовых сигналов существенно нарастает ошибка оценки ФВК, которая к тому же увеличивается за счет нелинейного увеличения значения нормировочного множителя, особенно при малом количестве отсчетов. Сохранение множителя постоянным в какой-то мере компенсирует этот эффект.

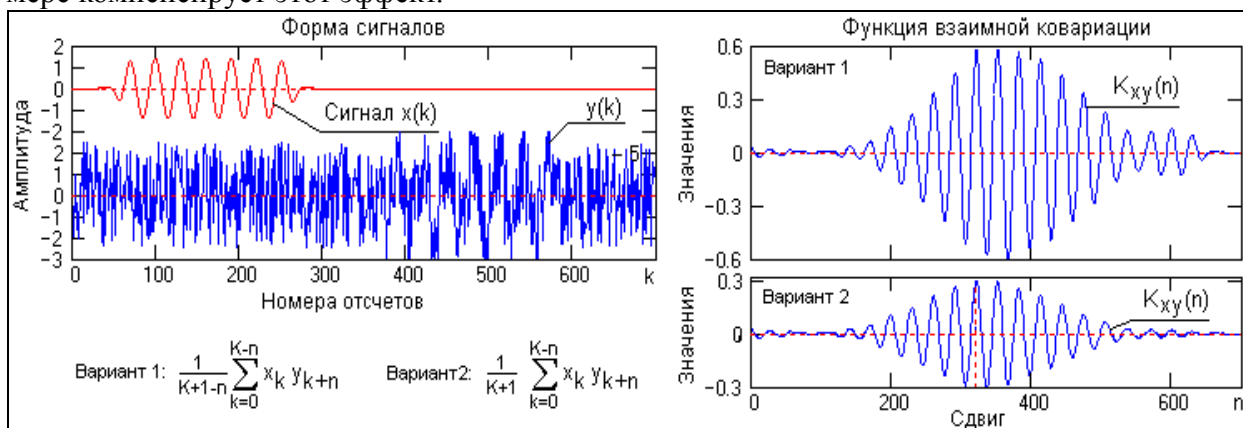


Рис. 1.2.3. ФВК двух сигналов, один из которых сильно зашумлен.

На рис. 1.2.4 приведен пример вычисления функции взаимной ковариации двух одинаковых сигналов, скрытых в шумах. ФВК позволяет не только определить величину сдвига между сигналами, но и уверенно оценить период колебаний в исследуемых радиоимпульсах.

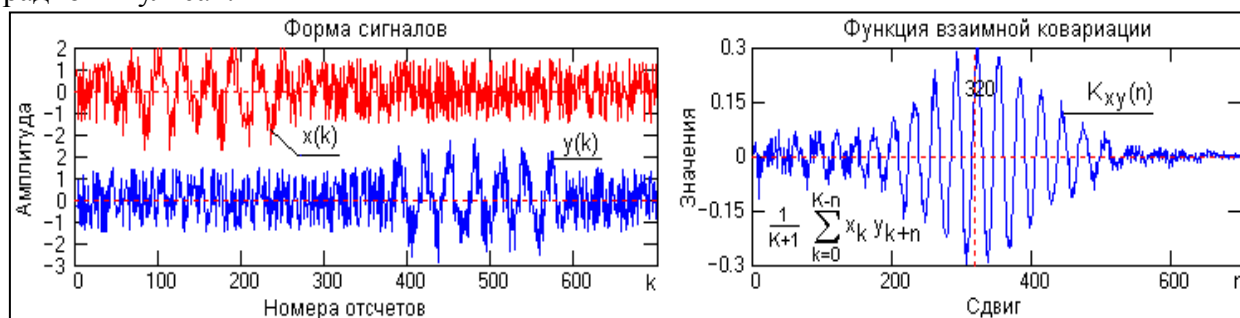


Рис. 1.2.4. ФВК двух зашумленных радиоимпульсов.

Относительный количественный показатель степени сходства двух сигналов  $x(k)$  и  $y(k)$  - функция взаимных корреляционных коэффициентов  $\rho_{xy}(n)$ . Она вычисляется через центрированные значения сигналов (для вычисления взаимной ковариации нецентрированных сигналов достаточно центрировать один из них), и нормируется на произведение значений стандартов (средних квадратических вариаций) функций  $x(k)$  и  $y(k)$ :

$$\rho_{xy}(n) = K_{xy}(n) / (\sigma_x \sigma_y). \quad (1.2.3)$$

$$\sigma_x^2 = K_{xx}(0) = (1/(K+1)) \sum_{k=0}^K (x(k))^2, \quad \sigma_y^2 = K_{yy}(0) = (1/(K+1)) \sum_{k=0}^K (y(k))^2. \quad (1.2.4)$$

Интервал изменения значений корреляционных коэффициентов при сдвигах  $n$  может изменяться от  $-1$  (полная обратная корреляция) до  $1$  (полное сходство или стопроцентная корреляция). При сдвигах  $n$ , на которых наблюдаются нулевые значения  $\rho_{xy}(n)$ , сигналы некоррелированы. Коэффициент взаимной корреляции позволяет устанавливать наличие определенной связи между сигналами вне зависимости от физических свойств сигналов и их величины.

Заметим, что в технической литературе в терминах "корреляция" и "ковариация" в настоящее время существуют накладки. Корреляционными функциями называют как функции по нецентрированным, так и по центрированным сигналам, а также и функцию взаимных корреляционных коэффициентов.

Автокорреляционная функция (АКФ, correlation function, CF) является количественной интегральной характеристикой формы сигнала, дает информацию о

структуре сигнала и его динамике во времени. Она, по существу, является частным случаем ВКФ для одного сигнала и представляет собой скалярное произведение сигнала и его копии в функциональной зависимости от переменной величины значения сдвига:

$$V_x(n) = (1/(K-n+1)) \sum_{k=0}^{K-n} x(k) x(k+n), \quad n = 0, 1, 2, \dots \quad (1.2.5)$$

АКФ имеет максимальное значение при  $n=0$  (умножение сигнала на самого себя), является четной функцией  $V_{xy}(-n)=V_{xy}(n)$ , и значения АКФ для отрицательных координат обычно не вычисляются. АКФ центрированного сигнала  $K_x(n)$  представляет собой функцию автоковариации (ФАК). ФАК, нормированная на свое значение  $K_x(0)=\sigma_x^2$  в  $n=0$ :

$$\rho_x(n) = K_x(n)/K_x(0) \quad (1.2.6)$$

называется функцией автокорреляционных коэффициентов.

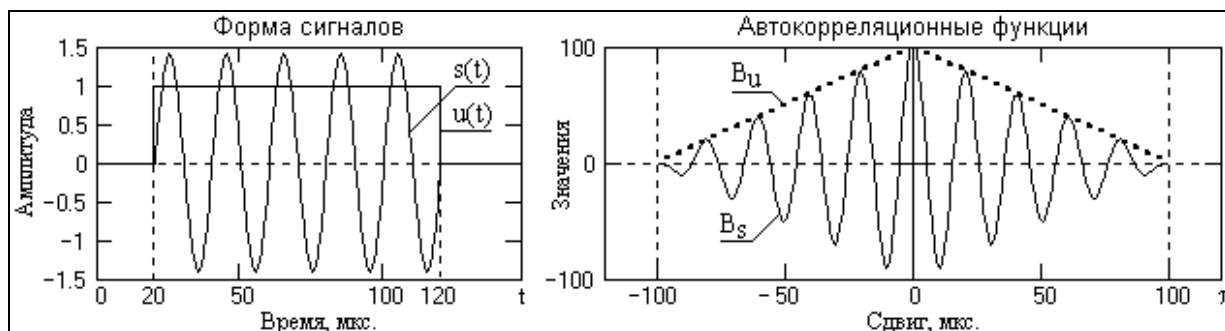


Рис. 1.2.5. Автокорреляционные функции.

В качестве примера на рис. 1.2.5 приведены два сигнала – прямоугольный импульс и радиоимпульс одинаковой длительности  $T$ , и соответствующие данным сигналам формы их АКФ. Амплитуда колебаний радиоимпульса установлена равной  $\sqrt{T}$  амплитуды прямоугольного импульса, при этом энергии сигналов будут одинаковыми, что подтверждается равными значениями максимумов АКФ. При конечной длительности импульсов длительности АКФ также конечны, и равны удвоенным значениям длительности импульсов (при сдвиге копии конечного импульса на интервал его длительности как влево, так и вправо, произведение импульса со своей копией становится равным нулю). Частота колебаний АКФ радиоимпульса равна частоте колебаний заполнения радиоимпульса (боковые минимумы и максимумы АКФ возникают каждый раз при последовательных сдвигах копии радиоимпульса на половину периода колебаний его заполнения).

**Линейная цифровая фильтрация** является одной из операций ЦОС, имеющих первостепенное значение, и определяется как

$$s(k) = \sum_{n=0}^N h(n) y(k-n), \quad (1.2.7)$$

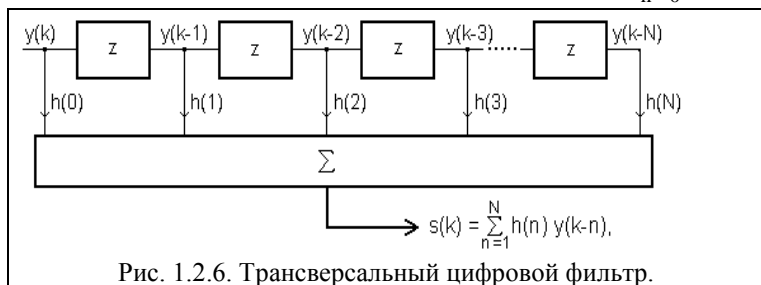


Рис. 1.2.6. Трансверсальный цифровой фильтр.

где:  $h(n)$ ,  $n=0, 1, 2, \dots, N$  – коэффициенты фильтра,  $y(k)$  и  $s(k)$  – вход и выход фильтра. Это по сути свертка сигнала с импульсной характеристикой фильтра.

На рис. 1.2.6 показана блок-схема фильтра, который в таком виде широко известен, как трансверсальный

( $z$  – задержка на один интервал дискретизации).

К основным операциям фильтрации информации относят операции сглаживания, прогнозирования, дифференцирования, интегрирования и разделения сигналов, а также выделение информационных (полезных) сигналов и подавление шумов (помех). Основными методами цифровой фильтрации данных являются частотная селекция

сигналов и оптимальная (адаптивная) фильтрация.

**Дискретные преобразования** позволяют описывать сигналы с дискретным временем в частотных координатах или переходить от описания во временной области к описанию в частотной. Переход от временных (пространственных) координат к частотным необходим во многих приложениях обработки данных.

Самым распространенным преобразованием является дискретное преобразование Фурье. При  $K$  отсчетов функции:

$$S(n) = \sum_{k=0}^{K-1} s(k) \exp(-j 2\pi kn/K). \quad (1.2.8)$$

Напомним, что дискретизация функции по времени приводит к периодизации ее спектра, а дискретизация спектра по частоте - к периодизации функции. Для дискретных преобразований  $s(k\Delta t) \Leftrightarrow S(n\Delta f)$ , и функция, и ее спектр дискретны и периодичны, а числовые массивы их представления соответствуют заданию на главных периодах  $T = K\Delta t$  (от 0 до  $T$  или от  $-T/2$  до  $T/2$ ), и  $2f_N = N\Delta f$  (от  $-f_N$  до  $f_N$ ), где  $K, N$  – количество отсчетов сигнала и его спектра соответственно. При этом:

$$\Delta f = 1/T = 1/(K\Delta t), \Delta t = 1/2f_N = 1/(N\Delta f), \Delta t\Delta f = 1/N, N = 2Tf_N = K. \quad (1.2.9)$$

Соотношения (1.2.9) являются условиями информационной равноценности динамической и частотной форм представления дискретных сигналов. Другими словами: для преобразований без потерь информации число отсчетов функции и ее спектра должны быть одинаковыми.

В принципе, согласно общей теории информации, последнее заключение действительно и для любых других видов линейных дискретных преобразований.

**Модуляция сигналов.** Системы регистрации, обработки, интерпретации, хранения и использования информационных данных становятся все более распределенными, что требует коммуникации данных по высокочастотным каналам связи. Как правило, информационные сигналы являются низкочастотными и ограниченными по ширине спектра, в отличие от широкополосных высокочастотных каналов связи, рассчитанных на передачу сигналов от множества источников одновременно с частотным разделением каналов. Перенос спектра сигналов из низкочастотной области в выделенную для их передачи область высоких частот выполняется операцией *модуляции*. При модуляции значения информационного (модулирующего) сигнала переносятся на определенный параметр высокочастотного (несущего) сигнала.

Самые распространенные схемы модуляции для передачи цифровой информации по широкополосным каналам – это амплитудная (amplitude shift keying – ASK), фазовая (phase shift keying – PSK) и частотная (frequency shift keying – FSK) манипуляции. При передаче данных по цифровым сетям используется также импульсно-кодовая модуляция (pulse code modulation – PCM).

### 1.3. ОБЛАСТИ ПРИМЕНЕНИЯ ЦИФРОВОЙ ОБРАБОТКИ

Нет смысла перечислять и давать оценку возможностей ЦОС в различных областях науки и техники. С весьма малой вероятностью можно попытаться найти отрасль, где ЦОС еще не получила широкого распространения. Поэтому коснемся только тех областей, где применение ЦОС развивается наиболее быстрыми темпами.

**Процессоры ЦОС.** Обработка данных в реальном времени обычно выполняется на специальных процессорах (чипах) ЦОС. Они, как правило, имеют:

- Встроенные умножители или умножители-накопители, работающие параллельно.
- Отдельные шины и области памяти для программ и данных.
- Команды организации циклов.
- Большие скорости обработки данных и тактовые частоты.
- Использование конвейерных методов обработки данных.

**Запись, воспроизведение, использование звука.**

*Цифровое микширование* – регулирование и смешивание многоканальных

аудиосигналов от различных источников. Это выполняется аудиоэквалайзерами (наборами цифровых полосовых фильтров с регулируемыми характеристиками), смесителями и устройствами создания специальных эффектов (реверберация, динамическое выравнивание и пр.).

*Синтезаторы речи* представляют собой достаточно сложные устройства генерации голосовых звуков. Микросхемы синтезаторов вместе с процессорами обычно содержат в ПЗУ словари слов и фраз в форме кадров (25 мс речи) с внешним управлением интонацией, акцентом и диалектом, что позволяет на высоком уровне имитировать человеческую речь.

*Распознавание речи* активно изучается и развивается, особенно для целей речевого ввода информации в компьютеры. Как правило, в режиме обучения выполняется их настройка на речь пользователя, в процессе которой система оцифровывает и создает в памяти эталоны слов. В режиме распознавания речь также оцифровывается и сравнивается с эталонами в памяти. Системы распознавания речи внедряются и в товары бытового назначения (набор телефонных номеров, включение/выключение телевизора, и пр.).

*Аудиосистемы воспроизведения компакт-дисков* при плотности записи выше  $10^6$  бит на  $\text{мм}^2$  обеспечивают очень высокую плотность хранения информации. Аналоговый звуковой сигнал в стереоканалах дискретизируется с частотой 44.1 кГц и оцифровывается 16-битным кодом. При записи на диск сигналы модулируются (EFM – преобразование 8-ми разрядного кода в 14-ти разрядный для надежности), при считывании сигналы демодулируются, исправляются ошибки (по возможности) и выполняется цифро-аналоговое преобразование.

#### **Применение ЦОС в телекоммуникациях.**

*Цифровая сотовая телефонная сеть* – двусторонняя телефонная система с мобильными телефонами через радиоканалы и связью через базовые радиостанции. Мировым стандартом цифровой мобильной связи является система GSM. Частотный диапазон связи 890-960 МГц, частотный интервал канала 200 кГц, скорость передачи информации 270 кбит/с. В мобильной связи ЦОС используется для кодирования речи, выравнивания сигналов после многолучевого распространения, измерения силы и качества сигналов, кодирования с исправлением ошибок, модуляции и демодуляции.

*Цифровое телевидение* дает потребителям интерактивность, большой выбор, лучшее качество изображения и звука, доступ в Интернет. ЦОС в цифровом телевидении играет ключевую роль в обработке сигналов, кодировании, модуляции/демодуляции видео- и аудиосигналов от точки захвата до момента появления на экране. ЦОС лежит в основе алгоритмов кодирования MPEG, которые используются для сжатия сигналов перед их передачей и при декодировании в приемниках.

*ЦОС в биомедицине.* Основное назначение – усиление сигналов, которые обычно не отличаются хорошим качеством, и/или извлечение из них информации, представляющей определенный интерес, на фоне существенного уровня шумов и многочисленных артефактов (ложных изображений как от внешних, так и от внутренних источников). Так, например, при снятии электрокардиограммы плода регистрируется электрическая активность сердца ребенка на поверхности тела матери, где также существует определенная электрическая активность. Применение ЦОС во многих областях медицины позволяет переходить от чисто качественных показателей к объективным количественным оценкам, как например, в анестезии к оценке глубины анестетического состояния пациента при операции по электрической активности мозга.

## **Тема 2. ЦИФРОВЫЕ ФИЛЬТРЫ ОБРАБОТКИ ОДНОМЕРНЫХ СИГНАЛОВ**

Задачей любого исследования является установление неизвестных свойств среды или отдельных конкретных объектов по данным наблюдения процессов, в них

происходящих. Изучаемые объекты могут оказаться труднодоступными или вовсе недоступными для непосредственного изучения методами прямого контакта. Например, о строении земных недр на глубинах более 10-15 км мы можем судить исключительно по данным сейсмических волн и по характеристикам гравитационного и магнитного полей Земли. По этой причине разработка методов математической обработки и интерпретации результатов наблюдений, установления взаимосвязи между физическими свойствами природных сред и происходящих в них процессов, имеет большое значение.

Естественным введением в широкую и фундаментальную область цифровой обработки информации является цифровая фильтрация данных. Под фильтрацией будем понимать любое преобразование информации (сигналов, результатов наблюдений), при котором во входной последовательности обрабатываемых данных целенаправленно изменяются определенные соотношения (динамические или частотные) между различными компонентами этих данных.

Преобразование сигналов осуществляется в системах. Системы, избирательно меняющие форму сигналов (амплитудно-частотную и/или фазово-частотную характеристику), подавление шумов, устранение помех, извлечение из сигналов определенной информации, разделение сигналов на составляющие, и т.п. называют фильтрами. Фильтры с любым целевым назначением являются частным случаем систем преобразования сигналов.

К основным операциям фильтрации информации относят операции сглаживания, прогнозирования, дифференцирования, интегрирования и разделения сигналов, а также выделение информационных (полезных) сигналов и подавление шумов (помех). Основными методами цифровой фильтрации данных являются частотная селекция сигналов и оптимальная фильтрация.

В настоящем курсе рассматриваются, в основном, методы линейной обработки сигналов линейными дискретными системами. Линейными называют системы, которые осуществляют преобразование линейных комбинаций входных сигналов в суперпозицию выходных сигналов. Принцип реализации линейных систем, физический - в виде специальных микропроцессорных устройств, или алгоритмический - в виде программ на ЭВМ, существенного значения не имеет и определяет только их потенциальные возможности.

В общем случае термином *Цифровой фильтр* (ЦФ) называют аппаратную или программную реализацию математического алгоритма, входом которого является цифровой сигнал, а выходом – другой цифровой сигнал с определенным образом модифицированной формой и/или амплитудной и фазовой характеристикой. Классификация цифровых фильтров обычно базируется на функциональных признаках алгоритмов цифровой фильтрации, согласно которому ЦФ подразделяются на 4 группы: фильтры частотной селекции, оптимальные (квазиоптимальные), адаптивные и эвристические. Наиболее изученными и опробованными на практике являются ЦФ частотной селекции.

## 2.1. ЦИФРОВЫЕ ФИЛЬТРЫ

**Общие понятия.** В одномерной дискретной линейной системе связь между входом и выходом (входной и выходной дискретными последовательностями значений сигнала – отсчетами), задается линейным оператором преобразования TL:

$$y(k\Delta t) = TL\{x(k\Delta t)\}.$$

Это выражение отображает краткую запись линейного разностного уравнения:

$$\sum_{m=0}^M a_m y(k\Delta t - m\Delta t) = \sum_{n=0}^N b_n x(k\Delta t - n\Delta t), \quad (2.1.1)$$

где  $k = 0, 1, 2, \dots$  - порядковый номер отсчетов,  $\Delta t$  - интервал дискретизации сигнала,  $a_m$  и  $b_n$  - вещественные или комплексные коэффициенты. Положим  $a_0 = 1$ , что всегда может

быть выполнено соответствующей нормировкой уравнения (2.1.1), и, принимая в дальнейшем  $\Delta t = 1$ , т.е. переходя к числовой нумерации цифровых последовательностей значений сигналов, приведем его к виду:

$$y(k) = \sum_{n=0}^N b_n x(k-n) - \sum_{m=1}^M a_m y(k-m). \quad (2.1.2)$$

При  $k < n$  и  $m$  проведение фильтрации возможно только при задании начальных условий для точек  $x(-k)$ ,  $k = 1, 2, \dots, N$ , и  $y(-k)$ ,  $k = 1, 2, \dots, M$ . Как правило, в качестве начальных условий принимаются либо нулевые значения, либо выполняется продление отсчетов входных сигналов или его тренда по отрицательным значениям аргумента.

Оператор, представленный правой частью данного уравнения, получил название цифрового фильтра, а выполняемая им операция - цифровой фильтрации данных (информации, сигналов). Если хотя бы один из коэффициентов  $a_m$  или  $b_n$  зависит от переменной  $k$ , то фильтр называется параметрическим, т.е. с переменными параметрами. Ниже мы будем рассматривать фильтры с постоянными коэффициентами (инвариантные по аргументу).

**Основные достоинства цифровых фильтров** по сравнению с аналоговыми.

- Цифровые фильтры могут иметь параметры, реализация которых невозможна в аналоговых фильтрах, например, линейную фазовую характеристику.
- ЦФ не требуют периодического контроля и калибровки, т.к. их работоспособность не зависит от дестабилизирующих факторов внешней среды, например, температуры.
- Один фильтр может обрабатывать несколько входных каналов или сигналов.
- Входные и выходные данные можно сохранять для последующего использования.
- Точность цифровых фильтров ограничена только разрядностью отсчетов.
- Фильтры могут использоваться при очень низких частотах и в большом диапазоне частот, для чего достаточно только изменять частоту дискретизации данных.

**Нерекурсивные фильтры.** При нулевых значениях коэффициентов  $a_m$  уравнение (2.1.2) переходит в уравнение линейной дискретной свертки функции  $x(k)$  с оператором  $b_n$ :

$$y(k) = \sum_{n=0}^N b_n x(k-n). \quad (2.1.3)$$

Значения выходных отсчетов свертки (2.1.3) для любого аргумента  $k$  определяются текущим и "прошлыми" значениями входных отсчетов. Такой фильтр называется нерекурсивным цифровым фильтром (НЦФ). Интервал суммирования по  $n$  получил название "окна" фильтра. Окно фильтра составляет  $N+1$  отсчет, фильтр является односторонним каузальным, т.е. причинно обусловленным текущими и "прошлыми" значениями входного сигнала, и выходной сигнал не может опережать входного. Каузальный фильтр может быть реализован физически в реальном масштабе времени.

При обработке данных на ЭВМ ограничение по каузальности снимается. В программном распоряжении фильтра могут находиться как "прошлые", так и "будущие" значения входной последовательности отсчетов относительно текущей точки вычислений  $k$ , при этом уравнение (2.1.3) будет иметь вид:

$$y(k) = \sum_{n=-N'}^N b_n x(k-n). \quad (2.1.4)$$

При  $N' = N$  фильтр называется двусторонним симметричным. Симметричные фильтры, в отличие от односторонних фильтров, не изменяют фазы обрабатываемого сигнала.

Так как реакция НЦФ на единичный входной импульс (а равно и на любой произвольный входной сигнал) всегда конечна и ограничена размером окна фильтра, такие фильтры называют также фильтрами с конечной импульсной характеристикой (КИХ-фильтры).

Техника выполнения фильтрации не отличается от техники выполнения обычной



дискретной свертки двух массивов данных.

Представим, что на одной полоске бумаги выписаны по порядку сверху вниз значения данных  $x(k) \equiv s_k$  (см. рис. 2.1.1). На второй полоске бумаги находятся записанные в обратном порядке значения коэффициентов фильтра  $b_n \equiv h_n$  (обозначение  $h$  для коэффициентов операторов НЦФ является общепринятым). Для вычисления  $y_k \equiv y(k)$  располагаем вторую полоску против первой таким образом, чтобы значение  $h_0$  совпало со значением  $s_k$ , перемножаем все значения  $h_n$  с расположенными против них значениями  $s_{k-n}$ , и суммируем все результаты перемножения. Результат суммирования является выходным значением сигнала  $y_k$ . Сдвигаем окно фильтра - полоску коэффициентов  $h_k$ , на один отсчет последовательности  $s_k$  вниз (или массив  $s_k$  сдвигаем на отсчет вверх) и вычисляем аналогично следующее значение выходного сигнала, и т.д.

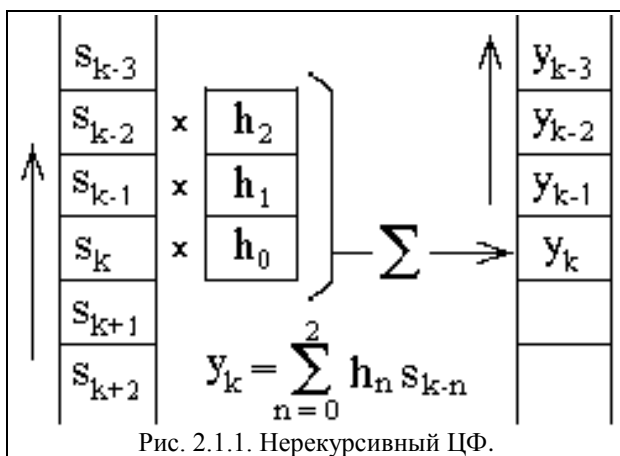


Рис. 2.1.1. Нерекурсивный ЦФ.

Описанный процесс является основной операцией цифровой фильтрации, и называется сверткой в вещественной области массива данных с оператором фильтра. Для математического описания наряду с формулами (2.1.3-2.1.4) применяются символические формы записи фильтрации:

$$y(k) = b(n) * x(k-n) \equiv b(n) \otimes x(k-n).$$

Сумма коэффициентов фильтра определяет коэффициент передачи (усиления) средних значений сигнала в окне фильтра и постоянной составляющей в

целом по массиву данных (с учетом начальных и конечных условий). Как правило, сумма коэффициентов фильтра нормируется к 1.

Имеется целый ряд методов обработки данных, достаточно давно и широко известных, которые по существу относятся к методам цифровой фильтрации, хотя и не называются таковыми. Например, методы сглаживания отсчетов в скользящем окне постоянной длительности. Так, для линейного сглаживания данных по пяти точкам с одинаковыми весовыми коэффициентами используется формула:

$$y_k = 0.2(x_{k-2} + x_{k-1} + x_k + x_{k+1} + x_{k+2}).$$

С позиций цифровой фильтрации это не что иное, как двусторонний симметричный нерекурсивный цифровой фильтр:

$$y_k = \sum_{n=-2}^2 b_n x_{k-n}, \quad b_n = 0,2. \quad (2.1.5)$$

Аналогично, при сглаживании данных методом наименьших квадратов (МНК) на основе кубического уравнения:

$$y_k = (-3x_{k-2} + 12x_{k-1} + 17x_k + 12x_{k+1} - 3x_{k+2})/35. \quad (2.1.6)$$

Это также НЦФ с коэффициентами:  $b_0 = 17/35$ ,  $b_1 = b_{-1} = 12/35$ ,  $b_2 = b_{-2} = -3/35$ .

---

**Пример.** Уравнение НЦФ:  $y_k = \sum_{n=-2}^2 b_n x_{k-n}$ ,  $b_n = 0,2$ . Начальные условия - нулевые.

Входной сигнал – скачок функции (ступень):  $x_k = \{0,0,0,0,0,10,10,10,10,\dots\}$ .

Выходной сигнал:  $y_k = \{0,0,0,0,2,4, 6, 8,10,10,10,\dots\}$ .

Результат фильтрации приведен на рис. 2.1.2(А). Проверьте результат (выполните фильтрацию, как это показано на рис. 2.1.1, с учетом четности фильтра).

---

Заметим: сумма коэффициентов сглаживающих НЦФ всегда должна быть равна 1, при этом сумма значений массива выходного сигнала равна сумме значений массива входного сигнала.

Координатная детальность выходного сигнала ниже входного, резкие изменения входных сигналов "размазываются" по аргументу.



Рис. 2.1.2. Сглаживание МНК в скользящем окне по пяти точкам

Повторите фильтрацию фильтром МНК на основе кубического уравнения. Сравните результаты фильтрации с результатами первого НЦФ (приведены на рис. 2.1.2(B)).

Для операции фильтрации характерны следующие основные свойства:

- Дистрибутивность:  $h(n) \otimes [a(k)+b(k)] = h(n) \otimes a(k)+h(n) \otimes b(k)$ .
- Коммутативность:  $h(n) \otimes a(k) \otimes b(k) = a(k) \otimes b(k) \otimes h(n)$ .
- Ассоциативность:  $[a(k) \otimes b(k)] \otimes h(n) = h(n) \otimes a(k) \otimes b(k)$ .

Фильтрация однозначно определяет выходной сигнал  $y(k)$  для установленного значения входного сигнала  $s(k)$  при известном значении импульсного отклика фильтра  $h(n)$ .

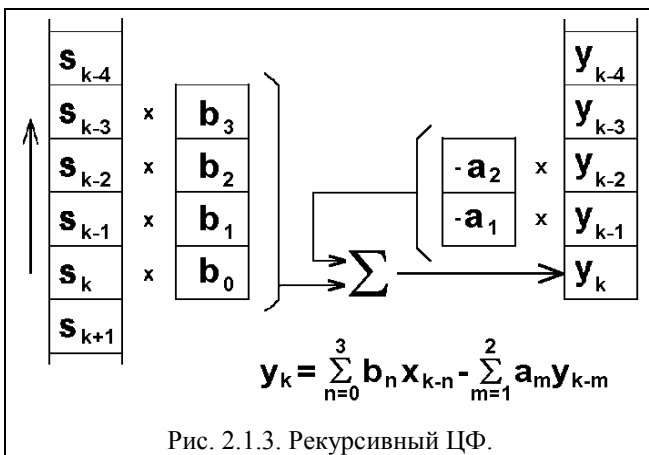


Рис. 2.1.3. Рекурсивный ЦФ.

**Рекурсивные фильтры.** Фильтры, которые описываются полным разностным уравнением (2.1.2)

$$y(k) = \sum_{n=0}^N b_n x(k-n) - \sum_{m=1}^M a_m y(k-m),$$

принято называть рекурсивными цифровыми фильтрами (РЦФ), так как в вычислении текущих выходных значений участвуют не только входные данные, но и значения выходных данных фильтрации, вычисленные в предшествующих циклах расчетов. С учетом последнего фактора рекурсивные фильтры называют также

фильтрами с обратной связью, положительной или отрицательной в зависимости от знака суммы коэффициентов  $a_m$ . Полное окно фильтра состоит из нерекурсивной части  $b_n$ , ограниченной в работе текущими и "прошлыми" значениями входного сигнала (на ЭВМ возможно использование и "будущих" отсчетов сигнала) и рекурсивной части  $a_m$ , которая работает с "прошлыми" значениями выходного сигнала. Техника вычислений приведена на рис. 2.1.3.

**Пример.** Уравнение РЦФ:  $y_k = b_0 x_k + a_1 y_{k-1}$ , при  $b_0 = a_1 = 0.5$ ,  $y_{-1} = 0$ .

Входной сигнал:  $x_k = \{0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, \dots\}$

Расчет выходного сигнала:

$$y_0 = 0,5x_0 + 0,5y_{-1} = 0; \quad y_1 = 0,5x_1 + 0,5y_0 = 0; \quad y_2 = 0,5x_2 + 0,5y_1 = 0,5; \quad y_3 = 0,5x_3 + 0,5y_2 = 0,25;$$

$$y_4 = 0,5x_4 + 0,5y_3 = 0,125; \quad y_5 = 0,5x_5 + 0,5y_4 = 0,0625; \quad y_6 = 0,5x_6 + 0,5y_5 = 0,03125; \quad \text{и т.д.}$$

Выходной сигнал:  $y_k = \{0, 0, 0, 5, 0, 25, 0, 125, 0, 0625, 0, 03125, 0, 015625, \dots\}$

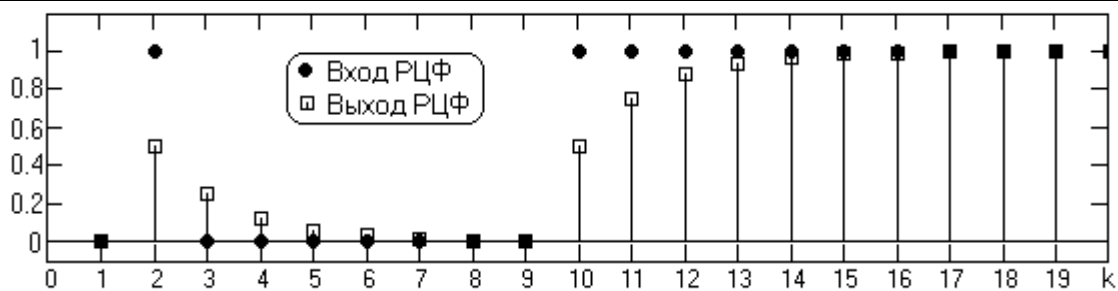


Рис. 2.1.4. Рекурсивная фильтрация.

Из примера рекурсивной фильтрации на рис. 2.1.4 можно видеть, что реакция РЦФ на входной сигнал (например, на единичный импульс Кронекера в точке 2), в результате действия обратной связи, в принципе, может иметь бесконечную длительность (в данном случае с близкими к нулю, но не нулевыми значениями), в отличие от реакции НЦФ, которая ограничена количеством членов  $b_k$  (окном фильтра). Фильтры такого типа называют фильтрами с бесконечной импульсной характеристикой (БИХ-фильтры). При положительной обратной связи (сумма коэффициентов  $a_m$  больше 1) фильтр становится неустойчивым (идет «в разнос» как на рис. 2.1.5)

**Пример.** Уравнение РЦФ:  $y_k = b_0 x_k - a_1 y_{k-1}$ , при  $b_0 = 0.5$ ,  $a_1 = 1.1$ ,  $y_{-1} = 0$   
 Входной сигнал:  $x_k = \{0, 10, 0, 0, 0, \dots\}$ .  
 Выходной сигнал:  $y_k = \{0, 0.5, -5.5, 6.05, -6.655, 7.321, -8.053, 8.858, -9.744, 10.718, -11.79, \dots \text{ и т.д.}\}$   
**Заметим:** коэффициент обратной связи больше  $a_1 > 1$  и выходной сигнал идет "в разнос".

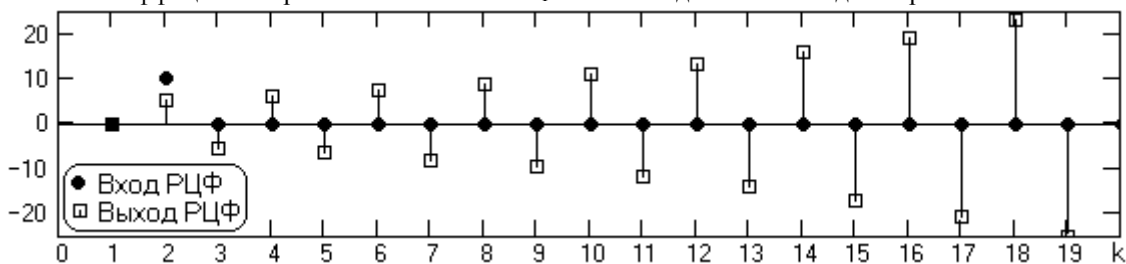


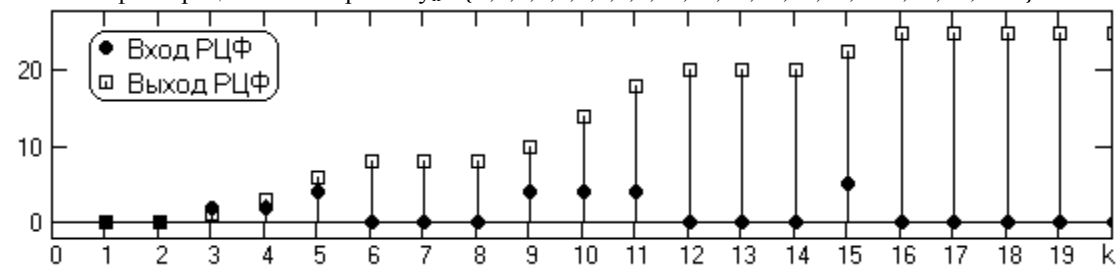
Рис. 2.1.5. Неустойчивый рекурсивный фильтр.

Операции, относящиеся к рекурсивной фильтрации, также известны в обычной практике, например - интегрирование. При интегрировании по формуле трапеций:

$$y_k = (x_k + x_{k-1})/2 + y_{k-1}, \quad (2.1.7)$$

т.е. здесь мы имеем РЦФ с коэффициентами:  $b_0 = b_1 = 0.5$ ,  $a_1 = 1$ .

**Пример.** Уравнение РЦФ:  $y_k = (x_k + x_{k-1})/2 + y_{k-1}$ , начальные условия - нулевые.  
 Входной сигнал:  $x_k = \{0, 0, 2, 2, 4, 0, 0, 0, 4, 4, 0, 0, 5, 0, 0, \dots\}$   
 Выполните фильтрацию. Контроль:  $y_k = \{0, 0, 0, 1, 3, 6, 8, 8, 10, 14, 18, 20, 20, 20, 22.5, 25, 25, \dots\}$



2.1.6. Интегрирующий рекурсивный фильтр.

## 2.2. ИМПУЛЬСНАЯ РЕАКЦИЯ ФИЛЬТРОВ.

**Функция отклика.** Если на вход нерекурсивного фильтра подать импульс Кронекера, расположенный в точке  $k = 0$ , то на выходе фильтра мы получим его реакцию на единичный входной сигнал (формула 2.1.3), которая определяется весовыми коэффициентами  $b_n$  оператора фильтра:

$$y(k) = \text{TL}[\delta(0)] = b_n \otimes \delta(k-n) = h(k) \equiv b_n. \quad (2.2.1)$$

Для рекурсивных фильтров реакция на импульс Кронекера зависит как от коэффициентов  $b_n$  фильтра, так и от коэффициентов обратной связи  $a_m$ . С использованием формулы (2.1.2):

$$y(k) = \sum_{n=0}^N b_n \delta(k-n) - \sum_{m=1}^M a_m y(k-m) = h_k. \quad (2.2.1')$$

Функция  $h(k)$ , которая связывает вход и выход фильтра по реакции на единичный входной сигнал и однозначно определяется оператором преобразования фильтра, получила название импульсного отклика фильтра (функции отклика). Для рекурсивных фильтров длина импульсного отклика, в принципе, может быть бесконечной.

Если произвольный сигнал на входе фильтра представить в виде линейной комбинации взвешенных импульсов Кронекера

$$x(k) = \sum_{n=-\infty}^{\infty} \delta(n) x(k-n),$$

то сигнал на выходе фильтра можно рассматривать как суперпозицию запаздывающих импульсных реакций на входную последовательность взвешенных импульсов:

$$y(k) = \sum_n h(n) (\delta(n) x(k-n)) \equiv \sum_n h(n) x(k-n).$$

Для нерекурсивных фильтров пределы суммирования в последнем выражении устанавливаются непосредственно по длине импульсного отклика  $h(n)$ .

**Определение импульсной реакции** на практике требуется, как правило, только для рекурсивных фильтров, так как импульсная реакция для НЦФ при известных значениях коэффициентов  $b(n)$ , как это следует из выражения (2.2.1), специального определения не требует:  $h(n) \equiv b(n)$ .

Если выражение для системы известно в общей форме (2.1.2), определение импульсной реакции производится подстановкой в уравнение системы импульса Кронекера с координатой  $k = 0$  при нулевых начальных условиях. В соответствии с выражением (2.2.1) сигнал на выходе системы будет представлять собой импульсную реакцию системы.

---

**Пример.** Уравнение РЦФ:  $y_k = x_k + 0.5y_{k-1}$ .  
 Входной сигнал:  $x_k = \delta_0 = \{1, 0, 0, 0, \dots\}$ .  
 Расчет выходного сигнала при нулевых начальных условиях:  
 $y_0 = x_0 + 0.5 y_{-1} = 1 + 0 = 1 = h_0$ .  $y_1 = x_1 + 0.5 y_0 = 0 + 0.5 = 0.5 = h_1$ .  $y_2 = x_2 + 0.5 y_1 = 0 + 0.25 = 0.25 = h_2$ .  
 $y_3 = x_3 + 0.5 y_2 = 0.125 = h_3$ .  $y_4 = x_4 + 0.5 y_3 = 0.0625 = h_4$ , и т.д.  
 Импульсный отклик фильтра:  $h_k = (0.5)^k$ ,  $k = 0, 1, 2, \dots$

---

Определение импульсной реакции физической системы обычно производится подачей на вход системы ступенчатой функции Хевисайда, которая равна  $u(k) = 1$  при  $k \geq 0$ , и  $u(k) = 0$  при  $k < 0$ :

$$g(k) = \sum_{n=0}^N h(n) u(k-n) = \sum_{n=0}^k h(n).$$

Отсюда:

$$h(k) = g(k) - g(k-1).$$

Функция  $g(k)$  получила название переходной характеристики системы (из одного статического состояния в другое). Форму реакции фильтра на функцию Хевисайда можно видеть на рис. 2.1.4 (с точки  $k = 10$  и далее) в сопоставлении с реакцией на импульс Кронекера в точке  $k=2$ .

### 2.3. ПЕРЕДАТОЧНЫЕ ФУНКЦИИ ФИЛЬТРОВ

**Z-преобразование.** Удобным методом решения разностных уравнений линейных систем является z-преобразование.

---

Напоминание.

$$y_n \leftrightarrow \sum_n y_n z^n = Y(z).$$

Аргумент  $z$  используется в двух вариантах:  $z^n$  или  $z^{-n}$ . В данном курсе по умолчанию используем аргумент  $z^n$ .

---

Применяя  $z$ -преобразование к обеим частям равенства (2.1.1), с учетом сдвига функций ( $y(k-m) \leftrightarrow z^m Y(z)$ ), получаем:

$$Y(z) \sum_{m=0}^M a_m z^m = X(z) \sum_{n=0}^N b_n z^n, \quad (2.3.1)$$

где  $X(z), Y(z)$ - соответствующие  $z$ -образы входного и выходного сигнала. Отсюда, полагая  $a_0 = 1$ , получаем в общей форме уравнение передаточной функции системы в  $z$ -области:

$$H(z) = Y(z)/X(z) = \sum_{n=0}^N b_n z^n / (1 + \sum_{m=1}^M a_m z^m). \quad (2.3.2)$$

Для НЦФ, при нулевых коэффициентах  $a_m$ :

$$H(z) = \sum_{n=0}^N b_n z^n. \quad (2.3.3)$$

При проектировании фильтров исходной, как правило, является частотная передаточная функция фильтра  $H(\omega)$ , по которой вычисляется ее  $Z$ -образ  $H(z)$  и обратным переходом в пространство сигналов определяется алгоритм обработки данных. В общей форме для выходных сигналов фильтра:

$$\begin{aligned} Y(z) &= H(z) \cdot X(z). \\ Y(z) \cdot (1 + \sum_{m=1}^M a_m z^m) &= X(z) \sum_{n=0}^N b_n z^n \\ Y(z) &= X(z) \sum_{n=0}^N b_n z^n - Y(z) \sum_{m=1}^M a_m z^m. \end{aligned} \quad (2.3.4)$$

После обратного  $Z$ -преобразования выражения (2.3.4):

$$y(k) = \sum_{n=0}^N b_n x(k-n) - \sum_{m=1}^M a_m y(k-m). \quad (2.3.5)$$

При подаче на вход фильтра импульса Кронекера  $\delta_0$ , имеющего  $z$ -образ  $\delta(z) = z^0 = 1$ , сигнал на выходе фильтра будет представлять собой импульсную реакцию фильтра  $y(k) \equiv h(k)$ , при этом:

$$H(z) = Y(z)/\delta(z) = Y(z) = TZ[y(k)] = \sum_{k=0}^{\infty} h(k) z^k, \quad (2.3.6)$$

т.е. передаточная функция фильтра является  $z$ -образом его импульсной реакции. При обратном  $z$ -преобразовании передаточной функции получаем импульсную характеристику фильтра:

$$h(k) \leftrightarrow H(z). \quad (2.3.7)$$

Если функция  $H(z)$  представлена конечным степенным полиномом, что характерно для НЦФ, являющихся КИХ-фильтрами, то обратное  $z$ -преобразование осуществляется элементарно - идентификацией коэффициентов по степеням  $z$ . Передаточная функция РЦФ также может быть представлена степенным полиномом прямым делением числителя на знаменатель правой части выражения (2.3.2), однако результат при этом может оказаться как конечным, так и бесконечным, т.е. система может иметь либо конечную, либо бесконечную импульсную характеристику. Практически используемые рекурсивные фильтры обычно имеют бесконечную импульсную характеристику (БИХ-фильтры) при конечном числе членов алгоритма фильтрации (2.3.5).

---

Примеры.

1. Передаточная функция РЦФ:  $H(z) = (1-z^5)/(1-z)$ .

---

Прямым делением числителя на знаменатель получаем:  $H(z) = 1+z+z^2+z^3+z^4$ .  
 $H(z) \Leftrightarrow h(n) = \{1,1,1,1,1\}$ . Фильтр РЦФ является КИХ-фильтром.  
 2. Передаточная функция:  $H(z) = 1/(1-2z)$ .  
 Методом обратного z-преобразования:  $h(n) = 2^n$ . Фильтр РЦФ является БИХ-фильтром.

**Устойчивость фильтров.** Фильтр называется устойчивым, если при любых начальных условиях реакция фильтра на любое ограниченное воздействие также ограничена. Критерием устойчивости фильтра является абсолютная сходимость отсчетов его импульсного отклика:

$$\sum_n |h(n)| < \infty. \quad (2.3.8)$$

Анализ устойчивости может быть проведен по передаточной функции. В устойчивой системе значение  $H(z)$  должно быть конечным во всех точках z-плоскости, где  $|z| \leq 1$ , а, следовательно, передаточная функция не должна иметь особых точек (полюсов) на и внутри единичного круга на z-плоскости. Полюсы  $H(z)$  определяются корнями знаменателя передаточной функции (2.3.2).

**Пример.**

Передаточная функция фильтра рис. 2.1.4:  $H(z) = b_0/(1-a_1z)$ . При  $a_1 = 0.5$  полюс знаменателя:  $z_p = 2$ .  $|z_p| > 1$ . Фильтр устойчив.

Передаточная функция фильтра рис. 2.1.5:  $H(z) = b_0/(1+a_1z)$ . При  $a_1 = 1.1$  полюс знаменателя:  $z_p = -0.909$ .  $|z_p| < 1$ . Фильтр неустойчив, что и подтверждает пример фильтрации.

Передаточная функция фильтра рис. 2.1.6:  $H(z) = 0.5(1+z)/(1-z)$ . Полюс знаменателя:  $z_p = 1$ . В принципе, фильтр неустойчив, но эта неустойчивость проявляется только при  $k = \infty$ . Импульсный отклик фильтра  $h(n) = \{0.5, 1, 1, 1, \dots\}$ , сумма которого равна  $\infty$  только при  $n = \infty$ , т.е. при интегрировании бесконечно больших массивов. При интегрировании конечных массивов результат всегда конечен.

Приведенный критерий устойчивости относится к несократимой дроби, т.к. в противном случае возможна компенсация полюса нулем передаточной функции, и следует проверить наличие однозначных нулей и полюсов.

Проверка на устойчивость требуется только для рекурсивных цифровых фильтров (систем с обратной связью), нерекурсивные системы всегда устойчивы.

**2.4. ЧАСТОТНЫЕ ХАРАКТЕРИСТИКИ ФИЛЬТРОВ**

**Общие понятия.** От z-образов передаточных функций подстановкой  $z = \exp(-j\omega\Delta t)$  в уравнение (2.3.2) можно перейти к Фурье-образам функций, т.е. к частотным спектрам сигналов и частотной характеристике фильтров, а точнее – к функциям их спектральных плотностей.

Можно применить и способ получения частотных характеристик непосредственно из разностного уравнения системы обработки данных. Так как цифровая фильтрация относится к числу линейных, то, принимая для сигнала на входе фильтра выражение  $x(k\Delta t) = B(\omega) \exp(j\omega k\Delta t)$ , мы вправе ожидать на выходе фильтра сигнал  $y(k\Delta t) = A(\omega) \exp(j\omega k\Delta t)$ . Подставляя эти выражения в разностное уравнение фильтра (2.1.1), получаем:

$$\sum_{m=0}^M a_m A(\omega) \exp(j\omega k\Delta t - j\omega m\Delta t) = \sum_{n=0}^N b_n B(\omega) \exp(j\omega k\Delta t - j\omega n\Delta t).$$

$$A(\omega) \exp(j\omega k\Delta t) \sum_{m=0}^M a_m \exp(-j\omega m\Delta t) = B(\omega) \exp(j\omega k\Delta t) \sum_{n=0}^N b_n \exp(-j\omega n\Delta t).$$

$$A(\omega) \sum_{m=0}^M a_m \exp(-j\omega m\Delta t) = B(\omega) \sum_{n=0}^N b_n \exp(-j\omega n\Delta t). \quad (2.4.1)$$

Передаточная частотная функция (частотная характеристика при  $a_0=1$ ):

$$H(\omega) = A(\omega)/B(\omega) = \sum_{n=0}^N b_n \exp(-j\omega n\Delta t) / [1 + \sum_{m=1}^M a_m \exp(-j\omega m\Delta t)]. \quad (2.4.2)$$

Полученная частотная характеристика повторяет функцию (2.3.2) при  $z = \exp(-j\omega\Delta t)$ , что и следовало ожидать. Аналогично z-преобразованию (2.3.7), частотная характеристика фильтра представляет собой Фурье-образ его импульсной реакции, и

наоборот. При  $\Delta t = 1$ :

$$H(\omega) = \sum_{n=-\infty}^{\infty} h(n) \exp(-j\omega n), \quad (2.4.3)$$

$$h(n) = (1/2\pi) \int_{-\pi}^{\pi} H(\omega) \exp(j\omega n) d\omega. \quad (2.4.4)$$

В общем случае  $H(\omega)$  является комплексной функцией, модуль которой  $R(\omega)$  называется амплитудно-частотной характеристикой (АЧХ), а аргумент  $\varphi(\omega)$  – фазово-частотной характеристикой (ФЧХ).

$$A(\omega) = |H(\omega)| = \sqrt{\text{Re}^2 H(\omega) + \text{Im}^2 H(\omega)}$$

$$\varphi(\omega) = \text{arctg}(-\text{Im } H(\omega)/\text{Re } H(\omega)).$$

Выбор знака фазового угла ориентирован на каузальные системы с отрицательным временным запаздыванием сигналов. Допустим, что система осуществляет только сдвиг сигнала  $x(t)$  вправо по временной оси, т.е.  $y(t) = x(t-\tau)$ . Для преобразования Фурье функции  $y(t)$  имеем:

$$\begin{aligned} Y(f) &= \int_{-\infty}^{\infty} y(t) \exp(-j2\pi ft) dt = \int_{-\infty}^{\infty} x(t-\tau) \exp(-j2\pi ft) dt = \\ &= \exp(-j2\pi f\tau) \int_{-\infty}^{\infty} x(t) \exp(-j2\pi ft) dt = \exp(-j2\pi f\tau) X(f). \end{aligned}$$

Отсюда:

$$H(f) = Y(f)/X(f) = \exp(-j2\pi f\tau), \quad |H(f)| = 1, \quad \varphi_h(f) = -2\pi f\tau.$$

Из последнего равенства следует, что фаза представляет собой прямую с отрицательным тангенсом угла наклона  $-2\pi f\tau$ . Соответственно, для всех каузальных фильтров, осуществляющих преобразование с определенной задержкой сигнала на выходе, при выполнении операции над частотными составляющими сигнала имеет место:

$$\begin{aligned} Y(f) &= H(f) X(f) = |H(f)| \exp(j\varphi_h(f)) |X(f)| \exp(j\varphi_x(f)) = |H(f)| |X(f)| \\ &\quad \exp\{j [\varphi_h(f) + \varphi_x(f)]\}, \\ |Y(f)| &= |H(f)| |X(f)|, \quad \varphi_y(f) = \varphi_h(f) + \varphi_x(f). \end{aligned}$$

С учетом отрицательного знака  $\varphi_h(f)$  фазовой характеристики каузальных фильтров это вызывает сдвиг в "минус" всех частотных составляющих сигнала и соответствующую задержку выходного сигнала относительно входного.

На рис. 2.4.1-2.4.3 приведены частотные характеристики фильтров (модули и аргументы спектральных плотностей), которые были рассмотрены выше в примерах и на рис. 2.1.4 – 2.1.6. Графики приведены в границах главных диапазонов спектров, и получены непосредственной подстановкой  $z = \exp(-j\omega\Delta t)$  при  $\Delta t = 1$  в уравнения передаточных функций  $H(z)$ .

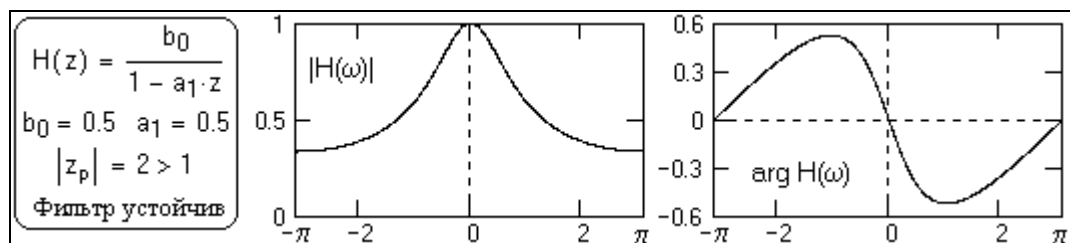


Рис. 2.4.1. Спектр не имеет особых точек.

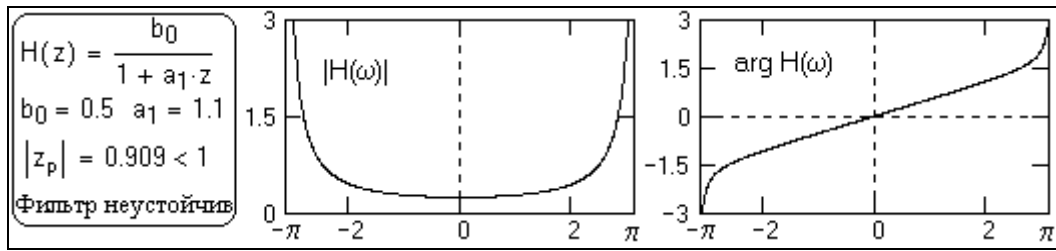


Рис. 2.4.2. Спектр имеет особые точки на границах диапазонов.

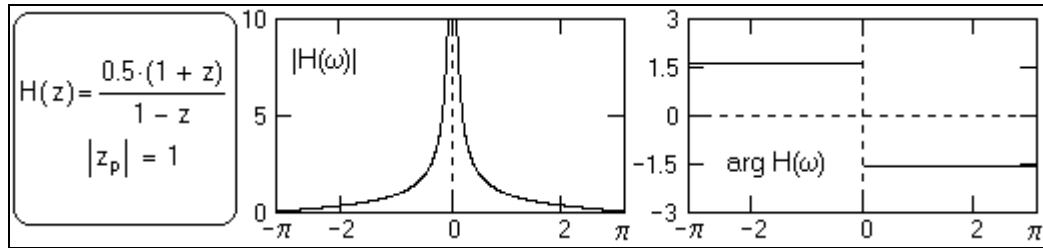


Рис. 2.4.3. Спектр интегрирующего фильтра. Особая точка на нулевой частоте.

При обработке ограниченных массивов амплитуда центрального пика равна количеству точек массива.

**Основные свойства** частотных характеристик цифровых фильтров:

1. Частотные характеристики являются непрерывными функциями частоты.
2. При дискретизации данных по интервалам  $\Delta t$  функция  $H(\omega)$  является периодической. Период функции  $H(\omega)$  равен частоте дискретизации входных данных  $F = 1/\Delta t$ . Первый низкочастотный период (по аргументу  $\omega$  от  $-\pi/\Delta t$  до  $\pi/\Delta t$ , по  $f$  от  $-1/2\Delta t$  до  $1/2\Delta t$ ) называется главным частотным диапазоном. Граничные частоты главного частотного диапазона соответствуют частоте Найквиста  $\pm\omega_N$ ,  $\omega_N = \pi/\Delta t$ . Частота Найквиста определяет предельную частоту данных, которую способен обрабатывать фильтр.

3. Для фильтров с вещественными коэффициентами импульсной реакции  $h(n\Delta t)$  функция АЧХ является четной, а функция ФЧХ - нечетной. С учетом этого частотные характеристики фильтров обычно задаются только на интервале положительных частот  $0 - \omega_N$  главного частотного диапазона. Значения функций на интервале отрицательных частот являются комплексно сопряженными со значениями на интервале положительных частот.

Как правило, при частотном анализе фильтров значение  $\Delta t$  интервала дискретизации принимают за 1, что соответственно определяет задание частотных характеристик на интервале  $(0, \pi)$  по частоте  $\omega$  или  $(0, 1/2)$  по  $f$ . При использовании быстрых преобразований Фурье (БПФ) вычисления спектров осуществляются в одностороннем варианте положительных частот в частотном интервале от 0 до  $2\pi$  (от 0 до 1 Гц), где комплексно сопряженная часть спектра главного диапазона (от  $-\pi$  до 0) занимает интервал от  $\pi$  до  $2\pi$  (для ускорения вычислений используется принцип периодичности дискретных спектров). Заметим, что при выполнении БПФ количество точек спектра равно количеству точек входной функции, а, следовательно, отсчет на частоте  $2\pi$ , комплексно сопряженный с отсчетом на частоте 0, отсутствует. При нумерации точек входной функции от 0 до  $N$  он принадлежит точке  $N+1$  - начальной точке следующего периода, при этом шаг по частоте равен  $2\pi/(N+1)$ . Современное программное обеспечение БПФ допускает любое количество точек входной функции, при этом для нечетного значения  $N$  частоте  $\pi$  соответствует отсчет на точке  $(N+1)/2$ , не имеющий сопряженного отсчета, а при четном значении  $N$  отсутствует отчет на частоте  $\pi$  (она располагается между отсчетами  $k=N/2$  и  $N/2 + 1$ ). Отсчетам с номерами  $k$  главного диапазона БПФ (за исключением точки  $k=0$ ) соответствуют комплексно сопряженные отсчеты  $N+1-k$  (за исключением точки  $k=(N+1)/2$  при нечетном  $N$ ).

**Фазовая и групповая задержка.** Задержка сигналов во времени относится к характерной особенности каузальных систем в целом, а, следовательно, рекурсивных и односторонних нерекурсивных фильтров.



Фазовая задержка, это прямая характеристика временной задержки фильтром гармонических колебаний. При подаче на вход фильтра гармоники  $\sin \omega t$ , сигнал на выходе каузального фильтра, без учета изменения его амплитуды, равен  $\sin(\omega t - \varphi)$ , при этом:

$$\sin(\omega t - \varphi) = \sin \omega(t - t_p), \quad \omega t - \varphi = \omega(t - t_p).$$

Отсюда, фазовая задержка  $t_p$  на частоте  $\omega$  равна:

$$t_p = \varphi / \omega. \quad (2.4.5')$$

При распространении (2.4.5) в целом на спектральную передаточную функцию фильтра получаем:

$$t_p(\omega) = \varphi(\omega) / \omega. \quad (2.4.5)$$

Постоянство значения  $t_p(\omega)$  в определенном частотном диапазоне обеспечивает для всех гармоник сигнала такое же соотношение их фазовых характеристик, какое было на входе системы, т.е. не изменяет формы сигнала, если его спектр полностью сосредоточен в этом частотном диапазоне, и значения АЧХ в этом диапазоне также имеют постоянное значение. Это условие является определяющим, например, для систем передачи данных, для сглаживающих и полосовых частотных фильтров.

Что касается каузальных фильтров, то они, как правило, имеют в рабочем диапазоне определенную зависимость значения  $t_p$  от частоты, которая характеризуется групповым временем задержки (ГВЗ). ГВЗ характеризует среднюю временную задержку составного сигнала.

Допустим, что сигнал на входе фильтра представляет собой сумму двух гармоник с близкими частотами:

$$s(t) = \cos \omega_1 t + \cos \omega_2 t.$$

Тождественная тригонометрическая запись:

$$s(t) = 2 \cos 0.5(\omega_1 + \omega_2)t \cdot \cos 0.5(\omega_1 - \omega_2)t.$$

Эта запись показывает, что сумму двух гармоник с частотами  $\omega_1$  и  $\omega_2$  можно рассматривать, как амплитудную модуляцию гармоники с частотой  $(\omega_1 + \omega_2)/2$  гармоникой с частотой  $(\omega_1 - \omega_2)/2$ .

При прохождении через фильтр каждая из гармоник  $\omega_1$  и  $\omega_2$  может получить различную задержку, при этом сигнал на выходе фильтра, без учета амплитудных изменений:

$$s(t) = \cos(\omega_1 t - \varphi_1) + \cos(\omega_2 t - \varphi_2).$$

Тождественная запись:

$$s(t) = 2 \cos[0.5((\omega_1 + \omega_2)t - (\varphi_1 + \varphi_2))] \cdot \cos[0.5((\omega_1 - \omega_2)t - (\varphi_1 - \varphi_2))].$$

Пульсацию колебаний выразим через групповую временную задержку  $t_g$ :

$$\cos[0.5((\omega_1 - \omega_2)t - (\varphi_1 - \varphi_2))] = \cos[0.5(\omega_1 - \omega_2) \cdot (t - t_g)].$$

Отсюда:

$$(\omega_1 - \omega_2) \cdot t_g = \varphi_1 - \varphi_2.$$

$$t_g = (\varphi_1 - \varphi_2) / (\omega_1 - \omega_2) = \Delta\varphi / \Delta\omega. \quad (2.4.6)$$

При распространении выражения на непрерывную частотную характеристику фильтра:

$$t_g(\omega) = d\varphi(\omega) / d\omega. \quad (2.4.7)$$

Для вычислений ГВЗ удобно использовать комплексный логарифм передаточной функции фильтра:

$$\text{Ln } H(\omega) = \ln |H(\omega)| + j \cdot \varphi(\omega), \quad \varphi(\omega) = \text{Im} [\text{Ln } H(\omega)].$$

$$t_g(\omega) = d\varphi / d\omega = \text{Im} \{d[\text{Ln}(H(\omega))] / d\omega\} = \text{Im} \{dH(\omega) / [H(\omega)d\omega]\}. \quad (2.4.8)$$

Приближение для дискретных спектральных функций:

$$t_g(k \cdot \Delta\omega) \approx (2/\Delta\omega) \text{Im} \{(H_{k+1} - H_k) / (H_{k+1} + H_k)\}. \quad (2.4.9)$$

Различают фильтры с линейной и нелинейной фазовой характеристикой.

В фильтрах с нелинейной фазовой характеристикой частотные компоненты сигнала задерживаются на величину, не пропорциональную частоте, и тем самым в выходном сигнале изменяется связь между его компонентами, что может быть недопустимо во многих случаях обработки сигналов (передача данных, обработка биосигналов, воспроизведение музыки и видео, и пр.).

Чтобы фильтр имел линейную фазовую характеристику необходимо и достаточно, если выполняется одно из следующих условий:

$$\varphi(\omega) = \alpha\omega, \quad (2.4.10)$$

$$\varphi(\omega) = \beta - \alpha\omega, \quad (2.4.11)$$

где  $\alpha$  и  $\beta$  - константы. Условие (2.4.10) обеспечивает постоянные значения групповой и фазовой задержки. Оно выполняется, если импульсная характеристика фильтра имеет положительную симметрию:

$$h(n) = h(N-n-1), \quad n = 0, 1, 2, \dots, (N-1)/2, \quad N - \text{нечетное}; \\ n = 0, 1, 2, \dots, (N/2)-1, \quad N - \text{четное}.$$

При этом фазовая характеристика является функцией длины фильтра:

$$\alpha = (N-1)/2.$$

**Пример.**

Импульсный отклик фильтра задан параметрами:  $N=7$ ,  $h(0)=h(6)$ ,  $h(1)=h(5)$ ,  $h(2)=h(4)$ ,  $h(3)$ .

Передаточная функция фильтра:  $H(z) = \sum_{k=0}^{N-1} h(k) z^k$ . Подставляем  $z=\exp(-j\omega\Delta t)$  при  $\Delta t = 1$  и

получаем частотную характеристику фильтра в главном диапазоне  $(-\pi, \pi)$ :

$$H(\omega) = h(0)+h(1)\exp(-j\omega)+h(2)\exp(-2j\omega)+h(3)\exp(-3j\omega)+h(4)\exp(-4j\omega)+h(5)\exp(-5j\omega)+h(6)\exp(-6j\omega) = \\ = \exp(-3j\omega) \{ h(0)[\exp(3j\omega)+\exp(-3j\omega)] + h(1)[\exp(2j\omega)+\exp(-2j\omega)] + h(2)[\exp(j\omega)+\exp(-j\omega)] + h(3) \} = \\ = \exp(-3j\omega) \{ 2h(0) \cos 3j\omega + 2h(1) \cos 2j\omega + 2h(2) \cos j\omega + h(3) \}.$$

Изменяя обозначения и переходя к индексации относительно центра симметрии  $a(0) = h(3)$ ,  $a(n) = 2h(3-n)$ ,  $n=1, 2, 3$ , записываем в компактной форме:

$$H(\omega) = \sum_{n=0}^3 a(n) \cos(nj\omega) \exp(-3j\omega) = |H(\omega)| \exp(j\varphi(\omega)), \quad \varphi(\omega) = -3\omega, \quad \alpha=3 \equiv (N-1)/2.$$

Частотная характеристика фильтра линейна.

Условие (2.4.11) обеспечивает постоянную групповую задержку и выполняется при отрицательной симметрии импульсной характеристики фильтра:

$$h(n) = -h(N-n-1), \\ \alpha = (N-1)/2, \quad \beta = \pi/2.$$

Для того чтобы убедиться в последнем, достаточно рассмотреть пример, аналогичный выше приведенному.

**Корреляция входа и выхода фильтров** может быть получена на основе следующих простых соображений.

Примем для входного сигнала  $x(t) \leftrightarrow X(f)$  и выходного сигнала  $y(t) \leftrightarrow Y(f)$  за основу выражение преобразования в частотной области

$$Y(f) = H(f) X(f). \quad (2.4.12)$$

Умножим обе части этого выражения на комплексно сопряженную функцию  $X^*(f)$  и найдем математические ожидания левой и правой части:

$$M\{X^*(f) Y(f)\} = M\{X^*(f) H(f) X(f)\} = H(f) M\{X^*(f) X(f)\}.$$

Но математические ожидания этих произведений спектров представляют собой спектры плотности мощности, и, при обратном преобразовании Фурье, зависимость взаимной корреляционной функции входного и выходного сигналов фильтра от корреляционной функции входного сигнала и функции импульсного отклика фильтра:

$$W_{xy} = H(f) W_x \leftrightarrow h(t) \otimes B_x(\tau) = B_{xy}(\tau).$$

Это выражение в спектральной области может использоваться для практического

определения частотных передаточных функций фильтров с неизвестной формой импульсных откликов.

Если математические ожидания взять от квадратов модулей левой и правой части исходного выражения (2.4.12), то в результате получим выражения:

$$W_y(f) = |H(f)|^2 W_x(f) \leftrightarrow B_h(\tau) \otimes B_x(\tau).$$

**Области применения НЦФ и РЦФ** обычно обуславливаются видом их передаточных функций.

В принципе, нерекурсивные цифровые фильтры универсальны и способны реализовать любые практические задачи обработки сигналов. Это и понятно, т.к. реакция РЦФ на импульс Кронекера представляет собой импульсный отклик НЦФ, а, следовательно, задачи, решаемые РЦФ, могут выполняться и НЦФ, но при условии отсутствия ограничений по размерам окна фильтра. В первую очередь это касается реализации БИХ-фильтров с незатухающим или слабо затухающим импульсным откликом, например, интегрирующих или фильтров рекурсивной деконволюции. Ограничение по размерам окна является скорее не теоретическим (бесконечных операторов НЦФ не требуется), а чисто практическим. Нет смысла применять НЦФ с огромными размерами операторов и тратить машинное время, если та же задача во много раз быстрее решается рекурсивным фильтром.

Существенным преимуществом НЦФ является их устойчивость, возможность выполнения в виде двусторонних симметричных фильтров, не изменяющих фазу выходных сигналов относительно входных, и реализации строго линейных фазовых характеристик.

С другой стороны, нерекурсивные фильтры могут быть преобразованы в рекурсивные фильтры, если есть возможность z-полином передаточной функции НЦФ выразить в виде отношения двух коротких z-полиномов РЦФ типа (2.3.2), что может дать существенное повышение производительности вычислений. Как правило, такая возможность имеется для сходящихся степенных рядов. Отношение двух z-полиномов позволяет реализовать короткие и очень эффективные фильтры с крутыми срезами на частотных характеристиках.

## 2.5. СТРУКТУРНЫЕ СХЕМЫ ЦИФРОВЫХ ФИЛЬТРОВ

**Структурные схемы.** Алгоритмы цифровой фильтрации сигналов (цифровых фильтров) представляются в виде структурных схем, базовые элементы которых показаны на рисунке 2.5.1 вместе с примерами структурных схем фильтров. Как правило, структурные схемы соответствуют программной реализации фильтров на ЭВМ, но не определяют аппаратной реализации в специальных радиотехнических устройствах, которая может отличаться от программной реализации.

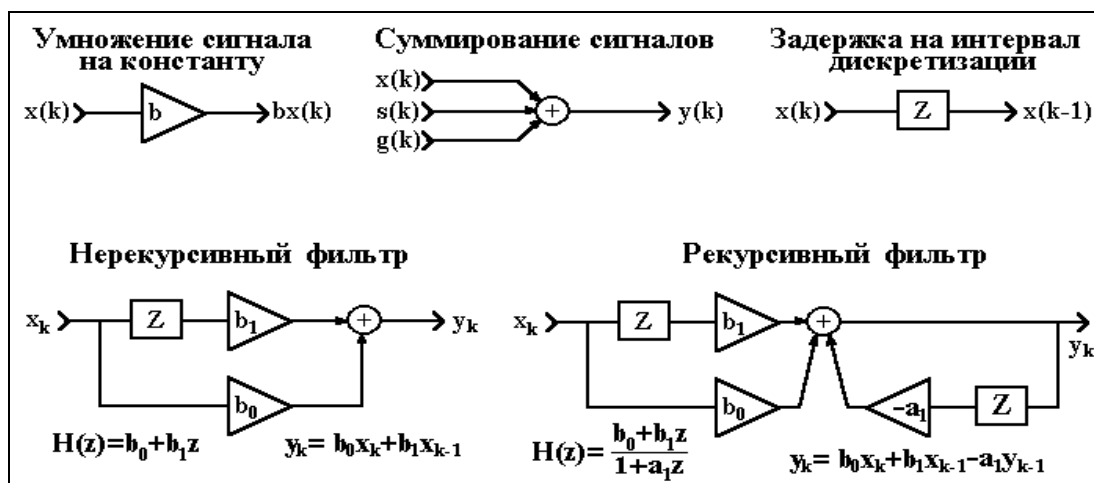
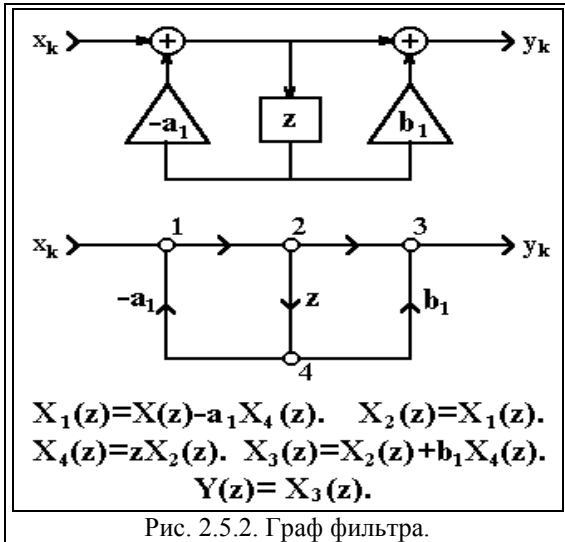
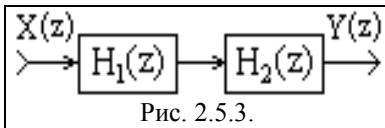


Рис. 2.5.1. Структурные схемы цифровых фильтров.



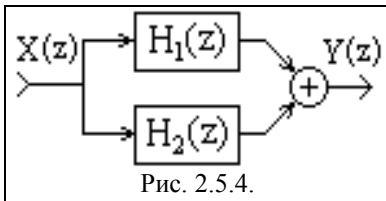
коэффициент.

**Соединения фильтров.** Различают следующие соединения фильтров.



1. **Последовательное соединение** (рис. 2.5.3). Выходной сигнал предшествующего фильтра является входным для последующего. Эквивалентная передаточная функция общей системы равна произведению передаточных функций фильтров, в нее входящих:

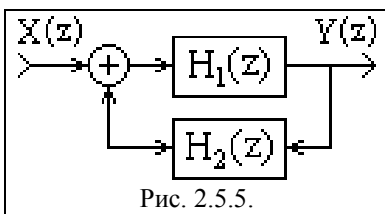
$$H(z) = H_1(z) \cdot H_2(z) \cdot \dots \cdot H_N(z).$$



2. **Параллельное соединение** (рис. 2.5.4). Сигнал подается на входы всех параллельно соединенных фильтров одновременно, выходные сигналы фильтров суммируются. Эквивалентная передаточная функция общей системы равна сумме передаточных функций фильтров, в нее входящих:

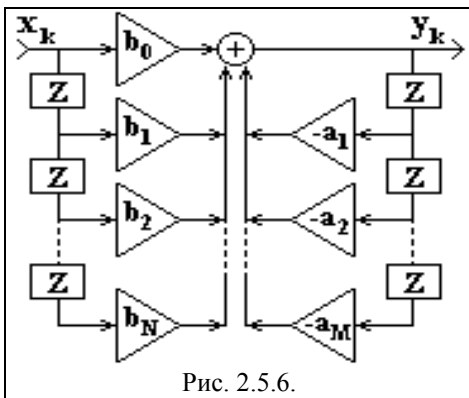
$$H(z) = H_1(z) + H_2(z) + \dots + H_N(z).$$

функций фильтров, в нее входящих:



3. **Соединение обратной связи** (рис. 2.5.5). Сигнал первого фильтра подается на выход системы и одновременно на вход фильтра обратной связи, выходной сигнал которого суммируется, со знаком плюс или минус в зависимости от вида связи (отрицательной или положительной), с входным сигналом системы. Эквивалентная передаточная функция системы:

$$H(z) = H_1(z) / (1 \pm H_1(z)H_2(z)).$$



минимальное число элементов

**Схемы реализации фильтров.** По принципам структурной реализации фильтров различают следующие схемы:

1. **Прямая форма** (рис. 2.5.6) реализуется непосредственно по разностному уравнению

$$y_k = \sum_{n=0}^N b_n x_{k-n} - \sum_{m=1}^M a_m y_{k-m},$$

или по передаточной функции

$$H(z) = \sum_{n=0}^N b_n z^n / (1 + \sum_{m=1}^M a_m z^m).$$

2. **Прямая каноническая форма** содержит задержки. Передаточную функцию РЦФ можно

представить в следующем виде:

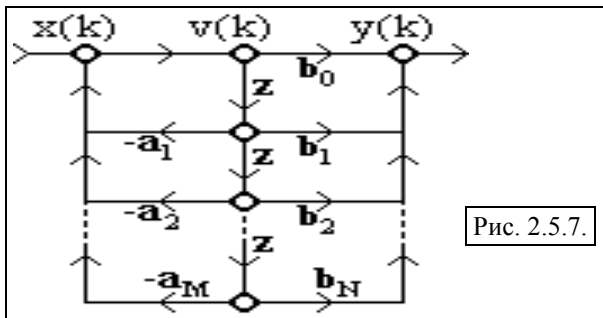
$$H(z) = Y(z)/X(z) = H_1(z)H_2(z),$$

$$H_1(z) = V(z)/X(z) = 1/(1 + \sum_{m=1}^M a_m z^m),$$

$$H_2(z) = Y(z)/V(z) = \sum_{n=0}^N b_n z^n.$$

Отсюда: 
$$v(k) = x(k) - \sum_{m=1}^M a_m v(k-m),$$
 (2.5.1)

$$y(k) = \sum_{n=0}^N b_n v(k-n). \quad (2.5.2)$$



В разностных уравнениях (2.5.1-2.5.2) осуществляется только задержка сигналов  $v(k)$ . Граф реализации РЦФ в прямой канонической форме приведен на рисунке 2.5.7.

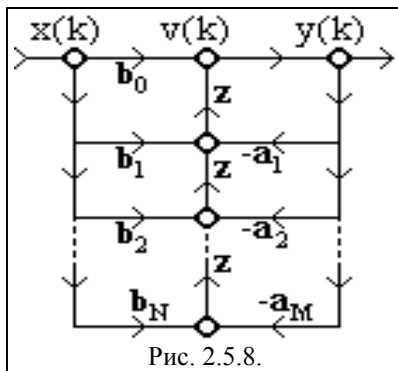
3. **Каскадная (последовательная) форма** соответствует представлению передаточной функции в виде произведения:

$$H(z) = \prod_{i=1}^k H_i(z).$$

$H_i(z)$  - составляющие функции вида  $(1-r_i z)/(1-p_i z)$  при представлении  $H(z)$  в факторизованной форме, где  $r_i$  и  $p_i$  - нули и полюсы функции  $H(z)$ . В качестве функций  $H_i(z)$  обычно используются передаточные функции биквадратных блоков - фильтров второго порядка:

$$H_i(z) = (b_{0i} + b_{1i} \cdot z + b_{2i} \cdot z^2) / (1 + a_{1i} \cdot z + a_{2i} \cdot z^2).$$

4. **Параллельная форма** используется много реже, и соответствует представлению передаточной функции в виде суммы биквадратных блоков или более простых функций.



**Обращенные формы.**

В теории линейных направленных сигнальных графов существуют процедуры преобразования исходных графов с сохранением передаточных функций. Одна из таких процедур - обращение (транспозиция) графов, которая выполняется путем изменения направления всех ветвей цепи, при этом вход и выход графа также меняются местами. Для ряда систем такая транспозиция позволяет реализовать более эффективные алгоритмы обработки данных. Пример обращения графа прямой канонической формы рекурсивной системы (с перестроением графа на привычное расположение входа с левой стороны) приведен на рис. 2.5.8.

расположение входа с левой стороны) приведен на рис. 2.5.8.

**Тема 3. ФИЛЬТРЫ СГЛАЖИВАНИЯ. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ.**

Основной инструмент цифровой фильтрации данных и проектирования цифровых фильтров – спектральный (частотный) анализ. Частотный анализ базируется на использовании периодических функций, в отличие от численных методов анализа и математической статистики, где предпочтение отдается полиномам. В качестве периодических используются гармонические функции синусов и косинусов. Спектральный состав сигналов – это тонкая внутренняя структура данных, которая практически скрыта в динамическом представлении данных даже для опытных обработчиков. Частотная характеристика цифрового фильтра – это его однозначный функциональный паспорт, полностью определяющий сущность преобразования входных данных.

Следует отметить, что хотя цель фильтрации сигналов состоит именно в направленном изменении частотного состава данных, которые несет сигнал, у начинающих специалистов существует определенное эмоциональное противодействие частотному подходу в анализе данных. Преодолеть это противодействие можно только одним путем – на опыте убедиться в эффективности частотного подхода.

Рассмотрим пример частотного анализа фильтров при сглаживании данных методом наименьших квадратов (МНК).

### 3.1. ФИЛЬТРЫ МНК 1-го ПОРЯДКА

Предположим, что требуется осуществить сглаживание (аппроксимацию) равномерного по аргументу массива данных методом наименьших квадратов (МНК).

**Расчет коэффициентов фильтра.** Простейший способ аппроксимации по МНК произвольной функции  $s(t)$  - с помощью полинома первой степени, т.е. функции вида  $y(t) = A+Bt$  (метод скользящих средних). Произведем расчет симметричного фильтра МНК на  $(2N+1)$  точек с окном от  $-N$  до  $N$ .

Для определения коэффициентов полинома найдем минимум функции остаточных ошибок приближения. С учетом дискретности данных по точкам  $t_n = n\Delta t$  и принимая  $\Delta t = 1$ , для симметричного НЦФ с нумерацией отсчетов по  $n$  от центра окна фильтра (в системе координат фильтра), функция остаточных ошибок записывается в форме:

$$\sigma(A, B) = \sum_n [s_n - (A+B \cdot n)]^2.$$

Дифференцируем функцию остаточных ошибок по аргументам  $A$ ,  $B$ , и, приравнявая полученные уравнения нулю, формируем 2 нормальных уравнения с двумя неизвестными:

$$\begin{aligned} \sum_{n=-N}^N (s_n - (A+B \cdot n)) &\equiv \sum_{n=-N}^N s_n - A \sum_{n=-N}^N 1 - B \sum_{n=-N}^N n = 0, \\ \sum_{n=-N}^N (s_n - (A+B \cdot n)) \cdot n &\equiv \sum_{n=-N}^N n \cdot s_n - A \sum_{n=-N}^N n - B \sum_{n=-N}^N n^2 = 0. \end{aligned}$$

С учетом равенства  $\sum_{n=-N}^N n = 0$ , решение данных уравнений относительно  $A$  и  $B$ :

$$A = \frac{1}{2N+1} \sum_{n=-N}^N s_n, \quad B = \sum_{n=-N}^N n \cdot s_n / \sum_{n=-N}^N n^2.$$

Подставляем значения коэффициентов в уравнение аппроксимирующего полинома, переходим в систему координат по точкам  $k$  массива  $y(k+\tau) = A+B \cdot \tau$ , где отсчет  $\tau$  производится от точки  $k$  массива, против которой находится точка  $n = 0$  фильтра, и получаем в общей форме уравнение фильтра аппроксимации:

$$y(k+\tau) = \frac{1}{2N+1} \sum_{n=-N}^N s_{k-n} + \tau \sum_{n=-N}^N n \cdot s_{k-n} / \sum_{n=-N}^N n^2.$$

Для сглаживающего НЦФ вычисления производятся непосредственно для точки  $k$  в центре окна фильтра ( $\tau = 0$ ), при этом:

$$y_k = \frac{1}{2N+1} \sum_{n=-N}^N s_{k-n}. \quad (3.1.1)$$

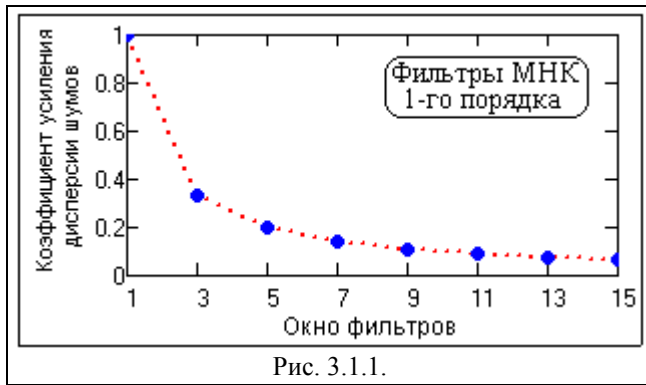


Рис. 3.1.1.

**Импульсная реакция фильтра**

соответственно определяется  $(2N+1)$  значениями коэффициентов  $b_n = 1/(2N+1)$ . Так, для 5-ти точечного НЦФ:

$$h(n) = \{0.2, 0.2, 0.2, 0.2, 0.2\}.$$

Передаточная функция фильтра в z-области:

$$H(z) = 0.2(z^{-2}+z^{-1}+1+z^1+z^2).$$

Коэффициент усиления дисперсии шумов:

$$K_q = \sum_n h^2(n) = 1/(2N+1),$$

т.е. обратно пропорционален ширине окна фильтра. Зависимость значения  $K_q$  от ширины окна приведена на рис. 3.1.1.

**Частотная характеристика фильтра** (передаточная функция фильтра в частотной области) находится преобразованием Фурье импульсной реакции  $h(n)$  (фильтр симметричный, начало координат в центре фильтра), или подстановкой  $z = \exp(-j\omega\Delta t)$  при  $\Delta t=1$  в выражение передаточной функции  $H(z)$ . И в том, и в другом случае получаем:

$$H(\omega) = 0.2[\exp(2j\omega)+\exp(j\omega)+1+\exp(-j\omega)+\exp(-2j\omega)]. \quad (3.1.2)$$

Можно использовать и непосредственно уравнение фильтра (3.1.1). Подадим на вход фильтра гармонический сигнал вида  $s_k = \exp(j\omega k)$ . Так как сигнальная функция относится к числу собственных, на выходе фильтра будем иметь сигнал  $u_k = H(\omega)\exp(j\omega k)$ . Подставляя выражения входного и выходного сигналов в уравнение (3.1.1), получаем:

$$H(\omega) \exp(j\omega k) = 0.2 \sum_{n=-2}^2 \exp(j\omega(k-n)) = 0.2 \exp(j\omega k) \sum_{n=-2}^2 \exp(-j\omega n).$$

Отсюда, выражение для передаточной функции:

$$H(\omega) = 0.2 \sum_{n=-2}^2 \exp(-j\omega n) = 0.2[\exp(2j\omega)+\exp(j\omega)+1+\exp(-j\omega)+\exp(-2j\omega)],$$

что полностью идентично выражению (3.1.2).

Так как импульсная реакция фильтра МНК симметрична (функция  $h(n)$  четная), частотное представление передаточной функции должно быть вещественным, в чем нетрудно убедиться, объединив комплексно сопряженные члены выражения (3.1.2):

$$H(\omega) = 0.2(1+2 \cos \omega+2 \cos 2\omega).$$

Альтернативное представление передаточной функции  $H(\omega)$  фильтра с произвольным количеством коэффициентов  $2N+1$  достаточно хорошо известно, как нормированный фурье-образ прямоугольной функции, каковой по существу и является селектирующее окно фильтра (3.1.1):

$$H(\omega) = \sin((N+1/2)\omega)/[(N+1/2)\omega] = \text{sinc}((N+1/2)\omega). \quad (3.1.3)$$

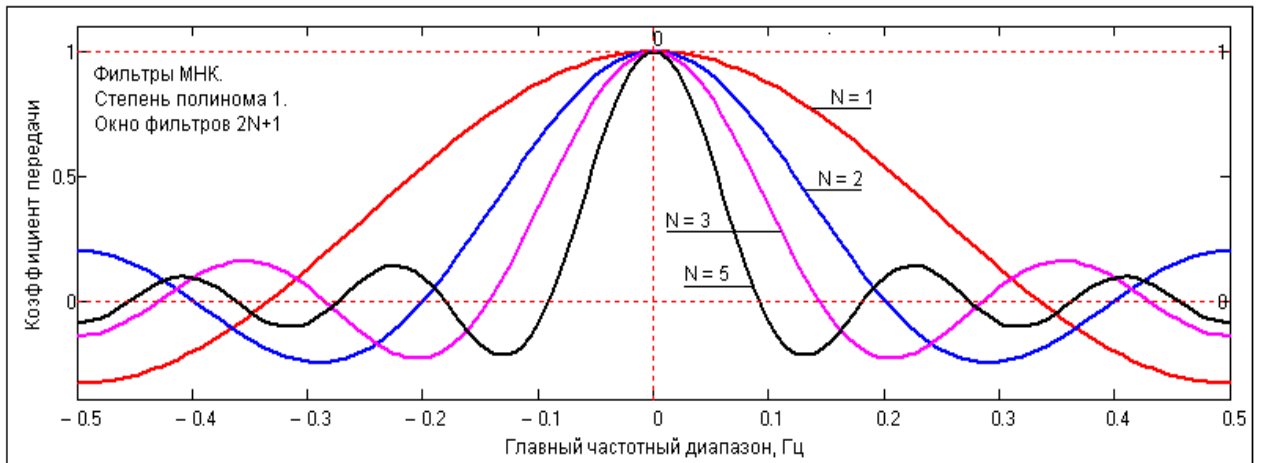


Рис. 3.1.2. Частотные характеристики фильтров МНК-1.

Графики передаточных функций (3.1.3) приведены на рисунке 3.1.2. По графикам можно видеть коэффициент передачи сигнала с входа на выход фильтра на любой частоте в главном частотном диапазоне. Без ослабления (с коэффициентом передачи 1) сглаживающим фильтром пропускается только сигнал постоянного уровня (нулевой частоты). Сумма коэффициентов сглаживающего НЦФ всегда должна быть равна 1 (отсчет дискретного фурье-преобразования на частоте  $\omega = 0$  равен сумме значений входной функции).

Чем больше число коэффициентов фильтра (шире окно фильтра), тем уже полоса пропускания низких частот. Подавление высоких частот довольно неравномерное, с осцилляциями передаточной функции относительно нуля. На рис. 3.1.3 приведен пример фильтрации случайного сигнала (шума) фильтрами с разным размером окна.

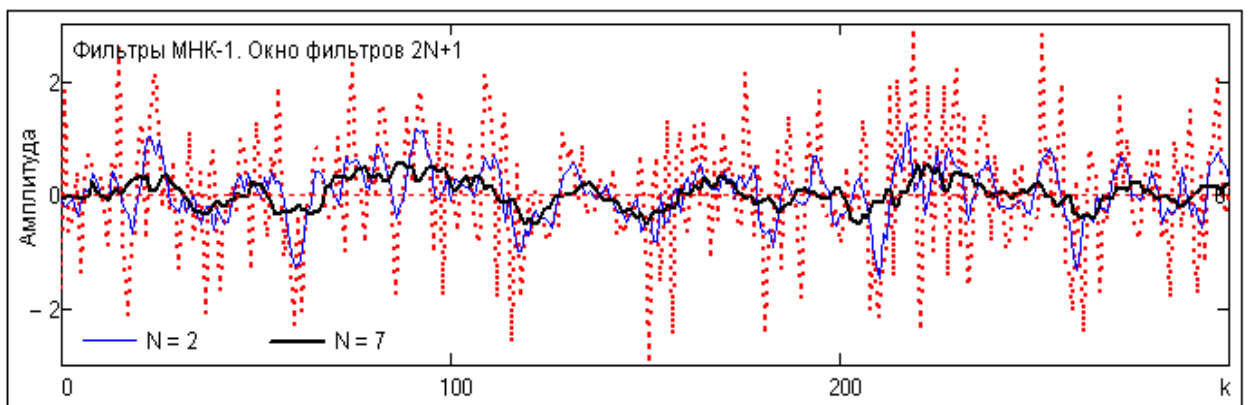


Рис. 3.1.3. Фильтрация шумов фильтрами МНК 1-го порядка.

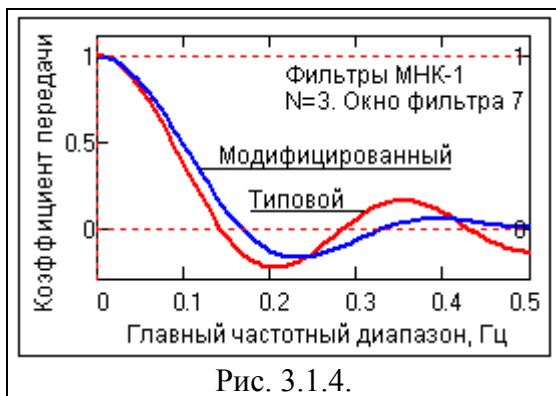


Рис. 3.1.4.

выражении (3.1.3) нормировку на  $2N+1$ , вычтем значение  $1/2$  крайних членов ( $\exp(-j\omega N) + \exp(j\omega N) / 2 = \cos \omega N$ ) и заново пронормируем полученное выражение к 1 (разделим на  $2N$ ). Пример новой передаточной функции при  $N=3$  приведен на рисунке 3.1.4.

### Модификация фильтра.

Частотное представление передаточных функций позволяет наглядно видеть особенности фильтров и целенаправленно улучшать их характеристики. Так, если в рассмотренном нами фильтре с однородной импульсной реакцией  $h_n = 1/(2N+1)$  уменьшить два крайних члена в 2 раза и заново нормировать к сумме  $\sum h_n = 1$ , то частотные характеристики фильтра заметно улучшаются. Для нахождения передаточной функции модифицированного фильтра снимем в



Передаточные функции модифицированных таким образом фильтров приводятся к нулю на частоте Найквиста, при этом несколько расширяется полоса пропускания низких частот и уменьшается амплитуда осцилляций в области подавления высоких частот. Если смотреть на сглаживание, как на операцию подавления высокочастотных помех, то модифицированные фильтры больше соответствует своему назначению.

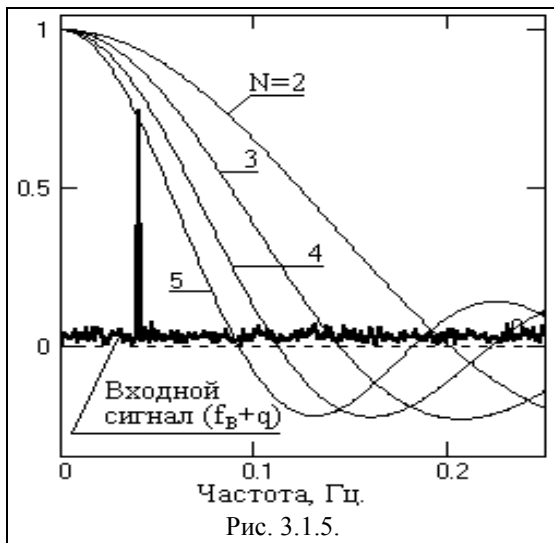


Рис. 3.1.5.

значение  $f_B$  равно 0.08 частоты Найквиста данных, т.е.  $f_B = 0.04$  Гц при  $\Delta t=1$ . Мощности гармоники и шума принимаем равными 1. Спектр модели сигнала плюс шума в сопоставлении с передаточными функциями фильтров приведен на рис. 3.1.5.

Таблица 3.1.1.

N	0	1	2	3	4	5	6	7
$K_y(f_B)$	1	0.98	0.94	0.88	0.8	0.7	0.6	0.51
$W_u(N)$	1	0.96	0.88	0.77	0.64	0.51	0.38	0.26
$W_q(N)$	1	0.33	0.2	0.14	0.11	0.09	0.08	0.07
$K_{с/ш}(N)$	1	2.88	4.4	5.4	5.8	5.6	4.89	3.85
$\delta^2(N)$	1	0.35	0.23	0.18	0.17	0.18	0.21	0.26
$\sigma^2(N)$	1	0.32	0.2	0.15	0.15	0.18	0.23	0.31

По формуле (3.1.3) вычисляем коэффициенты  $K_y(f_B)$  усиления фильтров с N от 0 до 7 на частоте  $f_B$  (см. таблицу 3.1.1). При мощности гармоники  $W_u = 1$  амплитудное значение гармоники на входе фильтра равно  $U = \sqrt{2W_u} = 1.41$ . Мощности гармоник на выходе фильтров в зависимости от N:

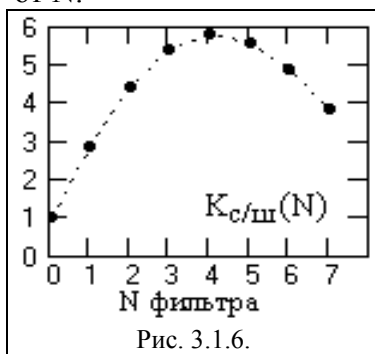


Рис. 3.1.6.

$$W_u(N) = 0.5 \cdot [U \cdot K_y(f_B)]^2$$

Соответственно, при мощности входного шума  $W_q=1$  мощности шумов на выходе фильтров будут численно равны коэффициентам усиления дисперсии шумов  $W_q(N) = W_q \cdot K_q(N)$ .

Максимум отношения

$$K_{с/ш}(N) = W_u(N) / W_q(N)$$

определяет оптимальный фильтр с максимальным увеличением отношения сигнал/шум, т.е., по существу, коэффициент увеличения отношения сигнал/шум при выполнении фильтрации с учетом изменения амплитудных значений полезной части сигнала на выходе фильтра.

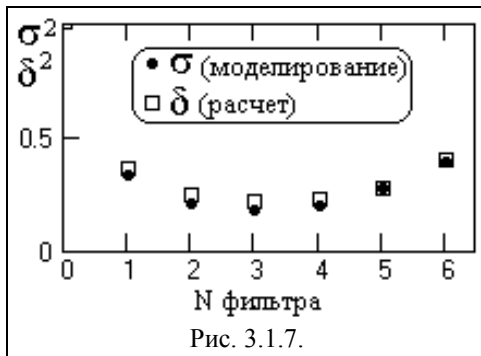


Рис. 3.1.7.

При  $K_y(f_b) > 0.5$  и  $W_u(N) = W_q(N) = 1$  численные значения величины  $\delta^2(N) = 1/K_{c/ш}(N)$  в первом приближении могут служить оценкой  $\sigma^2(N)$  квадрата среднего квадратического отклонения выходных сигналов от "чистой" гармонике  $f_b$ , заданной на входе. Свидетельством этому служат последние строки таблицы 3.1.1, где приведены результаты математического моделирования фильтрации по данным условиям на выборке 10000

точек. На рис. 3.1.7 приведены результаты сопоставления расчетных  $\delta^2(N)$  и модельных  $\sigma^2(N)$  значений данных коэффициентов.

Пример фильтрации зашумленного входного сигнала с полезной частотой  $f_b = 0.04$  Гц (синий пунктир) оптимальным фильтром МНК для данных условий ( $N=3$ ) приведен на рис. 3.1.8.

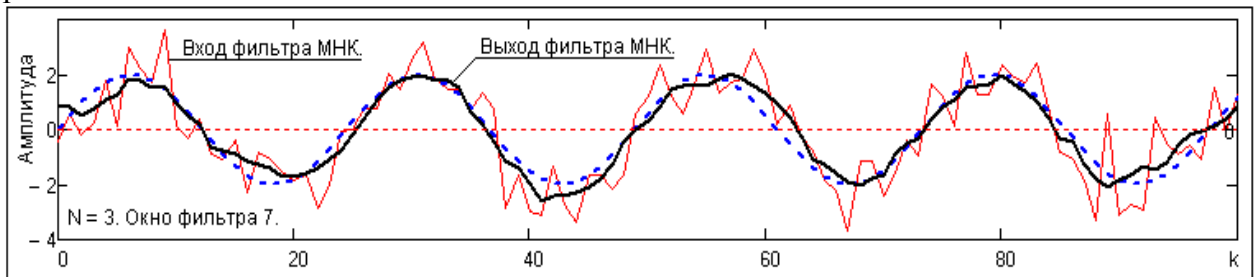


Рис. 3.1.8.

**Последовательная фильтрация.** Из фильтров МНК можно конструировать новые фильтры, частотные характеристики которых соответствуют последовательному применению «родительских» фильтров. Это выполняется последовательной  $n$ -кратной сверткой оператора исходного фильтра с самим собой, что дает эквивалент операторов  $n$ -кратной последовательной свертки данных с «родительским» оператором. Обычно применяется одно- и двукратная свертка, при этом окно нового фильтра расширяется, полоса пропускания по уровню 0.5 уменьшается (примерно на 25 и 40% соответственно), но резко уменьшается амплитуда пульсаций в зоне подавления (примерно в 4 и 16 раз соответственно). Пример конструирования новых фильтров из 7-ми точечного фильтра МНК-1 и их частотных характеристик приведен на рис. 3.1.9.

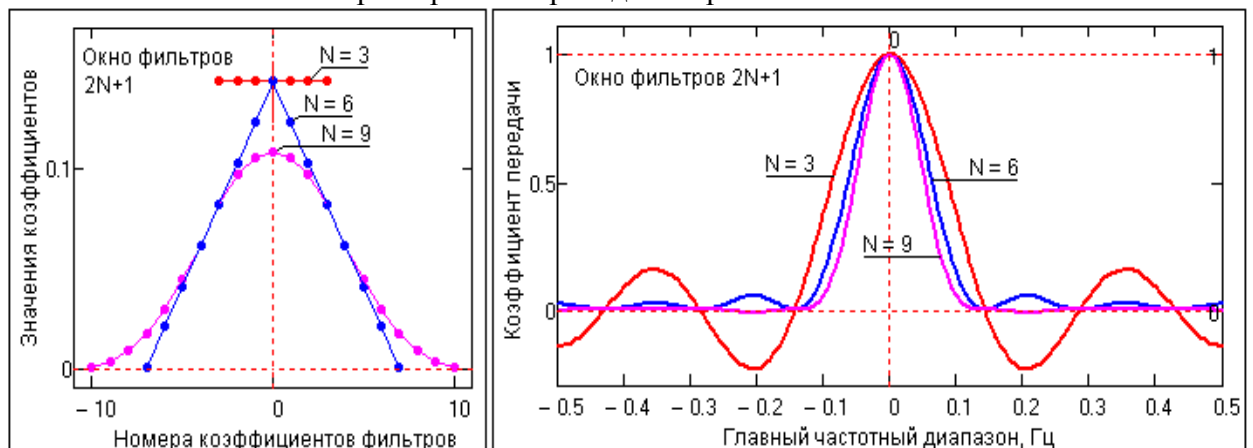


Рис. 3.1.9.

### 3.2. ФИЛЬТРЫ МНК 2-го ПОРЯДКА

**Расчет фильтров.** Фильтры МНК 2-го порядка (МНК-2) рассчитываются и анализируются аналогично. Рассмотрим квадратный многочлен вида  $y(t)=A+B \cdot t+C \cdot t^2$ . Для упрощения примера ограничимся симметричным сглаживающим НЦФ при  $\Delta t=1$ .

Уравнение суммы квадратов остаточных ошибок:

$$\sigma(A, B, C) = \sum_{n=-N}^N [s_n - (A + B \cdot n + C \cdot n^2)]^2. \quad (3.2.1)$$

Система уравнений после дифференцирования выражения (3.2.1) по A, B, C и приравнивания полученных выражений нулю:

$$A \sum_{n=-N}^N 1 + B \sum_{n=-N}^N n + C \sum_{n=-N}^N n^2 = \sum_{n=-N}^N s_n.$$

$$A \sum_{n=-N}^N n + B \sum_{n=-N}^N n^2 + C \sum_{n=-N}^N n^3 = \sum_{n=-N}^N n \cdot s_n.$$

$$A \sum_{n=-N}^N n^2 + B \sum_{n=-N}^N n^3 + C \sum_{n=-N}^N n^4 = \sum_{n=-N}^N n^2 \cdot s_n.$$

При вычислении значения квадратного многочлена только для центральной точки ( $t=0$ ) необходимости в значениях коэффициентов B и C не имеется. Решая систему уравнений относительно A, получаем:

$$A = \left\{ \sum_{n=-N}^N n^4 \sum_{n=-N}^N s_n - \sum_{n=-N}^N n^2 \sum_{n=-N}^N n^2 s_n \right\} / \left\{ \sum_{n=-N}^N 1 \sum_{n=-N}^N n^4 - \left[ \sum_{n=-N}^N n^2 \right]^2 \right\}. \quad (3.2.2)$$

При разворачивании выражения (3.2.2) для 5-ти точечного НЦФ:

$$y_0 = (17 \sum_{n=-2}^2 s_n - 5 \sum_{n=-2}^2 n^2 s_n) / 35 = (-3 \cdot s_{-2} + 12 \cdot s_{-1} + 17 \cdot s_0 + 12 \cdot s_1 - 3 \cdot s_2) / 35. \quad (3.2.3)$$

Импульсная реакция:  $h_n = \{(-3, 12, 17, 12, -3)/35\}$ .

Передаточная функция фильтра:

$$H(z) = (-3z^{-2} + 12z^{-1} + 17 + 12z^1 - 3z^2) / 35. \quad (3.2.4)$$

Аналогичным образом выражение (3.2.2) позволяет получить импульсную реакцию для 7, 9, 11 и т.д. точек фильтра:

$${}^3h_n = \{-2, 3, 6, 7, 6, 3, -2\} / 21.$$

$${}^4h_n = \{-21, 14, 39, 54, 59, 54, 39, 14, -21\} / 231.$$

$${}^5h_n = \{-36, 9, 44, 69, 84, 89, 84, 69, 44, 9, -36\} / 459.$$

**Частотные характеристики фильтров.** Подставляя значение  $z = \exp(-j\omega)$  в (3.2.4) или сигнал  $s_n = \exp(j\omega n)$  в (3.2.3) и объединяя комплексно сопряженные члены, получаем частотную характеристику 5-ти точечного сглаживающего фильтра МНК второго порядка:

$$H(\omega) = (17 + 24 \cos \omega - 6 \cos 2\omega) / 35.$$

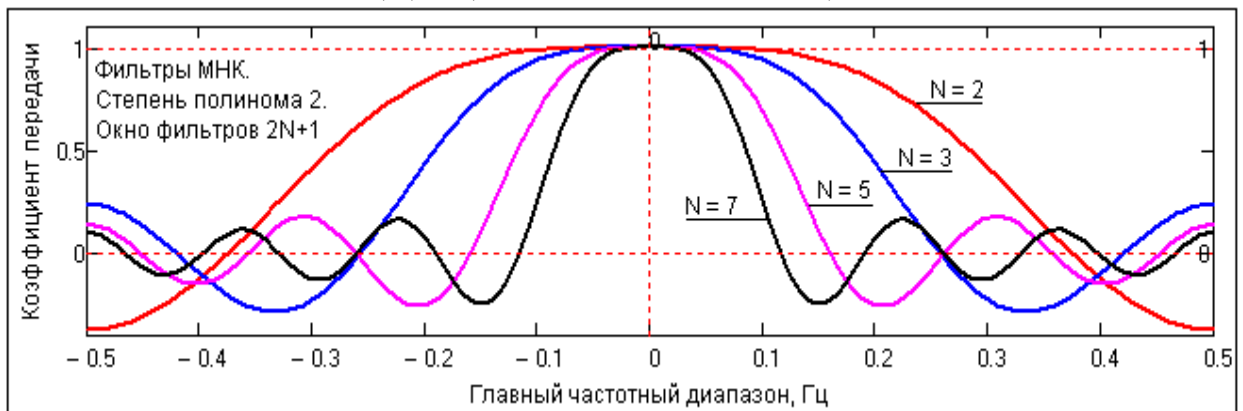


Рис. 3.2.1. Частотные характеристики сглаживающих фильтров МНК-2

Вывод формул передаточных функций для 7, 9, 11-ти точечных фильтров МНК-2 предлагается для самостоятельной работы.

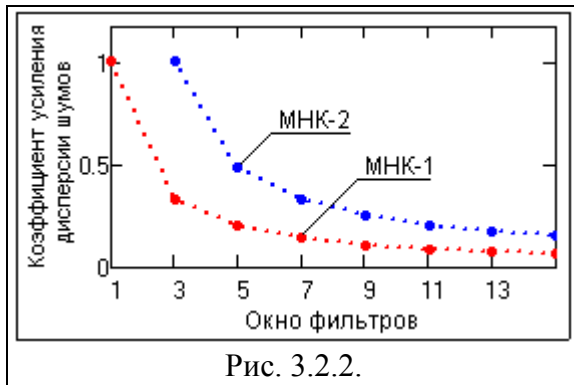


Рис. 3.2.2.

Вид частотных характеристик фильтров МНК-2 приводится на рис. 3.2.1. При сравнении характеристик с характеристиками фильтров МНК-1 можно видеть, что повышение степени полинома расширяет низкочастотную полосу пропускания фильтра и увеличивает крутизну ее среза. За счет расширения полосы пропускания главного частотного диапазона при тех же значениях  $N$  коэффициенты усиления дисперсии шумов фильтров МНК-2 выше, чем фильтров 1-го

порядка, что можно видеть на рис. 3.2.2.

Методика выбора окна фильтра под частотные характеристики входных сигналов не отличается от фильтров МНК 1-го порядка. Для получения примерно равных значений подавления шумов и коэффициента сигнал/шум на выходах фильтров, фильтры МНК-2 должны иметь в 2 раза большую ширину окна, чем фильтры МНК-1. Об этом свидетельствует пример моделирования фильтрации, приведенный на рис. 3.2.4, в котором графики, приведенные выше на рис. 3.1.8 для МНК-1, дополнены графиком фильтрации эквивалентным фильтром МНК-2. Нетрудно также заметить, что выходные данные МНК-2 имеют более высокую гладкость, чем на выходе МНК-1.

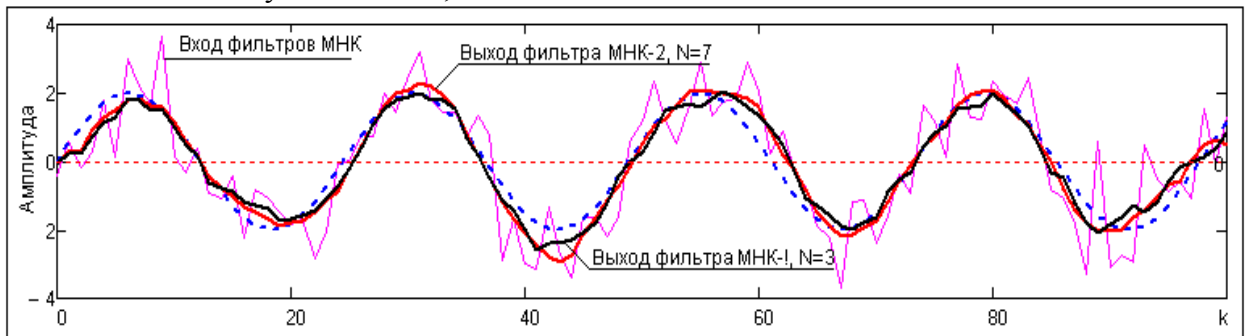


Рис. 3.2.3.

**Модификация фильтров.** Фильтры МНК второго порядка (равно как фильтры МНК-1 и другие фильтры подобного назначения) также можно модифицировать по условию  $H(\omega) \rightarrow 0$  при  $\omega \rightarrow \pi$ , что снижает пульсации передаточной функции фильтра в полосе подавления при небольшом увеличении зоны пропускания. Один из простейших методов модификации заключается в следующем. В выражение передаточной функции (со всеми коэффициентами фильтра, вида (3.2.4)) подставляем  $z = \exp(-j\omega)$ , заменяем значения концевых коэффициентов фильтра на параметры, принимаем  $\omega = \pi$ , и, приравняв полученное выражение нулю, находим новые значения концевых коэффициентов, после чего сумму всех коэффициентов нормируем к 1 при  $\omega = 0$ .

~> **Пример модификации фильтра МНК 2-го порядка.**

~> Передаточная функция: выражение (3.2.4). Частотная характеристика (нормировку можно снять):

~>  $H(\omega) = -3\exp(2j\omega) + 12\exp(j\omega) + 17 + 12\exp(-j\omega) - 3\exp(-2j\omega)$ .

~> Замена концевых коэффициентов {значение 3} на параметр  $b$  и упрощение:

~>  $H(\omega) = 17 + 24 \cos(\omega) + 2b \cos(2\omega)$ .

~> При  $\omega = \pi$ :  $H(\pi) = 17 - 24 + 2b = 0$ . Отсюда:  $b = 3.5$

~> Новая частотная характеристика (с приведением коэффициентов к целым числам):

~>  $H(\omega) = 68 + 96 \cos(\omega) + 14 \cos(2\omega)$ . Сумма коэффициентов при  $\omega = 0$ :  $H(0) = 68 + 96 + 14 = 178$ .

~> Нормированная частотная характеристика:  $H(\omega) = (68 + 96 \cos(\omega) + 14 \cos(2\omega)) / 178$ .

~> Коэффициенты фильтра:  $h_n = \{(7, 48, 68, 48, 7) / 178\}$ .

~> **Пример-** задание: Модифицировать 7, 9 и 11-ти точечные сглаживающие фильтры МНК 2-го порядка.

~> Контроль:  ${}^7h_n = \{(1, 6, 12, 14, 12, 6, 1) / 52\}$ .  ${}^9h_n = \{(-1, 28, 78, 108, 118, 108, 78, 28, -1) / 548\}$ .

$${}^{11}h_n = \{(-11, 18, 88, 138, 168, 178, 168, 138, 88, 18, -11)/980\}.$$

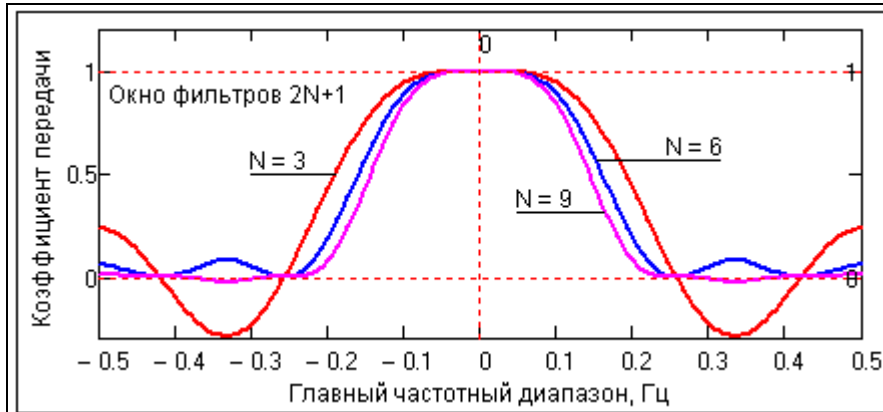


Рис. 3.2.4.

**Последовательная фильтрация.** Из фильтров МНК-2, как и из фильтров МНК-1, можно конструировать новые фильтры, частотные характеристики которых соответствуют последовательному применению «родительских» фильтров. Методика конструирования

аналогична. Пример частотных характеристик конструирования новых фильтров из 7-ми точечного фильтра МНК-2 приведен на рис. 3.2.4.

Фильтры МНК третьего порядка по своим частотным характеристикам эквивалентны фильтрам второго порядка.

### 3.3. ФИЛЬТРЫ МНК 4-го ПОРЯДКА

**Фильтры МНК 4-го порядка.** Расчет по аналогичной методике сглаживающих фильтров МНК 4-ой степени дает следующие результаты:

$$h_{0-3} = (131,75, -30,5)/231,$$

$$h_{0-4} = (179,135,30, -55,15)/429,$$

$$h_{0-5} = (143,120,60, -10, -45,18)/429.$$

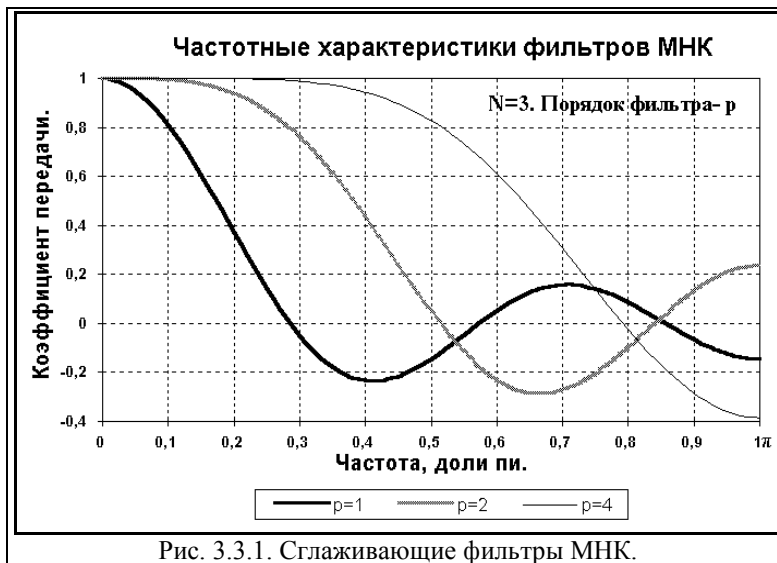


Рис. 3.3.1. Сглаживающие фильтры МНК.

На рис. 3.3.1 приведено сопоставление частотных характеристик одномерных фильтров МНК 1-го, 2-го и 4-го порядка.

В целом, по сглаживающим фильтрам МНК можно сделать следующие выводы:

1. Повышение порядка фильтра увеличивает степень касания частотной характеристикой уровня коэффициента передачи  $N=1$  на частоте  $\omega = 0$  и расширяет полосу пропускания фильтра.

2. Увеличение количества членов фильтра приводит к сужению полосы пропускания и увеличивает крутизну ее среза.

3. Модификация фильтров уменьшает осцилляции передаточной функции в полосе подавления сигналов.

Совместное изменение этих параметров позволяет подбирать для сглаживания данных такой фильтр МНК, частотная характеристика которого наилучшим образом удовлетворяет частотному спектру сигналов при минимальном количестве коэффициентов фильтра.

### 3.4. РАСЧЕТ ПРОСТОГО ФИЛЬТРА ПО ЧАСТОТНОЙ ХАРАКТЕРИСТИКЕ.

Если шумы в обрабатываемых сигналах сосредоточены в основном в высокочастотной области, то достаточно простые фильтры сглаживания без значительных осцилляций могут быть синтезированы непосредственно по частотной характеристике. В качестве примера проведем расчет простого симметричного сглаживающего НЦФ с окном в пять точек:

$$y_k = as_{k-2} + bs_{k-1} + cs_k + bs_{k+1} + as_{k+2}. \quad (3.4.1)$$

Полагаем  $s_k = \exp(j\omega k)$ , при этом  $y_k = H(\omega) \exp(j\omega k)$ . Подставляем значения входного и выходного сигнала в уравнение фильтра, сокращаем левую и правую части на общий член  $\exp(j\omega k)$  и, объединяя комплексно сопряженные члены в правой части, получаем уравнение передаточной функции:

$$H(\omega) = 2a \cos 2\omega + 2b \cos \omega + c.$$

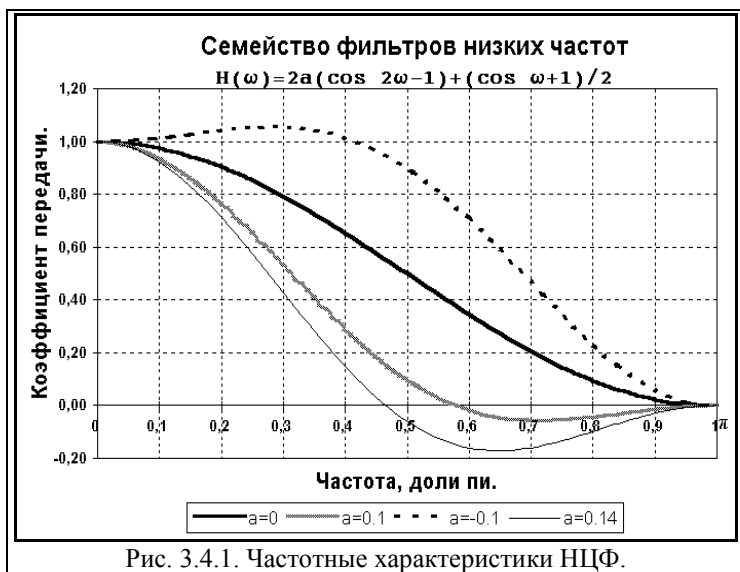


Рис. 3.4.1. Частотные характеристики НЦФ.

значений 'a' и выбрать фильтр, удовлетворяющий заданию. Пример семейства частотных характеристик приведен на рисунке 3.4.1.

Можно наложить еще одно дополнительное условие и определить все коэффициенты фильтра непосредственно. Так, например, если к двум граничным условиям задать третье условие сбалансированности:  $H(\omega) = 0.5$  при  $\omega = \pi/2$ , то из трех полученных уравнений сразу же получим все три коэффициента фильтра:  $a = 0$ ,  $b = 1/4$ ,  $c = 1/2$  (фильтр сокращается до трех точек).

В принципе, таким методом можно задать любую произвольную форму частотной характеристики симметричного НЦФ с произвольным количеством N точек дискретизации, что определит полное уравнение (3.4.1) с окном 2N+1 точка и соответствующую передаточную функцию фильтра, по которой можно составить и решить N+1 уравнение для определения коэффициентов фильтра.

#### Тема 4. РАЗНОСТНЫЕ ФИЛЬТРЫ И ФИЛЬТРЫ ИНТЕГРИРОВАНИЯ.

Основной инструмент проектирования цифровых фильтров – частотный (спектральный) анализ. Частотный анализ базируется на использовании периодических функций синусов и косинусов. По-существу, спектральная характеристика цифрового фильтра – это тонкая внутренняя структура системы, его однозначный функциональный паспорт направленного изменения частотного состава данных, полностью определяющий сущность преобразования фильтром входных данных.

Рассмотрим примеры синтеза и частотного анализа фильтров применительно к известным способам дифференцирования и интегрирования цифровых данных.

#### 4.1. РАЗНОСТНЫЕ ОПЕРАТОРЫ

Примеры частотного подхода при анализе разностных операторов.

**Разностный оператор** 1-го порядка имеет вид:

$$\Delta s_k = s_{k+1} - s_k.$$

Последовательное  $n$ -кратное применение оператора записывается в виде оператора  $n$ -го порядка:

$$\Delta^n(s_k) = \Delta[\Delta^{n-1}(s_k)] = \Delta s_k \otimes \Delta^{n-1}(s_k) \quad (4.1.1)$$

k	$s_k$	$\Delta(s_k)$	$\Delta^2(s_k)$	$\Delta^3(s_k)$	$\Delta^4(s_k)$	$\Delta^5(s_k)$	$\Delta^6(s_k)$
-7	0	0	0	0	0	0	0
-6	0	0	0	0	0	0	1
-5	0	0	0	0	0	1	-6
-4	0	0	0	0	1	-5	15
-3	0	0	0	1	-4	10	-20
-2	0	0	1	-3	6	-10	15
-1	0	1	-2	3	-4	5	-6
0	1	-1	1	-1	1	-1	1
1	0	0	0	0	0	0	0
$K_q$		2	6	20	70	252	924

Выходные значения импульсной реакции разностных операторов на единичный импульсный сигнал Кронекера приведены в таблице. Ряды последовательных разностей содержат знакопеременные биномиальные коэффициенты. В представленной форме разностные операторы являются каузальными фазосдвигающими (односторонними) фильтрами, но нетрудно заметить, что

операторы четных степеней могут быть переведены в симметричную форму сдвигом влево на половину окна оператора.

В последней строке таблицы приводятся коэффициенты усиления дисперсии шумов, значение которых резко нарастает по мере увеличения порядка оператора. Это позволяет использовать разностные операторы с порядком выше 1 для определения местоположения статистически распределенных шумов в массивах данных. Особенно наглядно эту возможность можно видеть на частотных характеристиках операторов.

Подставляя сигнал  $s(k) = \exp(j\omega k)$  в (4.1.1) и упрощая, получаем:

$$\begin{aligned} \Delta^n s(k) &= (j^n) \exp(j\omega n/2) [2 \sin(\omega/2)]^n \exp(j\omega k). \\ H(\omega) &= (j^n) \exp(j\omega n/2) [2 \sin(\omega/2)]^n \end{aligned} \quad (4.1.2)$$

Так как модуль первых двух множителей в выражении (4.1.2) равен 1, зависимость коэффициента передачи разностного оператора от частоты определяется вторым сомножителем  $(2 \sin(\omega/2))^n$  и представлена на рисунке 4.1.1.

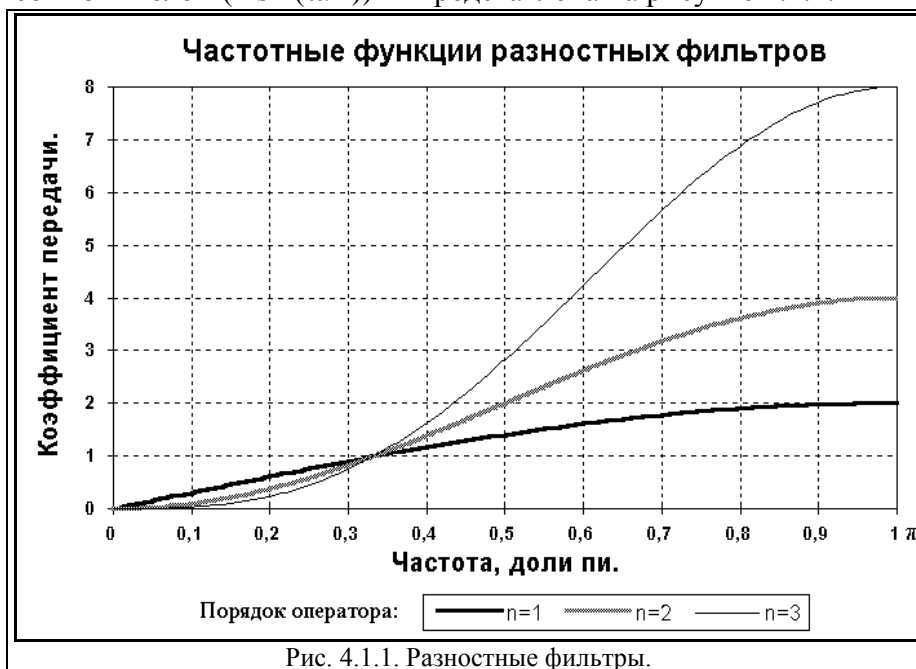


Рис. 4.1.1. Разностные фильтры.

**Выделение в сигналах шумов.** Как следует из графиков на рис. 4.1.1, разностные операторы подавляют постоянную составляющую сигнала и его гармоники в первой трети интервала Найквиста и увеличивают высокочастотные составляющие сигнала в остальной части интервала тем больше, чем больше порядок оператора.

Как правило, эту часть главного интервала спектра сигналов занимают высокочастотные статистические шумы.

Шумы при анализе данных также могут представлять собой определенную информацию, например, по стабильности условий измерений и по влиянию на измерения внешних дестабилизирующих факторов. На рис. 4.1.2 приведен пример выделения интервалов интенсивных шумов в данных акустического каротажа, что может свидетельствовать о сильной трещиноватости пород на этих интервалах. Такая информация относится уже не шумовой, а к весьма полезной информации при поисках и разведке нефти, газа и воды.

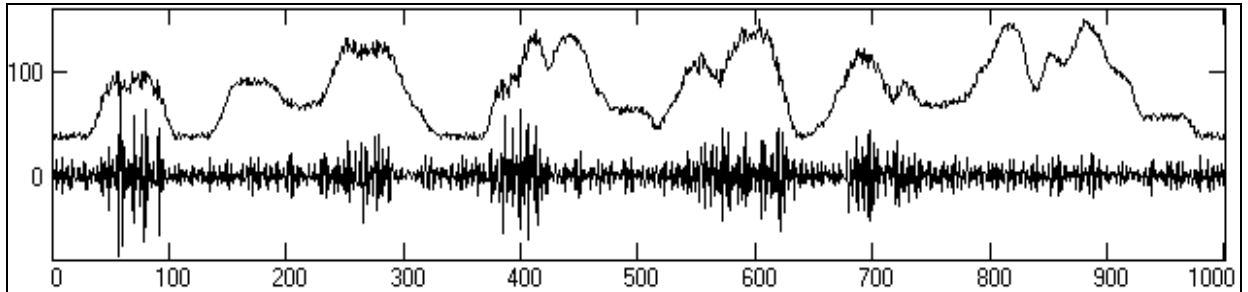


Рис. 4.1.2.

**Восстановление утраченных данных.** Разностные операторы имеют одну особенность: оператор  $n+1$  порядка аннулирует полином степени  $n$ , т.е. свертка оператора порядка  $n+1$  с полиномом  $n$ -ой степени дает нулевые значения:  $\Delta^{n+1} \otimes P_n(k) = 0$ .

---

**Пример.**  $P_2(k) = x_k = 1+2k-k^2$ ,  $k = 0,1,2,\dots$   $x_k = 1,2,1,-2,-7,-14,-23,-34,\dots$   $y_k = x_k \otimes \Delta^3 = 0,0,0,0,\dots$

---

Эту особенность можно использовать для создания очень простых и достаточно надежных операторов восстановления в массивах пропущенных и утраченных значений или для замены аннулированных при обработке величин (например, явных выбросов).

Если считать, что отрезок данных, содержащий пропуск, является многочленом некоторой степени, то свертка данных с разностным оператором следующего порядка должна быть равна нулю. Так, при аппроксимации данных многочленом третьей степени для любой точки массива должно выполняться равенство:

$$\Delta^4 \otimes (s_k) = s_{k-2} - 4s_{k-1} + 6s_k - 4s_{k+1} + s_{k+2} = 0.$$

Интерполяционный фильтр восстановления утраченной центральной точки данных:

$$s_k = (-s_{k-2} + 4s_{k-1} + 4s_{k+1} - s_{k+2})/6. \quad (4.1.3)$$

Соответственно, оператор фильтра восстановления данных  $h(n) = (-1, 4, 0, 4, -1)/6$ . Коэффициент усиления шумов  $\sigma^2 = 17/18 = 0.944$ .

---

**Пример.** Фактический отрезок массива данных:  $x_k = \{3, 6, 8, 8, 7, 5, 3, 1\}$ .  
 Допустим, что на отрезке был зарегистрирован явный выброс:  $x_k = \{3, 6, 8, 208, 7, 5, 3, 1\}$ .  
 Отсчет с выбросом аннулирован. Замена отсчета:  $x_3 = (-x_1 + 4x_2 + 4x_4 - x_5)/6 = (-6 + 32 + 28 - 5)/6 \approx 8.17$ .  
 В массиве утрачен 5-й отсчет. Восстановление:  $x_4 = (-x_2 + 4x_3 + 4x_5 - x_6)/6 = (-8 + 32 + 20 - 3)/6 \approx 6.83$ .

---



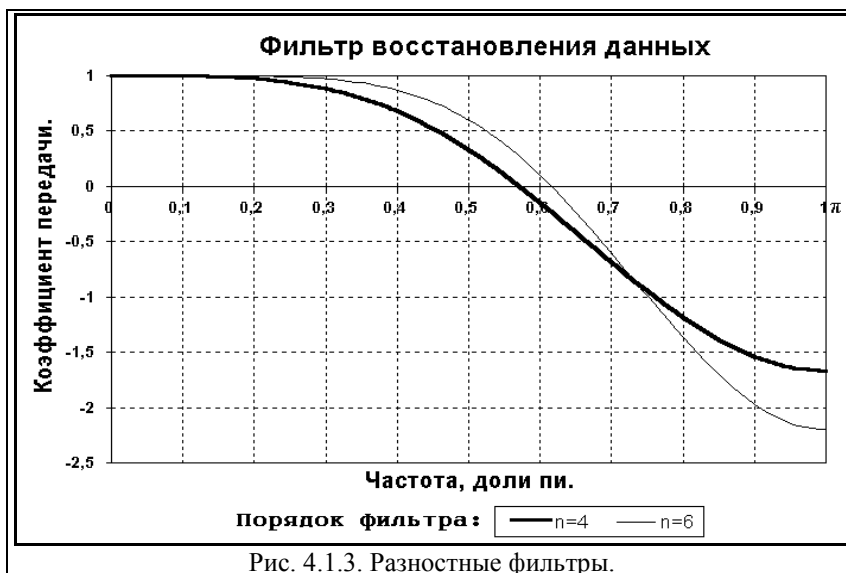


Рис. 4.1.3. Разностные фильтры.

Принимая в (4.1.3)  $k = 0$  и подставляя сигнал  $s_k = \exp(j\omega k)$ , получаем частотную характеристику фильтра восстановления данных 4-го порядка:

$$H(\omega) = (4 \cos \omega - \cos 2\omega)/3.$$

Вид частотной характеристики для фильтров восстановления пропущенных данных 4-го и 6-го порядков приведен на рис. 4.1.3. Графики наглядно показывают, что применение разностных

интерполяционных фильтров восстановления данных возможно только для сигналов, высокочастотные составляющие которых минимум в три раза меньше частоты Найквиста. Интерполяционные фильтры выше 4-го порядка применять не рекомендуется, т.к. они имеют коэффициент усиления шумов более 1.

На рис. 4.1.4 – 4.1.6 приведены примеры восстановления утраченных данных во входных сигналах оператором 3-го порядка и спектры сигналов в сопоставлении с передаточной функцией оператора восстановления данных.

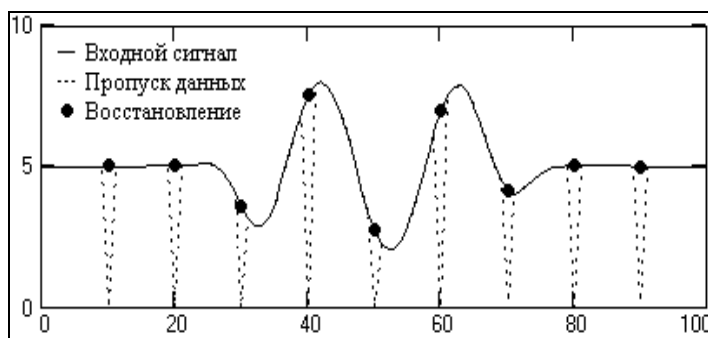


Рис. 4.1.4. Восстановление незашумленных данных.

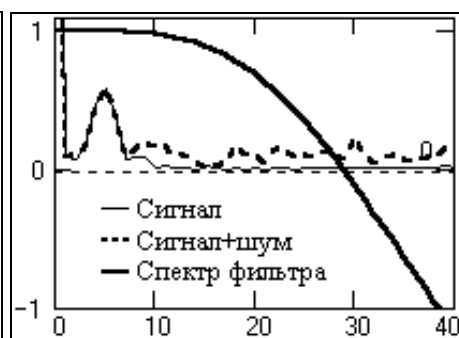


Рис.4.1.5. Спектры.

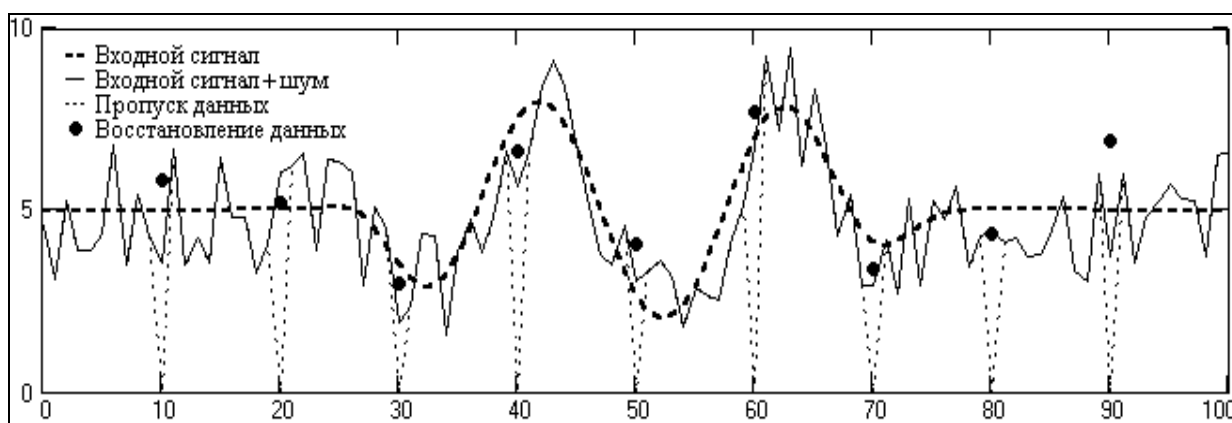


Рис. 4.1.6. Восстановление зашумленных данных.

В сигналах, представленных на рисунках, утрачен каждый 10-ый отсчет (например, при передаче данных) при сохранении тактовой частоты нумерации данных. Учитывая, что все значения входных сигналов положительны, индикатор пропуска данных для работы оператора служат нулевые значения. В любых других случаях для оператора восстановления данных необходимо предусматривать специальный маркер (например,

заменять аннулированные данные или выбросы определенным большим или малым значением отсчетов).

Как следует из рис. 4.1.5, спектр полезного сигнала полностью находится в зоне единичного коэффициента частотной характеристики оператора, и восстановление данных выполняется практически без погрешности (рис. 4.1.4). При наложении на сигнал статистически распределенных шумов (рис. 4.1.6) погрешность восстановления данных увеличивается, но для информационной части полного сигнала она, как и во входных данных, не превышает среднеквадратического значения (стандарта) флюктуаций шума.

**Аппроксимация производных** - вторая большая область применения разностных операторов. Оценки первой, второй и третьей производной можно производить по простейшим формулам дифференцирования:

$$(s_n)' = (s_{n+1} - s_{n-1}) / 2\Delta t. \quad h1 = \{-0.5, 0, 0.5\}. \quad (4.1.4)$$

$$(s_n)'' = (s_{n+1} - 2s_n + s_{n-1}) / \Delta t. \quad h2 = \{1, -2, 1\}.$$

$$(s_n)''' = (-s_{n+2} + 2s_{n+1} - 2s_{n-1} + s_{n-2}) / 2\Delta t. \quad h3 = \{0.5, -1, 0, 1, -0.5\}.$$

Оператор первой производной является нечетной функцией и имеет мнимый спектр. Если принять  $s(t) = \exp(j\omega t)$ , то истинное значение первой производной должно быть равно:  $s'(t) = j\omega \exp(j\omega t)$ . Передаточная функция  $H(\omega) = j\omega$ . Оценка первой производной в точке  $n = 0$  по разностному оператору при  $\Delta t = 1$ :  $s'(0) = (\exp(j\omega) - \exp(-j\omega)) / 2 = j \sin \omega = H1(\omega)$ . Отношение расчетного значения к истинному на той же точке:  $K1(\omega) = \sin(\omega) / \omega$ . Графики функций в правой половине главного диапазона приведены на рис. 4.1.7.

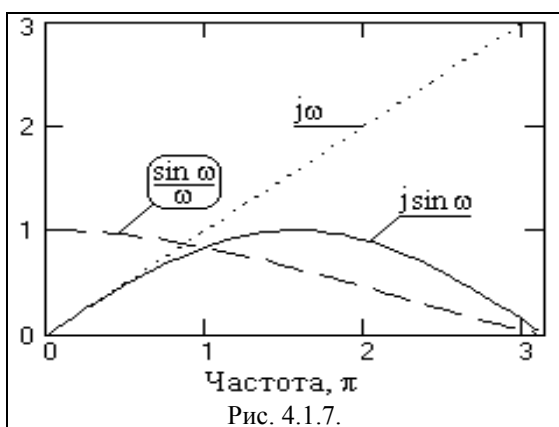


Рис. 4.1.7.

Как следует из приведенных выражений и графиков, значение  $K(\omega)$  равно 1 только на частоте  $\omega = 0$ . На всех других частотах в интервале Найквиста формула дает заниженные значения производных. Однако при обработке практических данных последний фактор может играть и положительную роль, если сигнал низкочастотный (не более 1/3 главного диапазона) и зарегистрирован на уровне высокочастотных шумов. Любое дифференцирование поднимает в спектре сигнала долю его высокочастотных составляющих. Коэффициент усиления дисперсии шумов разностным оператором дифференцирования непосредственно по его спектру в главном диапазоне:

$$K_q = (1/\pi) \int_0^\pi (\sin \omega)^2 d\omega = 0.5.$$

При точном дифференцировании по всему главному диапазону:

$$K_q = (1/\pi) \int_0^\pi \omega^2 d\omega = 3.29.$$

Следовательно, разностный оператор имеет практически в шесть раз меньший коэффициент усиления дисперсии шумов, чем точный оператор дифференцирования.

На рис. 4.1.8 показан пример дифференцирования гармоники с частотой 0.1 частоты Найквиста (показана пунктиром) и этой же гармоники с наложенными шумами (сплошная кривая).

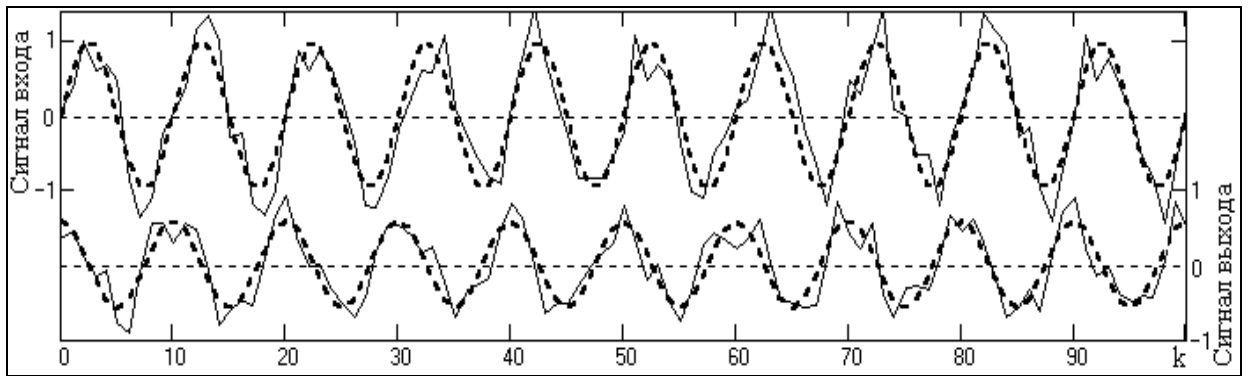


Рис. 4.1.8. Пример дифференцирования (входные сигналы – вверху, выходные – внизу).

Оператор второй производной относится к типу четных функций. Частотная функция оператора:  $H_2(\omega) = -2(1 - \cos \omega)$ . Собственное значение операции  $H(\omega) = -\omega^2$ . Отношение фактического значения к собственному

$$K_2(\omega) = [\sin(\omega/2)/(\omega/2)]^2,$$

и также равно 1 только на частоте  $\omega = 0$ . На всех других частотах в интервале Найквиста формула дает заниженные значения производных, хотя и меньшие по относительным значениям, чем оператор первой производной. Частотные графики дифференцирования приведены на рис. 4.1.9. Коэффициент усиления дисперсии шумов оператором второй производной равен 6 при собственном значении дифференцирования, равном 19.5. Эти значения показывают, что операция двойного дифференцирования может применяться только для данных, достаточно хорошо очищенных от шумов, с основной энергией сигнала в первой трети интервала Найквиста.

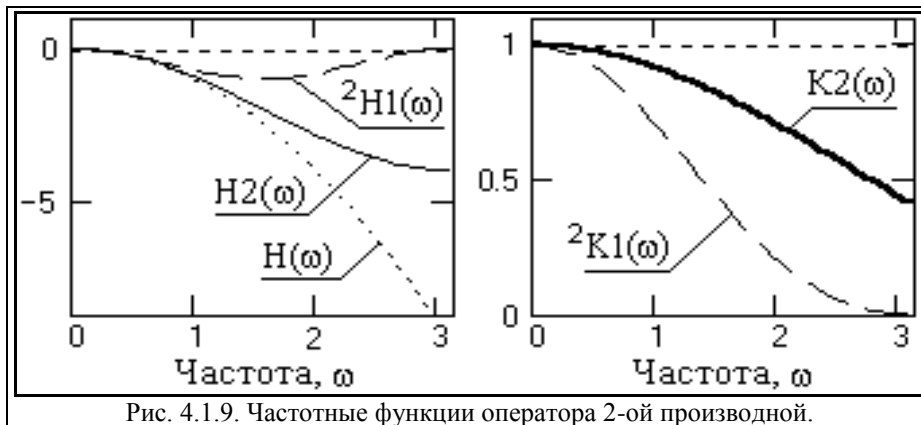


Рис. 4.1.9. Частотные функции оператора 2-ой производной.

Оператор последовательного двойного дифференцирования можно получить сверткой оператора первой производной с самим собой:

$${}^2h_1 = h_1 \otimes h_1 = \{0.25, 0, -0.5, 0, 0.25\},$$

и имеет коэффициент усиления дисперсии шумов всего 0.375. Частотная характеристика оператора:

$${}^2H_1(\omega) = -0.5[1 - \cos(2\omega)].$$

Графики  ${}^2H_1(\omega)$  и коэффициента соответствия  ${}^2K_1(\omega)$  приведены пунктиром на рис. 4.1.9. Из их сопоставления с графиками второй производной можно видеть, что последовательное двойное дифференцирование возможно только для данных, спектральный состав которых занимает не более пятой начальной части главного диапазона, и по точности хуже оператора второй производной.

В принципе, вторую производную можно получать и последовательным двойным дифференцированием данных оператором первой производной. Однако для таких простых операторов эти две операции не тождественны.

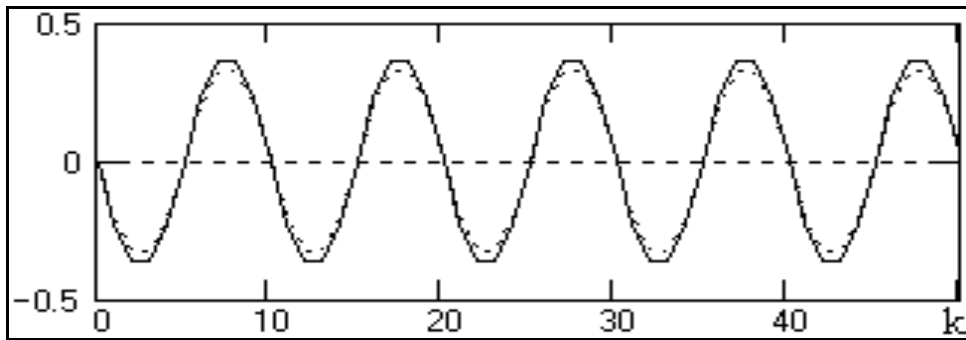


Рис. 4.1.10. Вторая производная гармоники с частотой  $\omega=0.2\pi$  при  $\Delta t=1$  (пунктир – двойное последовательное дифференцирование)

Пример применения двух операторов второй производной приведен на рис. 4.1.10.

Частота Найквиста главного диапазона обратно пропорциональна интервалу  $\Delta t$  дискретизации данных ( $\omega_N = \pi/\Delta t$ ), а, следовательно, интервал дискретизации данных для корректного использования простых операторов дифференцирования должен быть в 3-5 раз меньше оптимального для сигналов с известными предельными частотами спектрального состава.

Частотные функции для третьей производной предлагается получить самостоятельно.

#### 4.2. ИНТЕГРИРОВАНИЕ ДАННЫХ

Интегрирование сигналов реализуется рекурсивными цифровыми фильтрами. Рассмотрим примеры анализа интегрирующих операторов.

Как известно, для точной операции интегрирования финитных сигналов в общем случае действительно преобразование:

$$\int_t s(t) dt \leftrightarrow (1/j\omega) S(\omega).$$

Это выражение в правой части имеет особую точку при  $\omega = 0$  и, соответственно, весовой дельта-импульс на нулевой частоте. Оператор интегрирования в частотной области  $(1/j\omega)$  при  $\omega > 1$  ослабляет в амплитудном спектре высокие частоты, а при  $0 < \omega < 1$  усиливает низкие. Фазовый спектр сигнала смещается на  $-90^\circ$  для положительных частот и на  $90^\circ$  для отрицательных.

Наиболее простыми и распространенными на практике алгоритмами интегрирования являются цифровые аналоги формул трапеций, прямоугольников и Симпсона.

*Алгоритм интегрирования по формуле трапеций* при нулевых начальных условиях:

$$y_{k+1} = y_k + (s_{k+1} + s_k)/2. \quad (4.2.1)$$

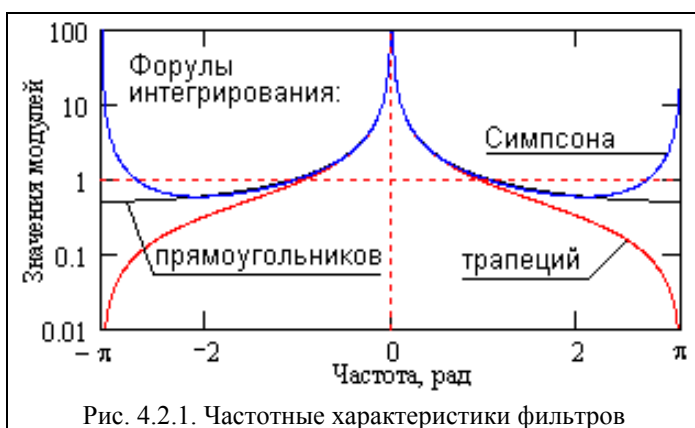


Рис. 4.2.1. Частотные характеристики фильтров

Принимая  $s_k = \exp(j\omega t)$  и  $y_k = H(\omega) \exp(j\omega t)$ , подставляем сигналы в (4.2.1) при  $t_k = k\Delta t$ ,  $\Delta t = 1$  и решаем относительно  $H(\omega)$ . Получаем:

$$H(\omega) = \cos(\omega/2) / [2j \sin(\omega/2)].$$

Частотная характеристика фильтра (в логарифмическом масштабе), а также фильтров интегрирования по другим формулам, приведена на рис. 4.2.1.

В связи с накоплением результатов по всему предыдущему циклу суммирования и большим диапазоном значений модуля АЧХ фильтра более

удобными, представительными и информационными являются частотные функции коэффициентов соответствия фактического интегрирования истинному:

$$K(\omega) = H(\omega)\exp(j\omega t)/[(1/j\omega)\exp(j\omega t)].$$

$$K(\omega) = \cos(\omega/2)[(\omega/2)/\sin(\omega/2)]. \quad (4.2.2)$$



Графики коэффициентов соответствия всех фильтров интегрирования приведены на рис. 4.2.2

**Оператор интегрирования по формуле прямоугольников** (интерполяционное средноточечное):

$$y_{k+1} = y_k + s_{k+1/2}. \quad (4.2.3)$$

После аналогичных подстановок сигнала и преобразований получаем:

$$K(\omega) = (\omega/2)/\sin(\omega/2).$$

**При численном**

**интегрировании по формуле Симпсона** уравнение фильтра имеет вид:

$$y_{k+1} = y_{k-1} + (s_{k+1} + 4s_k + s_{k-1})/6. \quad (4.2.4)$$

Частотный анализ фильтра проведите самостоятельно. Контроль:

$$K(\omega) = (2 + \cos \omega) / [3 \sin(\omega)/\omega].$$

Наиболее простые формулы цифрового интегрирования (трапеций и прямоугольников) ведут себя различным образом в главном частотном диапазоне. Формула прямоугольников завышает результаты на высоких частотах, а формула трапеций - занижает. Эти особенности легко объяснимы. Для одиночной гармоники площадь трапеции по двум последовательным отсчетам всегда меньше, чем площадь с выпуклой дугой гармоники между этими отсчетами, и разница тем больше, чем больше частота. В пределе, для гармоники с частотой Найквиста, отсчеты соответствуют знаочередующемуся ряду (типа 1, -1, 1, -1, ... или любые другие значения в зависимости от амплитуды и начального фазового угла) и при нулевых начальных условиях суммирование двух последовательных отсчетов в формуле (4.2.1) будет давать 0 и накопления результатов не происходит. Интегрирование по площади прямоугольников с отчетом высоты по центральной точке между двумя отсчетами всегда ведет к завышению площади прямоугольника относительно площади, ограниченной выпуклой дугой гармоники.

Формула Симпсона отличается от формул трапеций и прямоугольников более высокой степенью касания единичного значения, что обеспечивает более высокую точность интегрирования в первой половине главного диапазона. Однако на высоких частотах погрешность начинает резко нарастать вплоть до выхода на бесконечность на конце диапазона (полюс в знаменателе передаточной функции рекурсивного фильтра на частоте Найквиста).

Эти особенности интегрирования следует учитывать при обработке данных сложного спектрального состава. Пример интегрирования сигнала и изменения его спектра приведен на рис. 4.2.3.

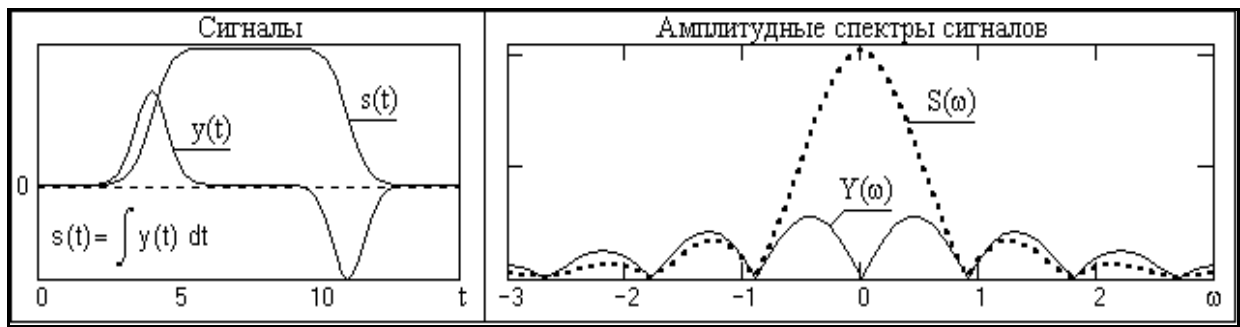


Рис. 4.2.3.

## ТЕМА 5. ФИЛЬТРАЦИЯ СЛУЧАЙНЫХ СИГНАЛОВ

Если сигнал на входе фильтра является детерминированным, то его соотношение с выходным сигналом однозначно определяется импульсным откликом фильтра. Таким же однозначным является соотношение входа - выхода и для случайных сигналов, однако в силу природы последних аналитическое представление как входного сигнала, так и отклика системы, не представляется возможным. Для описания реакции фильтра на случайный входной сигнал используется статистический подход.

### 5.1. ФИЛЬТРАЦИЯ СЛУЧАЙНЫХ СИГНАЛОВ

Если параметры случайного входного сигнала специально не оговариваются, то по умолчанию принимается, что на вход фильтра поступает реализация случайного стационарного процесса  $x(k\Delta t)$  с нулевым средним, которая вызывает сигнал  $y(k\Delta t)$  на выходе фильтра. Значение  $\Delta t$ , как обычно, принимаем равным 1.

**Сохранение природы сигнала.** Допустим, что фильтр имеет импульсный отклик  $h(n) = \exp(-a \cdot n)$ ,  $n \geq 0$ . Зададим на входе фильтра стационарный квазидетерминированный случайный сигнал, который не обладает свойством эргодичности, но имеет все свойства случайного сигнала, и может быть описан в явной математической форме:

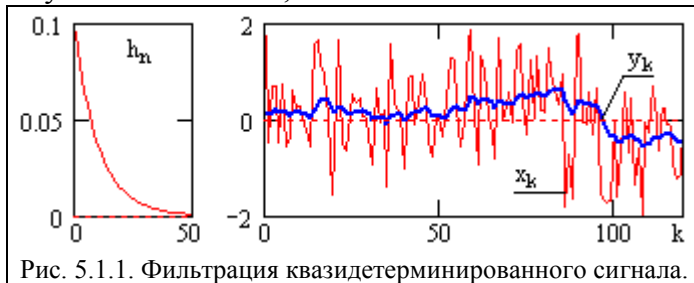


Рис. 5.1.1. Фильтрация квазидетерминированного сигнала.

$$x(k) = A + \cos(2k + \varphi),$$

где  $A$  и  $\varphi$  - взаимно независимые случайные величины, причем значение  $\varphi$  равномерно распределено в интервале  $[0, 2\pi]$ . При этом выходной сигнал определяется выражением:

$$y(k) = h(n) \otimes x(k-n) \equiv \sum_{n=0}^N h(n) x(k-n)$$

$$y(k) = A/3 + [3 \cos(2k + \varphi) + 2 \sin(2k + \varphi)]/13.$$

Из этого выражения следует, что выходной сигнал фильтра также является случайным и содержит те же самые случайные параметры, что и входной сигнал, а, следовательно, для него существуют определенные статистические характеристики. Пример реализации квазидетерминированного случайного сигнала и его фильтрации аналогом сглаживающего RC-фильтра приведен на рис. 5.1.1.

**Математическое ожидание** (индекс операции -  $M$ ) произвольного входного случайного стационарного сигнала  $x(k)$  на выходе фильтра определяется выражением:

$$\begin{aligned} \bar{y} &= M\{y(k)\} = M\left\{\sum_n h(n) x(k-n)\right\} = \sum_n M\{x(k-n)\} h(n) = \\ &= \bar{x} \sum_n h(n) = \bar{x} K_{\text{пс}}. \end{aligned} \quad (5.1.1)$$

Отсюда следует, что математическое ожидание выходных сигналов фильтра равно математическому ожиданию входных сигналов, умноженному на коэффициент усиления фильтром постоянной составляющей. При  $K_{\text{пс}} = 1$  среднее значение выходных сигналов не изменяется и равно среднему значению входных сигналов. Если фильтр не пропускает

постоянную составляющую сигналов (сумма коэффициентов импульсного отклика фильтра равна нулю), то случайный выходной сигнал всегда будет иметь нулевое математическое ожидание.

**Корреляционные соотношения.** Для нецентрированных входных сигналов  $x(k)$  размером  $(0-K)$  автокорреляционная функция (АКФ), а равно и функция автоковариации  $K_x(n)$  (ФАК) для центрированных случайных сигналов, вычисляется по формуле:

$$R_x(n) = [1/(K+1-n)] \sum_{k=0}^{K-n} x(k) x(k+n). \quad (5.1.2)$$

Формула применяется довольно редко, в основном для детерминированных сигналов с небольшим числом отсчетов. Для случайных и зашумленных сигналов уменьшение знаменателя  $(K-n)$  и числа перемножаемых отсчетов по мере увеличения сдвига приводит к нарастанию статистических флуктуаций вычисления АКФ. Большую достоверность в этих условиях обеспечивает вычисление АКФ в единицах мощности сигнала по формуле:

$$R_s(n) = \frac{1}{K} \sum_{k=0}^K s_k \cdot s_{k+n}, \quad s_{k+n} = 0 \text{ при } k+n > K, \quad (5.1.3)$$

т.е. с нормированием на постоянный множитель  $1/K$  и с продлением сигнала нулевыми значениями (в левую сторону при сдвигах  $k-n$  или в правую сторону при использовании сдвигов  $k+n$ ). Эта оценка является смещенной и имеет несколько меньшую дисперсию, чем по формуле (5.1.2). Разницу между нормировками по формулам (5.1.2) и (5.1.3) можно наглядно видеть на рис. 5.1.2.

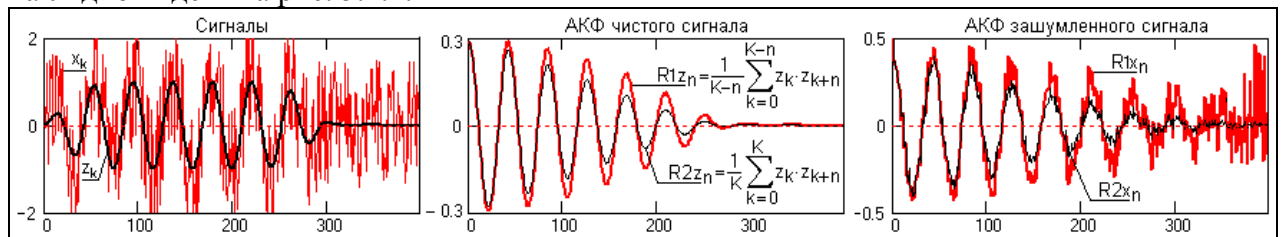


Рис. 5.1.2.

Формулу (5.1.3) можно рассматривать, как усреднение суммы произведений, т.е. как оценку математического ожидания:

$$R_s(n) = M\{s_k s_{k+n}\} \cong \overline{s_k s_{k+n}}. \quad (5.1.4)$$

По аналогичной формуле может быть вычислена и АКФ выходных сигналов. Для произведения выходных сигналов  $y(k)$  и  $y(k+n)$ , образующих функцию автокорреляции выходных сигналов, можно также записать (без дополнительных множителей):

$$y(k) y(k+n) = \sum_j \sum_i h(i)h(j) x(k-i)x(k+n-j).$$

Если взять математические ожидания от обеих частей этого равенства, то, с учетом соотношения в правой части под знаками сумм

$$M\{x(k-i) x(k+n-j)\} = -R_x(k-i-k-n+j) = R_x(n+i-j),$$

получим:

$$R_y(n) = \sum_i \sum_j h(i)h(j) R_x(n+i-j) \equiv R_x(n) \otimes h(n+i) \otimes h(n-j). \quad (5.1.5)$$

Таким образом, функция автокорреляции выходного сигнала равна АКФ входного сигнала, свернутой дважды, в прямом и обратном направлении, с импульсным откликом фильтра, что сохраняет четность АКФ выходного сигнала. Для центрированных процессов аналогичное заключение действительно и для ковариационных функций. На рис. 5.1.3 приведен пример нормированных АКФ входной и выходной случайных последовательностей при фильтрации RC-фильтром, форма импульсного отклика которого также приведена на рисунке.

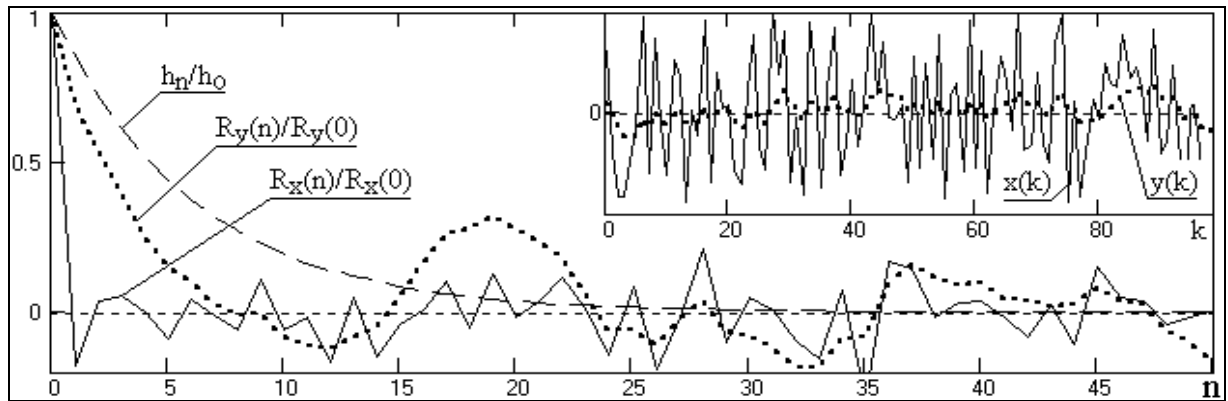


Рис. 5.1.3. Функции корреляционных коэффициентов.

Заметим, что для свертки импульсных откликов, производя замену  $n-j = m$ , мы имеем равенство:

$$h(n+i) \otimes h(n-j) = h(m+i+j) \otimes h(m) = h(m) \otimes h(m+p) = R_h(m),$$

где  $R_h(m)$  - функция корреляции импульсного отклика фильтра. Отсюда:

$$R_y(n) = R_x(n) \otimes R_h(m). \quad (5.1.5')$$

Это означает появление в случайном сигнале на выходе фильтра определенной корреляционной зависимости, определяемой инерционностью фильтра. Эффективный интервал  $\tau_k$  корреляции данных в сигнале тем меньше, чем выше верхняя граничная частота  $\omega_b$  его спектра (по уровню 0.5):

$$\tau_k = \pi/\omega_b = 1/2f_b.$$

Оценка интервала корреляции для конечных (непериодических) функций, как правило, производится непосредственно по функциям автокорреляции  $R(n)$ :

$$\tau_k = 2\sum_n |R(n)/R(0)| - 1, \quad (5.1.6)$$

где значение  $n$  ограничивается величиной 3-5 интервалов спада центрального пика до величины порядка  $0.1 \cdot R(0)$ . Без такого ограничения за счет суммирования модуля флюктуаций, не несущих информации, значение  $\tau_k$  завышается относительно расчетного по спектральной характеристике сигнала. Значение  $\tau_k$  может определяться также непосредственно по координате пересечения нулевой линии функцией автоковариации  $K(n)$ . Дальше обычно начинаются статистические флюктуации значения  $K(n)$  около нулевой линии, вызванные ограниченностью выборки.

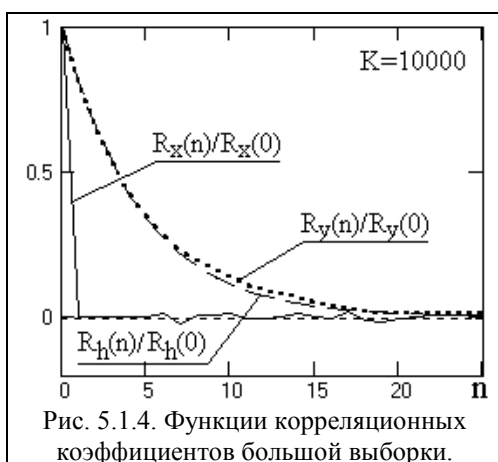


Рис. 5.1.4. Функции корреляционных коэффициентов большой выборки.

Функция  $R_x(n)$  случайных статистически независимых отсчетов близка к  $\delta$ -функции, свертка которой с  $R_h(m)$  приведет к формированию на выходе выходного сигнала, нормированная форма АКФ которого будет стремиться к форме  $R_h(m)$ . При достаточно большой выборке случайных отсчетов входного сигнала это означает практически полное повторение функцией  $R_y(n)$  формы корреляционной функции импульсного отклика, как это можно видеть на рис. 5.1.4, который отличается от рис. 5.1.3 только количеством выборки  $K=10000$ . Соответственно, интервал корреляции выходных сигналов для случайной входной

последовательности можно определять непосредственно по функции (5.1.6) непосредственно импульсного отклика фильтра.

Для взаимной корреляционной функции (ВКФ)  $R_{xy}$  входного и выходного сигналов соответственно имеем:



$$x(k) \otimes y(k+n) = \sum_i h(i) x(k) x(k+n-i).$$

$$R_{xy}(n) = \sum_i h(i) R_x(n-i) \equiv h(i) \otimes R_x(n-i). \quad (5.1.7)$$

т.е. функция взаимной корреляции входного и выходного сигналов равна свертке АКФ входного сигнала с функцией импульсного отклика фильтра. Заключение действительно и для функций ковариации.

Другая взаимно корреляционная функция  $R_{yx}$  может быть получена из соотношения:

$$R_{yx}(n) = R_{xy}(-n) \equiv h(i) \otimes R_x(n+i). \quad (5.1.7')$$

Отметим, что для статистически независимых случайных величин при одностороннем импульсном отклике ( $h(i) = 0$  при  $i < 0$ ) функция  $R_{xy}(n)$  также является односторонней, и равна 0 при  $n < 0$ , а функция  $R_{yx}$  соответственно равна 0 при  $n > 0$ .

## 5.2. СПЕКТРЫ МОЩНОСТИ СЛУЧАЙНЫХ СИГНАЛОВ

**Спектр мощности выходного сигнала.** Если на вход фильтра с импульсным откликом  $h(k) \Leftrightarrow H(f)$  поступает случайный стационарный эргодический сигнал  $x(k) \Leftrightarrow X_T(f)$ , имеющий на интервале  $T$  функцию автокорреляции  $R_x(n)$  и спектр мощности  $W_x(f)$ , то на выходе фильтра регистрируется стационарный эргодический сигнал  $y(k) \Leftrightarrow Y_T(f) = X_T(f)H(f)$ . Соответственно, энергетический спектр выходного сигнала на том же интервале:

$$|Y_T(f)|^2 = |X_T(f)|^2 |H(f)|^2. \quad (5.2.1)$$

Оценка спектра мощности (спектральной плотности энергии):

$$W_y(f) \approx (1/T) |X_T(f)|^2 |H(f)|^2 = W_x(f) |H(f)|^2. \quad (5.2.2)$$

Спектр мощности сигнала на выходе фильтра равен спектру мощности входного сигнала, умноженному на квадрат модуля частотной характеристики фильтра. С учетом четности корреляционных функций спектр мощности выходного сигнала также является четной действительной функцией и не имеет фазовой характеристики процесса.

Спектр мощности сигнала и его функция автокорреляции связаны преобразованием Фурье:

$$R_y(n) \Leftrightarrow |Y(\omega)|^2 = W_y(\omega).$$

**Средняя мощность выходного сигнала** определяется с использованием формулы (5.2.1):

$$W_y = R_y(0) = \int_{-\infty}^{\infty} W_x(f) |H(f)|^2 df \equiv R_x(0) \sum_n h^2(n) = W_x \sum_n h^2(n). \quad (5.2.3)$$

Если значение мощности входного сигнала неизвестно, то вычисляется непосредственно *средний квадрат* значений выходного сигнала:

$$\overline{y^2(t)} = R_y(0) \equiv \overline{x^2} \sum_n h^2(n) \equiv \int_{-\infty}^{\infty} W_x(f) |H(f)|^2 df.$$

Вывод: средняя мощность выходного сигнала равна средней мощности входного сигнала, умноженной на сумму квадратов коэффициентов импульсного отклика фильтра.

**Дисперсия выходного сигнала.** Для центрированных случайных сигналов средняя мощность равна дисперсии сигналов. Для нецентрированных выходных сигналов:

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2 \equiv (\overline{x^2} - \bar{x}^2) \sum_n h^2(n). \quad (5.2.4)$$

**Взаимный спектр мощности** входного и выходного сигнала:

$$W_{xy}(f) \approx (1/T) X_T(f) Y_T(f) = (1/T) |X_T(f)|^2 H(f) = W_x(f) H(f). \quad (5.2.5)$$

Осуществляя преобразование Фурье левой и правой части выражения, получаем:

$$R_{xy}(n) = R_x(n) \otimes h(n), \quad (5.2.6)$$

что повторяет формулу (5.1.5).

**Усиление шумов.** Критерием качества при использовании любого метода фильтрации информации можно считать выполнение целевого назначения с минимальным

усилением шумов (максимальным их подавлением). Обозначим через  $\varepsilon(k)$  аддитивный шум во входном сигнале с математическим ожиданием  $M\{\varepsilon(k)\} = 0$  и дисперсией  $\sigma^2$ . Значения  $\varepsilon(k)$  статистически независимы. С учетом помехи во входном сигнале значение сигнала на выходе:

$$y(k) = \sum_n h(n)[x(k-n) + \varepsilon(k-n)].$$

Математическое ожидание значений выходного сигнала:

$$M\{y(k)\} = \sum_n h(n)[x(k-n) + M\{\varepsilon(k-n)\}] = \sum_n h(n) x(k-n).$$

Вычислим дисперсию распределения отсчетов выходного сигнала:

$$\begin{aligned} D\{y(k)\} &= M\{[\sum_n h(n)[x(k-n) + \varepsilon(k-n)] - M\{y(k)\}]^2\} = \\ &= M\{[\sum_n h(n) \varepsilon(k-n)]^2\} = \sum_n h^2(n) M\{\varepsilon^2(k-n)\} = \sigma^2 \sum_n h^2(n). \end{aligned} \quad (5.2.7)$$

Отсюда следует, что сумма квадратов значений импульсного отклика цифрового фильтра представляет собой коэффициент усиления шумов, равномерно распределенных в главном частотном диапазоне фильтра. Это полностью соответствует прямому использованию выражения (5.2.7) при  $W_x(f) = \sigma^2$ :

$$\sigma_y^2 = \sigma^2 \int_{-\infty}^{\infty} |H(f)|^2 df \equiv \sigma^2 \sum_n h^2(n). \quad (5.2.7')$$

Таким образом, коэффициент усиления фильтром дисперсии статистически распределенных шумов при расчете по импульсному отклику:

$$K_q = \sum_n h^2(n). \quad (5.2.8)$$

По дискретной частотной функции фильтра:

$$K_q = [1/(N+1)] \sum_n H_n^2. \quad (5.2.8')$$

>>> **Пример.** Сглаживающий фильтр:  $y(k) = 0.2 \sum_{k=-2}^2 x(k-n)$ .

>> Коэффициент усиления шумов:  $5(0,2^2) = 0,2$ . Дисперсия шумов уменьшается в  $1/0,2 = 5$  раз.

>> Выполните расчет коэффициента усиления шумов для пятиточечного фильтра МНК.

>> Контрольный ответ: 0.486.

**Функция когерентности** входного и выходного сигналов фильтра оценивается по формуле:

$$\gamma_{xy}^2(f) = |W_{xy}(f)|^2 / [W_x(f) \cdot W_y(f)]. \quad (5.2.9)$$

Если функции  $W_x(f)$  и  $W_y(f)$  отличны от нуля и не содержат дельта-функций, то для всех частот  $f$  значения функции когерентности заключены в интервале:

$$0 \leq \gamma_{xy}^2(f) \leq 1.$$

Для исключения дельта-функции на нулевой частоте (постоянная составляющая сигнала) определение функции когерентности производится по центрированным сигналам. Для фильтров с постоянными параметрами функция когерентности равна 1, в чем нетрудно убедиться, если в формулу (5.2.9) подставить выражения  $W_{xy}$  и  $W_y$ , определенные через  $W_x$ . Для совершенно не связанных сигналов функция когерентности равна нулю. Промежуточные между 0 и 1 значения могут соответствовать трем ситуациям:

1. В сигналах (или в одном из них) присутствует внешний шум (например, шум квантования при ограничении по разрядности).

2. Фильтр не является строго линейным. Это может наблюдаться, например, при определенном ограничении по разрядности вычислений, при накоплении ошибки в рекурсивных системах и т.п.

3. Выходной сигнал  $y(t)$  помимо  $x(t)$  зависит еще от каких-то входных или внутренних системных процессов.

Величина  $1 - \gamma_{xy}^2(f)$  задает долю среднего квадрата сигнала  $y(t)$  на частоте  $f$ , не связанную с сигналом  $x(t)$ .

Использование функций когерентности в практических методах анализа случайных данных подробно рассмотрено в работе /4/.

## Тема 6. ВЕСОВЫЕ ФУНКЦИИ.

Большинство методов анализа и обработки данных имеют в своем составе операцию свертки множества данных  $s(k)$  с функцией оператора свертки  $h(n)$ . Как множество данных  $s(k)$ , так и оператор  $h(n)$ , выполняющий задачу обработки данных и реализующий определенную частотную передаточную функцию системы (фильтра), могут быть бесконечно большими. Практика цифровой обработки имеет дело только с ограниченными множествами и данных, и коэффициентов оператора. В общем случае, эти ограниченные множества "вырезаются" из бесконечных множеств  $s(k)$  и  $h(n)$ , что равносильно их умножению на прямоугольную функцию с единичным амплитудным значением, которую называют естественным временным окном или естественной весовой функцией. Учитывая, что произведение функций отображается в спектральной области сверткой их фурье-образов, это может весьма существенно сказаться как на спектральных характеристиках функций, так и на результатах их последующих преобразований и обработки. Основное назначение весовых функций – сведение к минимуму нежелательных эффектов усечения сигналов.

### 6.1. ЯВЛЕНИЕ ГИББСА

Чаще всего с изменением частотных характеристик функций приходится сталкиваться при усечении операторов фильтров. При расчетах фильтров, как правило, задается определенная передаточная характеристика  $H(\omega)$  фильтра, и по ней производится расчет оператора фильтра  $h(n)$ , количество членов которого может оказаться очень большим, в пределе - бесконечным. Усечение может рассматриваться, как результат умножения функции оператора фильтра на селективирующее весовое окно длиной  $2N+1$ . В простейшем случае это окно представляет собой П-образную селективирующую функцию:

$$h_n = h(n) \Pi_N(n), \quad \Pi_N(n) = 1 \text{ при } |n| \leq N, \quad \Pi_N(n) = 0 \text{ при } |n| > N.$$

Функция  $h(n)$  оператора фильтра обуславливает определенную частотную передаточную характеристику фильтра  $H(\omega)$ . Полному оператору  $h(n)$  соответствует исходная частотная характеристика  $H(\omega)$ :

$$H(\omega) = \sum_{n=-\infty}^{\infty} h(n) \exp(-j\omega n). \quad (6.1.1)$$

**Сущность явления Гиббса.** Усеченной функции  $h_n$  во временном окне селекции  $\Pi_N(n)$  в частотном пространстве соответствует спектральная функция, которая отличается от функции  $H(\omega)$ . Очевидно, что при усечении ряда Фурье (6.1.1), до конечного числа членов  $\pm N$  мы будем иметь усеченный ряд Фурье:

$$H_N(\omega) = \sum_{n=-N}^N h(n) \exp(-j\omega n), \quad (6.1.2)$$

при этом сходимость суммы остающихся членов ряда  $H_N(\omega)$  к исходной передаточной функции  $H(\omega)$  ухудшается. Отклонение частотной характеристики фильтра от первоначально заданной тем больше, чем меньше значение  $N$ . Особенно ярко это проявляется на крутых перепадах (разрывах, скачках) в передаточных функциях:

- крутизна перепадов "размывается", т.к. она не может быть больше, чем крутизна последней сохраненной гармоники ряда (6.1.2);

- по обе стороны "размытых" перепадов появляются выбросы и затухающие осцилляции с частотой, близкой к частоте первого отброшенного члена ряда (6.1.1).

Эти эффекты (см. рис. 6.1.2) при усечении рядов Фурье получили название явления Гиббса.

Рассмотрим явление Гиббса более подробно на примере разложения в ряд Фурье частотной функции единичного скачка  $G(\omega)$ , которая является Фурье-образом какой-то дискретной временной функции  $g_n$ . Уравнение функции единичного скачка:

$$G(\omega) = -0.5, \quad -\pi \leq \omega < 0, \quad G(\omega) = 0.5, \quad 0 \leq \omega \leq \pi, \quad (6.1.3)$$

Функция (6.1.3) имеет разрыв величиной 1 в точке  $\omega = 0$ , и в точках  $\pm \pi, \pm 2\pi, \dots$ , в силу дискретности временной функции и периодичности ее спектра. Поскольку функция  $G(\omega)$  является нечетной, ее ряд Фурье не содержит косинусных членов:

$$g_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\omega) \sin n\omega \, d\omega = \frac{1}{\pi} \int_0^{\pi} \sin n\omega \, d\omega.$$

$$g_n = 2/(n \cdot \pi), \quad n - \text{нечетное},$$

$$g_n = 0, \quad n - \text{четное}.$$

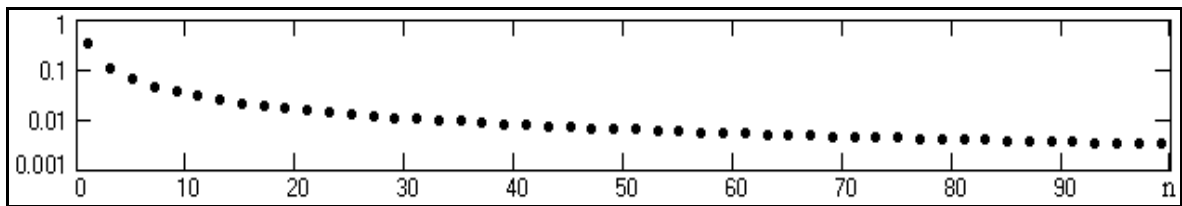


Рис. 6.1.1. Значения коэффициентов  $g_n$ .

Как видно на рис. 6.1.1, ряд коэффициентов  $b_n$  затухает очень медленно. Соответственно, медленно будет затухать и ряд Фурье функции  $G(\omega)$ :

$$G(\omega) = (2/\pi)[\sin \omega + (1/3) \cdot \sin 3\omega + (1/5) \cdot \sin 5\omega + \dots],$$

$$G(\omega) = \frac{2}{\pi} \sum_{n=0}^{\infty} \sin[(2n+1)\omega]/(2n+1). \quad (6.1.4)$$

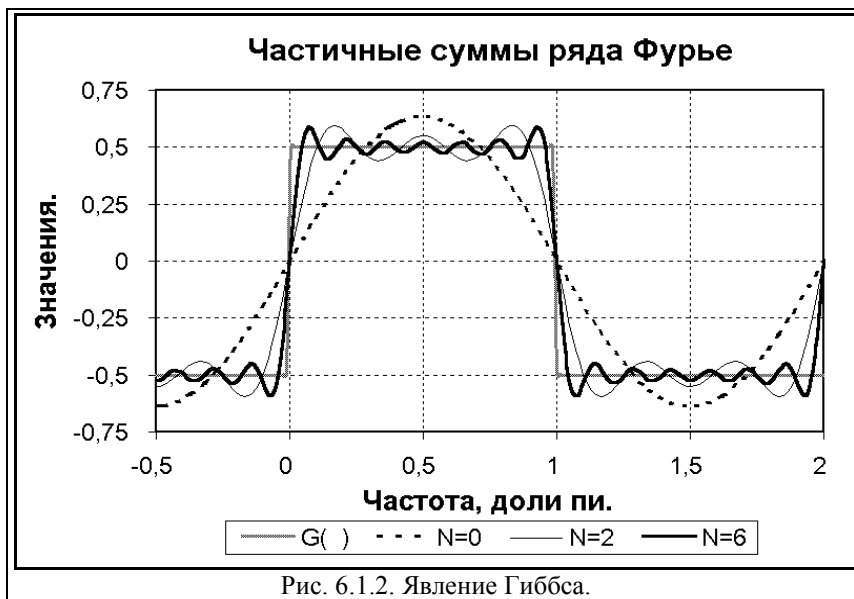


Рис. 6.1.2. Явление Гиббса.

Если мы будем ограничивать количество коэффициентов  $g_n$ , т.е. ограничивать значение  $N$  ряда Фурье функции  $G(\omega)$ , то суммирование в (6.1.4) будет осуществляться не до  $\infty$ , а до значения  $N$ . Графики частичных сумм ряда (6.1.4) в сопоставлении с исходной функцией приведены на рис. 6.1.2. Они наглядно показывают сущность явления Гиббса.

При усечении рядов Фурье определенное искажение функции, разложенной в ряд Фурье, существует всегда. Но при малой доле энергии отсекаемой части сигнала этот эффект может быть и мало заметен. На скачках и разрывах функций он проявляется наиболее выразительно.

**Параметры эффекта.** Ряд (6.1.4) при усечении можно записать в следующем виде:

$$G_N(\omega) = \frac{2}{\pi} \sum_{n=0}^N \left[ \int_0^{\omega} \cos((2n+1)\omega) \, d\omega \right] = \frac{2}{\pi} \int_0^{\omega} \left[ \sum_{n=0}^N \cos((2n+1)\omega) \right] \, d\omega.$$

Сумма косинусного ряда равна  $\sin[2(N+1)\omega]/(2\sin \omega)$ . Отсюда:

$$G_N(\omega) = \frac{1}{\pi} \int_0^{\omega} \frac{\sin 2(N+1)\omega}{\sin \omega} d\omega. \quad (6.1.5)$$

Для определения местоположения максимумов и минимумов осцилляций функции (6.1.5) приравняем к нулю ее первую производную (подынтегральную функцию), при этом:

$$\omega_k = \pm k\pi / (2(N+1)), \quad k = 1, 2, \dots$$

Соответственно, амплитудные значения первых (максимальных) осцилляций функции приходится на точки  $\omega_1 = \pm\pi / (2(N+1))$ , вторых (противоположных по полярности) - на точки  $\omega_2 = \pm\pi / (N+1)$ . Период пульсаций равен  $2\omega_1 = \pi / (N+1) = \Delta\omega$ , т.е. интервалу дискретизации спектра при равном количестве отсчетов оператора фильтра и его спектра. Функция пульсаций (при ее выделении) является нечетной относительно скачка. Соответственно, при скачке функции  $G(\omega)$  на произвольной частоте главного частотного диапазона значения  $\omega_k$  являются значениями относительно положения скачка. Амплитудные значения функции в точках  $\omega_1$  и  $\omega_2$  (при подстановках  $\omega_1$  и  $\omega_2$  верхним пределом в (6.1.5)) практически не зависят от количества членов ряда  $N$  и равны:

$$G_N(\omega_1) \approx 0.5 + 0.09, \quad G_N(\omega_2) \approx 0.5 - 0.05.$$

Амплитуда последующих осцилляций постепенно затухает.

Таким образом, для усеченных рядов Фурье предельные значения максимальных выбросов по обе стороны от скачка и следующих за ними обратных выбросов при единичной амплитуде разрыва функции достигают соответственно 9% и 5% значения амплитуды скачка. Кроме того, сам скачок функции преобразуется в переходную зону, длина которой между точками максимальных выбросов по обе стороны скачка равна  $\pi / (N+1)$ , т.е.  $2\Delta\omega$ , а по уровню исходных значений функции на скачке (в данном случае от -0.5 до 0.5) порядка  $(2/3)\pi / (N+1)$ . Это явление типично для всех функций с разрывом.

Можно рассмотреть это явление и с других позиций. Как известно, произведение функций отображается в частотном представлении сверткой их фурье-образов. Отсюда:

$$h_n = h(n) \Pi_N(n) \Leftrightarrow H(\omega) \otimes \Pi_N(\omega) = H_N(\omega). \quad (6.1.6)$$

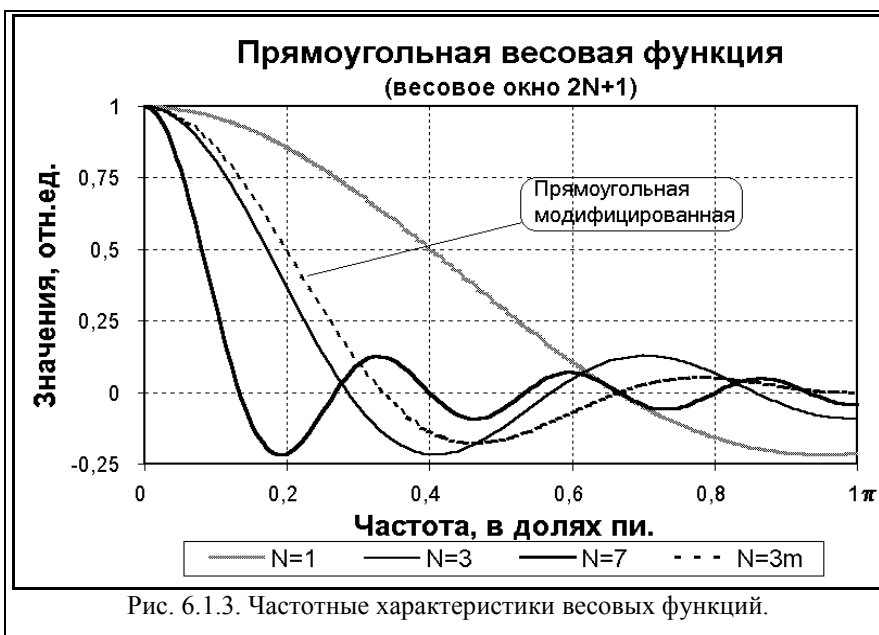


Рис. 6.1.3. Частотные характеристики весовых функций.

Правая часть выражения (6.1.6) и отражает математическую сущность явления Гиббса. Ограничение массива функции определенным количеством членов (умножением на  $\Pi$ -окно, прямоугольную селективную функцию) отображается сверткой частотной характеристики функции с частотной характеристикой селективной

функции. Частотная характеристика прямоугольной функции хорошо известна, как интегральный синус или функция отсчетов  $\text{sinc}(x) = \sin(x)/x$ ,  $x = \omega(2N+1)/2$ , и для  $\Pi$ -импульса длиной  $2N+1$  приведена на рис. 6.1.3 (для ряда значений  $N$ ). Свертка этой частотной функции (Фурье-образа селективной функции) с частотной характеристикой

усекаемых функций и порождает явление Гиббса на резких скачках частотных характеристик. Чем больше  $N$ , тем уже центральный пик спектра прямоугольного импульса, и, соответственно, меньше ширина переходной зоны, которая формируется вместо скачков функций. Амплитуда осцилляций имеет постоянное значение и не зависит от  $N$ .

**Последствия для практики.** При расчетах фильтров и усечении размеров их операторов явление Гиббса является весьма нежелательным, т.к. приводит к искажению формы передаточных характеристик фильтров. В качестве примера рассмотрим явление Гиббса применительно к фильтру низких частот.

Попытаемся реализовать передаточную функцию фильтра следующего вида:

$$H(f) = 1, \text{ при } -0.2 \leq f \leq 0.2, \quad H(f) = 0, \text{ при } -0.2 > f > 0.2,$$

в главном частотном диапазоне от  $-0.5$  до  $0.5$ . Функция четная, коэффициенты ряда Фурье представлены только косинусными членами:

$$a_n = 4 \int_0^{0.2} \cos(2\pi fn) df = 2 \sin(0.4\pi n) / (\pi n).$$

Передаточная функция:

$$H(f) = 0.4 + 2 \sum_{n=1}^{\infty} \sin(0.4\pi n) \cos(2\pi fn) / (\pi n). \quad (6.1.7)$$

Результат усечения ряда Фурье (6.1.7) до  $N = 7$  приведен на рис. 6.1.4.

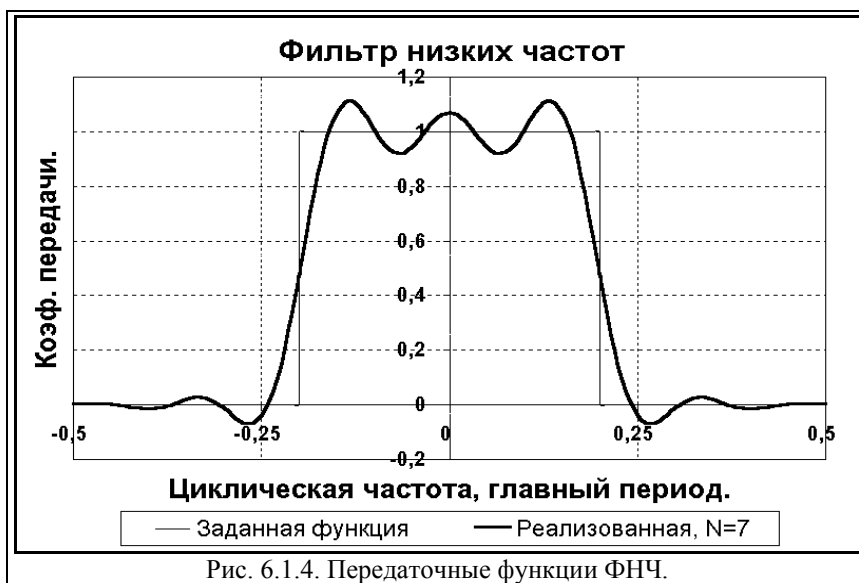


Рис. 6.1.4. Передаточные функции ФНЧ.

Как видно на рисунке, явление Гиббса существенно искажает передаточную функцию фильтра. Однако при реализации фильтров ограничение длины операторов фильтров является правилом их конструирования исходя из чисто практических соображений реализации.

Явление Гиббса имеет место при усечении любых числовых массивов. При обработке геофизических данных

операция усечения числовых массивов, как одномерных, так и многомерных, относится к числу типовых. Вырезаются из профилей и площадей участки съемки с аномальными данными для их более детальной обработки и интерпретации. При анализе усекаются корреляционные функции, и соответственно свертываются с частотным образом весового окна вычисляемые спектры мощности, и пр. Во всех этих случаях мы можем столкнуться как с явлением Гиббса, так и с другими последствиями свертки функций в частотной области, в частности с циклическостью свертки, с определенным сглаживанием спектров усекаемых данных, которое может быть и нежелательным (снижение разрешающей способности), и полезным (повышение устойчивости спектров). В самих усекаемых данных мы не видим этих явлений, т.к. они проявляется в изменении их частотного образа, но при обработке данных, основной целью которой, как правило, и является изменение частотных соотношений в сигналах, последствия этих явлений могут сказаться самым неожиданным образом.

На рис. 6.1.5 показан другой пример искажений сигнала при усечении. Исходный

аналоговый сигнал был вырезан из массива данных на интервале  $k = \{0..60\}$ , дискретизирован и переведен в цифровой форме в спектральную область для обработки. Дискретизация сигнала вызвала периодизацию его спектра, а дискретизация спектра вызвала периодизацию его динамического представления. Но на точках  $k=0$  и  $k=60$  в периодическом повторении исходного сигнала при усечении образовался скачок функции с бесконечным частотным спектром, а главный диапазон спектра дискретизированного сигнала ограничен интервалом его дискретизации ( $\omega_N = \pi/\Delta t$ ). Следовательно, спектр сигнала является искаженным за счет наложения спектров боковых периодов на главный диапазон, а при восстановлении аналогового сигнала по спектру главного диапазона он восстанавливается из искаженного спектра. Это приводит к появлению явления Гиббса на обоих концах вырезанного сигнала, что наглядно видно на рис. 6.1.5.

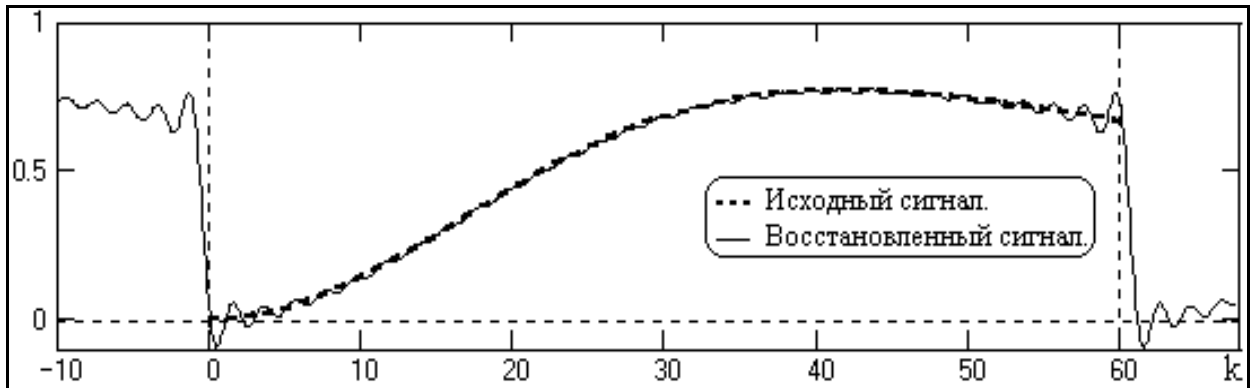


Рис. 6.1.5.

Практически это означает, что при частотной обработке вырезанного сигнала будет обрабатываться не спектр исходного сигнала, а спектр, которому соответствует сигнал, восстанавливаемый по данному спектру с наложенным явлением Гиббса.

## 6.2. ВЕСОВЫЕ ФУНКЦИИ

Естественным методом нейтрализации нежелательных эффектов усечения сигналов является изменение окна селекции сигнала таким образом, чтобы частотная характеристика окна селекции при свертке как можно меньше искажала спектр сигнала. Что последнее возможно, показывает даже такая простая модификация прямоугольной функции, как уменьшение в два раза значений ее крайних членов. Фурье-образ модифицированной П-функции рассматривался нами в составе сглаживающих фильтров МНК 1-го порядка. Он отличается от обычной П-функции с тем же размером окна выходом в ноль на частоте Найквиста и несколько меньшей амплитудой осцилляций при небольшом расширении главного максимума.

**Нейтрализация явления Гиббса** в частотной области. Рассмотрение продолжим с формулы (6.1.2) при усечении произвольного оператора фильтра  $h(n)$  прямоугольным селектирующим окном  $\Pi_N(n)$ . Период осцилляций суммы усеченного ряда Фурье (6.1.2) примерно равен периоду первого отброшенного члена ряда. С учетом этого фактора осцилляции частотной характеристики могут быть существенно сглажены путем усреднения в скользящем окне по периоду осцилляций, т.е. при нормированной свертке оператора  $H_N(\omega)$  с  $\Pi_r(\omega)$  — импульсом, длина которого равна периоду осцилляций  $r = 2\pi/(N+1)$ . Эта свертка отобразится во временной области умножением коэффициентов фильтра  $h(n)$  на множители, которые являются коэффициентами преобразования Фурье частотной П-образной сглаживающей функции  $\Pi_r(\omega)$ :

$$H'_N(\omega) = H_N(\omega) \otimes \Pi_r(\omega) \Leftrightarrow h_n \sigma_N(n) = h(n) \Pi_N(n) \sigma_N(n),$$

$$p(n) = \Pi_N(n) \sigma_N(n) = \text{sinc}(\pi n/(N+1)), \quad |n| \leq N. \quad (6.2.1)$$

Эта операция носит название сглаживания Ланцоша. Произведение  $\Pi_N(n) \sigma_N(n) \equiv \sigma_N(n)$  представляет собой новое весовое окно селекции  $p(n)$  взамен прямоугольного окна.

Функцию  $\sigma_N(n)$  обычно называют временной весовой функцией (окном). Вид и частотная характеристика весового окна Ланцоша в сопоставлении с прямоугольным окном приведены на рис. 6.2.1.

Как видно на рисунке, частотная характеристика весовой функции Ланцоша по сравнению с П-образной функцией имеет почти в 4 раза меньшую амплитуду осцилляций, но при этом ширина главного максимума увеличилась примерно на четверть. Отметим, однако, что если амплитуда осцилляций (в единицах амплитуды главного максимума) определяется выбранным типом весовой функции, то ширина главного максимума, которой определяется ширина переходной зоны (вместо скачка функции), зависит от размеров весового окна и соответственно может изменяться под поставленные условия (уменьшаться увеличением размера  $2N+1$  весового окна).

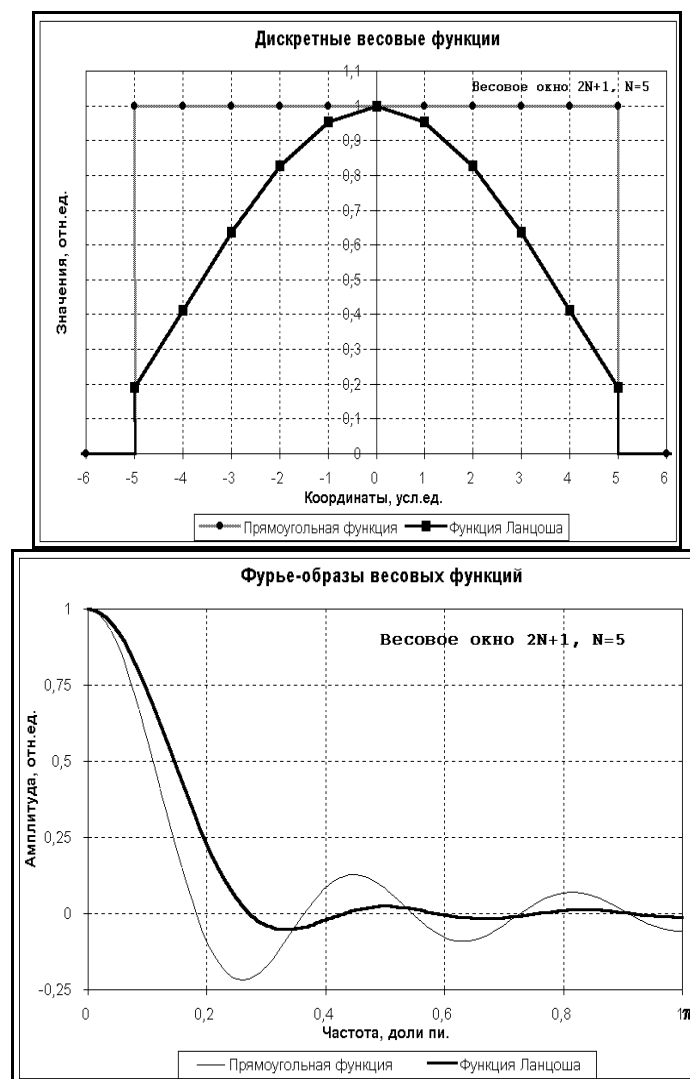


Рис. 6.2.1. Весовая функция Ланцоша.

**Основные весовые функции.** В настоящее время известны десятки различных по эффективности весовых функций. В идеальном случае хотелось бы иметь весовую функцию с минимальной амплитудой осцилляций, высокую и узкую в главном максимуме.

В таблицах 6.2.1 и 6.2.2 приведены формулы и основные спектральные характеристики наиболее распространенных и часто используемых весовых окон. Носители весовых функций, в принципе, являются неограниченными и при использовании в качестве весовых окон действуют только в пределах окна и обнуляются за его пределами (как и в (6.2.1)), что выполняется без дальнейших пояснений. Для упрощения записи формулы приводятся в аналитической, а не в дискретной форме, с временным окном  $2\tau$ ,



симметричным относительно нуля ( $0 \pm \tau$ ). При переходе к дискретной форме окно  $2\tau$  заменяется окном  $2N+1$  (полное количество точек дискретизации выделяемой сигнальной функции), а значения  $t$  - номерами отсчетов  $n$  ( $t = n\Delta t$ ). Следует заметить, что большинство весовых функций на границах окна ( $n = \pm N$ ) принимают нулевые или близкие к нулевым значения, т.е. фактическое окно усечения данных занижается на 2 точки. Последнее исключается, если принять  $2\tau = (2N+3)\Delta t$ .

Таблица 6.2.1.

Основные весовые функции

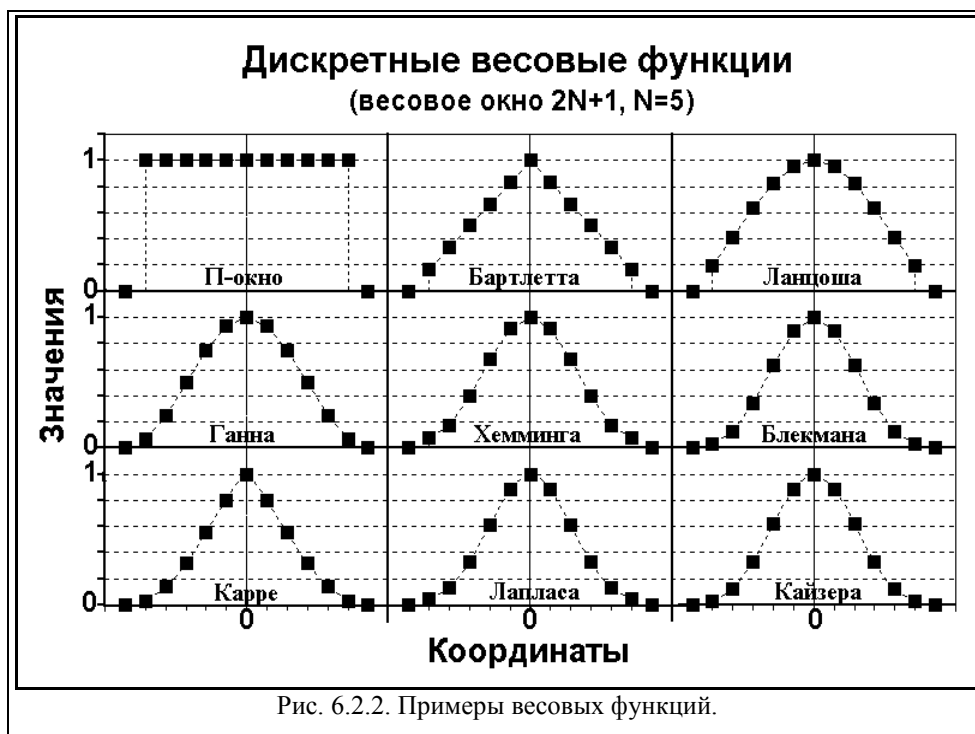
Временное окно	Весовая функция	Фурье-образ
Естественное (П)	$\Pi(t) = 1,  t  \leq \tau; \Pi(t) = 0,  t  > \tau$	$\Pi(\omega) = 2\tau \operatorname{sinc}[\omega\tau]$
Бартлетта (Δ)	$b(t) = 1 -  t /\tau$	$B(\omega) = \tau \operatorname{sinc}^2(\omega\tau/2)$ .
Хеннинга, Ганна	$p(t) = 0.5[1 + \cos(\pi t/\tau)]$	$0.5\Pi(\omega) + 0.25\Pi(\omega + \pi/\tau) + 0.25\Pi(\omega - \pi/\tau)$
Хемминга	$p(t) = 0.54 + 0.46 \cos(\pi t/\tau)$	$0.54\Pi(\omega) + 0.23\Pi(\omega + \pi/\tau) + 0.23\Pi(\omega - \pi/\tau)$
Карре (2-е окно)	$p(t) = b(t) \operatorname{sinc}(\pi t/\tau)$	$\tau \cdot B(\omega) * \Pi(\omega), \Pi(\omega) = 1$ при $ \omega  < \pi/\tau$
Лапласа-Гаусса	$p(t) = \exp[-\beta^2(t/\tau)^2/2]$	$[(\tau/\beta) \sqrt{2\pi} \exp(-\tau^2\omega^2/(2\beta^2))] * \Pi(\omega)$
Кайзера-Бесселя	$p(t) = \frac{J_0[\beta\sqrt{1-(t/\tau)^2}]}{J_0[\beta]}$ , $J_0[x] = \sum_{k=1}^{\infty} [(x/2)^k/k!]^2$	Вычисляется преобразованием Фурье. $J_0[x]$ - модифицированная функция Бесселя нулевого порядка

Таблица 6.2.2.

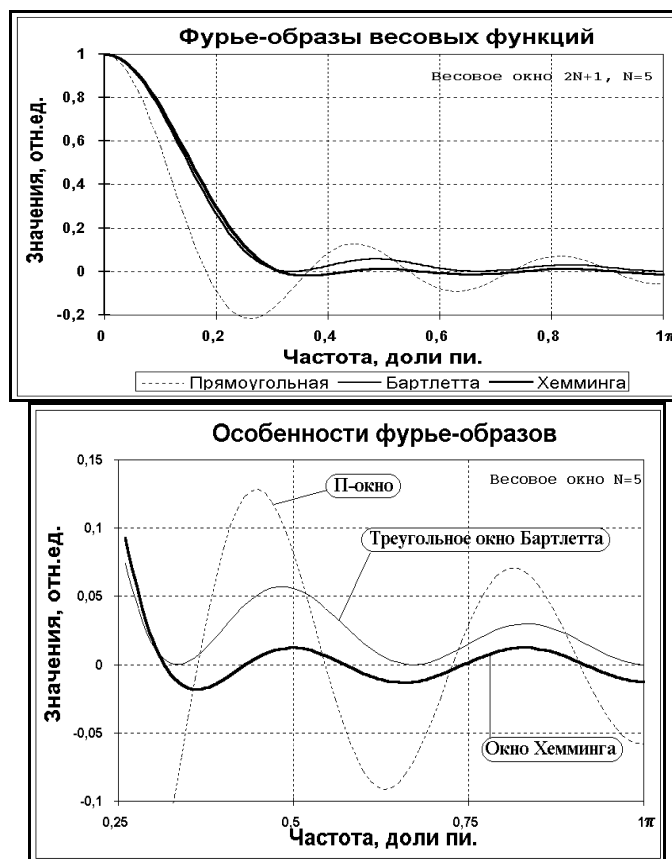
Характеристики спектров весовых функций

Параметры	Ед. изм.	П-окно	Бартлетт	Ланцош	Хеннинг	Хемминг	Карре	Лаплас	Кайзер
Амплитуда:									
Главный пик	$\tau$	2	1	1.18	1	1.08	0.77	0.83	0.82
1-й выброс(-)	%Гл.п.	0.217	-	0.048	0.027	0.0062	-	0.0016	.00045
2-й выброс(+)	- " -	0.128	0.047	0.020	0.0084	0.0016	-	0.0014	.00028
Ширина Гл. пика	$\omega\tau/2\pi$	0.60	0.89	0.87	1.00	0.91	1.12	1.12	1.15
Положения:									
1-й нуль	$\omega\tau/2\pi$	0.50	1.00	0.82	1.00	1.00	-	1.74	1.52
1-й выброс	$\omega\tau/2\pi$	0.72	-	1.00	1.19	1.09	-	1.91	1.59
2-й нуль	$\omega\tau/2\pi$	1.00	-	1.29	1.50	1.30	-	2.10	1.74
2-й выброс	$\omega\tau/2\pi$	1.22	1.44	1.50	1.72	1.41	-	2.34	1.88

Сравнительный вид весовых функций приведен на рис. 6.2.2. Расчет функций проведен с исключением нулевых значений на границах весового окна.



Спектральные окна Бартлетта и Карре не имеют отрицательных выбросов и применяются, в основном, для усечения корреляционных функций. Функция Карре не имеет нулей и представляет собой положительно убывающую функцию. Функции Хеннинга и Хемминга примерно одного класса, функция Хемминга является улучшенным вариантом функции Хеннинга. Частотные образы функций Бартлетта и Хемминга приведены на рис. 6.2.3.



Весовые окна Лапласа и Кайзера - усеченные функции соответственно Гаусса и Бесселя. Степень усечения зависит от параметра  $\beta$ . Характеристики функций, приведенные в таблице 6.2.2, действительны при  $\beta=3$  для окна Лапласа и  $\beta=9$  для окна Кайзера. При уменьшении значения  $\beta$  крутизна главного максимума сглаживающих функций увеличивается (ширина пика уменьшается), но платой за это является увеличение амплитуды осцилляций.

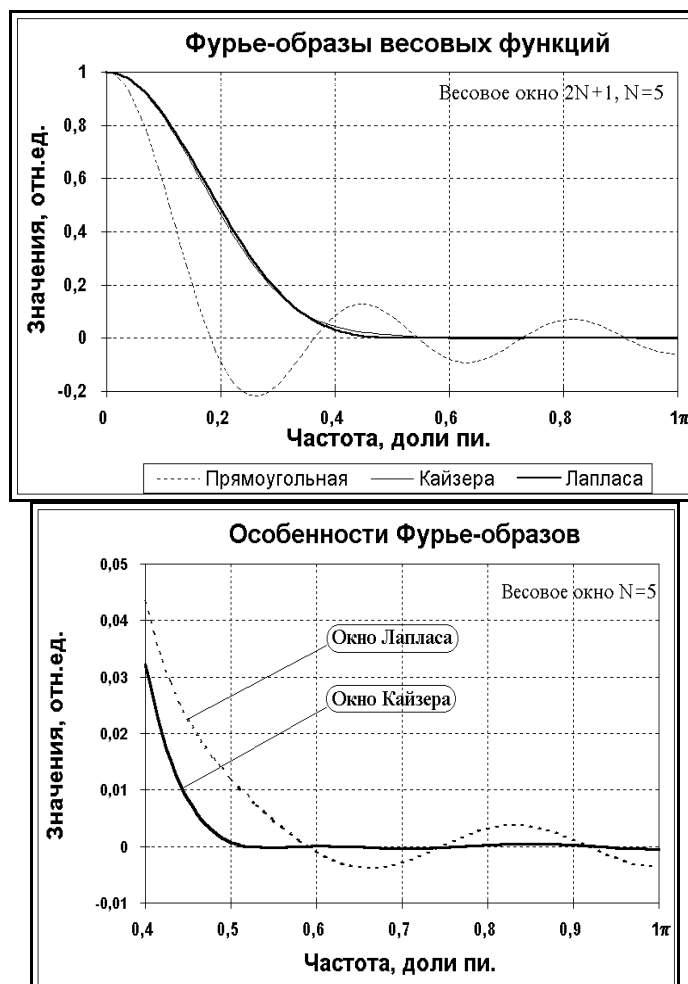


Рис. 6.2.4. Частотные функции весовых окон.

Функции Лапласа и Кайзера являются универсальными функциями. По-существу, их можно отнести к числу двухпараметровых: размером окна  $2\tau$  (числом  $2N+1$ ) может устанавливаться ширина главного максимума, а значением коэффициента  $\beta$  - относительная величина осцилляций на частотной характеристике весовых функций, причем, вплоть до осцилляций П-окна при  $\beta=0$ . Это обусловило их широкое использование, особенно при синтезе операторов фильтров.

Попутно заметим, что достаточно гладкие частотные характеристики весовых функций позволяют использовать их в качестве сглаживающих низкочастотных НЦФ.

## Тема 7. НЕРЕКУРСИВНЫЕ ЧАСТОТНЫЕ ЦИФРОВЫЕ ФИЛЬТРЫ

Нерекурсивные фильтры реализуют алгоритм свертки двух функций:  $y_k = h_n \otimes x_{k-n}$ , где  $x_k$  - массив входных данных фильтра,  $h_n$  - оператор (ядро, импульсный отклик) фильтра,  $k$  и  $n$  - нумерация числовых значений массива данных и числовых значений коэффициентов фильтра,  $k = 0, 1, 2, \dots, K$ ;  $n = 0, 1, 2, \dots, N$ ;  $K \geq N$ . Значения выходных отсчетов свертки  $y_k$  для любого аргумента  $k$  определяются текущим и "прошлыми" (до  $k$ -

N) значениями входных отсчетов. Такой фильтр называется нерекурсивным цифровым фильтром (НЦФ). Интервал [0-N] оператора получил название "окна" фильтра. Окно фильтра составляет N+1 отсчет, фильтр является односторонним каузальным, т.е. причинно обусловленным текущими и "прошлыми" значениями входного сигнала, и выходной сигнал не опережает входного. В общем случае, каузальный фильтр меняет в спектре сигнала состав гармоник, их амплитуды и фазы.

Каузальный фильтр может быть реализован физически в реальном масштабе времени. Начало фильтрации возможно только при задании определенных начальных условий – N значений отсчетов для точек  $x(k-n)$  при  $k < n$ . Как правило, в качестве начальных условий принимаются нулевые значения, тренд сигнала или значения отсчета  $x(0)$ , т.е. продление отсчета  $x(0)$  назад по аргументу.

При обработке данных на ЭВМ ограничение по каузальности снимается. В программном распоряжении фильтра могут находиться как "прошлые", так и "будущие" ( $k+n$ , до  $k+N'$ ) значения входной последовательности отсчетов относительно текущей точки вычислений  $k$ , при этом для завершения свертки (аналогично началу) требуется  $N'$  точек конечных условий при  $(k+n) > K$ . При  $N' = N$  и  $h(-n) = h(n)$  фильтр называется двусторонним симметричным фильтром. Симметричные фильтры, в отличие от односторонних, не изменяют фазы обрабатываемого сигнала.

### 7.1. ОБЩИЕ СВЕДЕНИЯ.

Основное свойство любого фильтра – его *частотная (frequency response) и фазовая характеристики*. Они показывают, какое влияние фильтр оказывает на амплитуду и фазу различных гармоник обрабатываемого сигнала.

К наиболее известным типам нерекурсивных цифровых фильтров (НЦФ) относятся частотные фильтры, алгоритм которых для симметричных НЦФ, не изменяющих фазу сигналов, имеет вид:

$$y_k = \sum_{n=-N}^N h_n s_{k-n}.$$

**Типы фильтров.** В зависимости от вида частотной характеристики выделяют три основных группы частотных фильтров: ФНЧ - фильтры низких частот (*low-pass filters*) - пропускание низких и подавление высоких частот во входном сигнале, ФВЧ - фильтры высоких частот (*high-pass filters*) - пропускание высоких и подавление низких частот, и ПФ - полосовые фильтры, которые пропускают (*band-pass filters*) или подавляют (*band-reject filters*) сигнал в определенной частотной полосе. Среди последних в отдельную группу иногда выделяют РФ - режекторные фильтры, понимая под ними фильтры с подавлением определенной гармоники во входном сигнале, и СФ – селекторные фильтры, обратные РФ.

Если речь идет о подавлении определенной полосы частот во входном сигнале, то такие фильтры называют заградительными. Ни теоретического, ни практического интереса к методам их расчета обычно не проявляется, так как их частотная характеристика обычно задается инверсией характеристики полосового фильтра ( $1-H_n(\omega)$ ) и каких-либо дополнительных особенностей при своем проектировании не имеет.

Схематические частотные характеристики фильтров приведены на рисунке 7.1.1. Между частотными интервалами пропускания и подавления сигнала существует зона, которая называется переходной. Ширина переходной зоны определяет резкость характеристики фильтра. В этой зоне амплитудная характеристика монотонно уменьшается (или увеличивается) от полосы пропускания до полосы подавления (или наоборот).

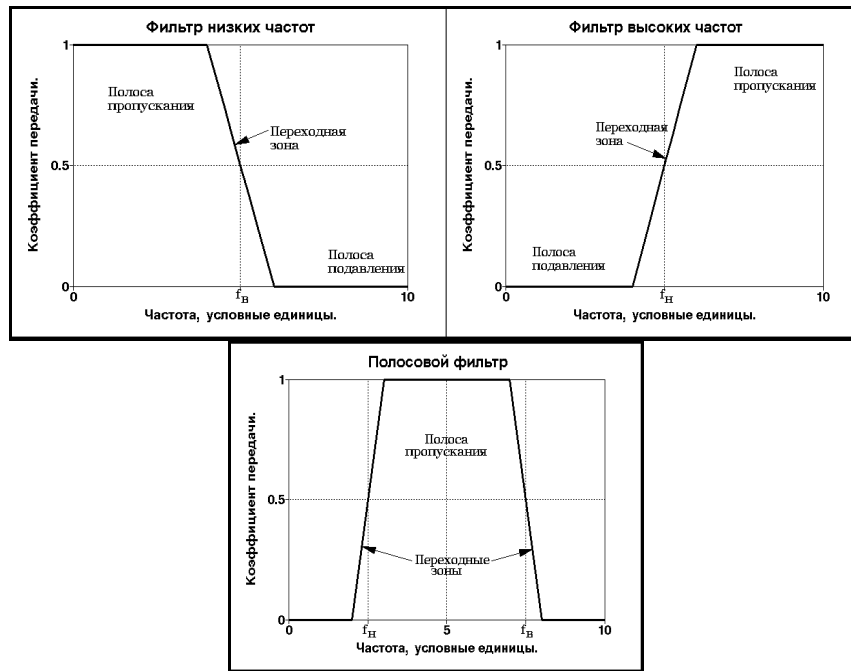


Рис. 7.1.1. Типы основных частотных фильтров.

Практика проектирования цифровых фильтров базируется, в основном, на синтезе фильтров низких частот. Все другие виды фильтров могут быть получены из фильтров низких частот соответствующим преобразованием.

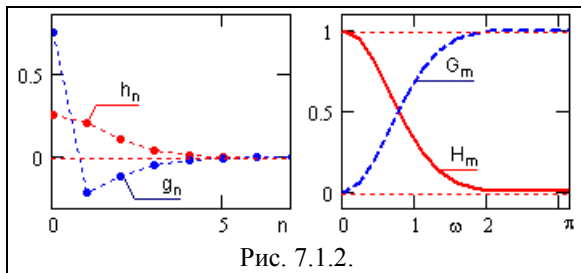


Рис. 7.1.2.

Так, например, фильтр высоких частот  $g(n)$  может быть получен инверсией фильтра низких частот  $h(n)$  - вычислением разности между исходным сигналом и результатом его фильтрации низкочастотным НЦФ:

$$y(k) = s(k) - \sum_{n=-N}^N h(n) s(k-n).$$

Отсюда, условие инверсии симметричного низкочастотного фильтра в высокочастотный:

$$g(0) = 1-h(0), \quad g(n) = -h(n) \quad \text{при } n \neq 0.$$

Пример обращения и спектры фильтров приведены на рис. 7.1.2 (в правой части главных диапазонов).

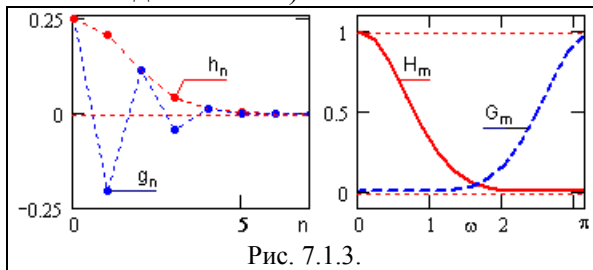


Рис. 7.1.3.

Применяется также способ получения фильтров высоких частот из низкочастотных фильтров путем реверса частоты в передаточной функции низкочастотного фильтра, т.е. заменой переменной  $\omega$  на переменную  $\omega' = \pi - \omega$  (при  $\Delta t = 1$ ). Для симметричных фильтров, содержащих в передаточной функции только косинусные члены аргумента  $\omega$ , в результате такой операции будем иметь:

$$\cos n(\pi - \omega) = \cos n\pi \cos n\omega = (-1)^n \cos n\omega.$$

Последнее означает смену знака всех нечетных гармоник передаточной характеристики фильтра и, соответственно, всех нечетных членов фильтра:

$$g(n) = h(n) \quad \text{при } n = \pm 1, \pm 3, \dots$$

Пример частотного реверса приведен на рис. 7.1.3. Физическую сущность такой операции инверсии спектра легко понять на постоянной составляющей сигнала. При изменении на противоположный знака каждого второго отсчета постоянной величины это

постоянной значение превращается в "пилу", частота которой равна частоте Найквиста главного частотного диапазона (отсчеты по амплитудным значениям этой частоты), равно как и наоборот, отсчеты гармоника сигнала на частоте Найквиста (знакопередающиеся в силу сдвига по интервалам дискретизации на  $\pi$ ) превращаются в постоянную составляющую.

Полосовой фильтр может реализоваться последовательным применением ФНЧ и ФВЧ с соответствующим перекрытием частот пропускания. В математическом представлении это означает последовательную свертку массива данных с массивами коэффициентов  $h_n$  - низкочастотного, и  $h_b$  - высокочастотного фильтров:

$$v_k = h_n(n) \otimes s(k-n), \quad y_k = h_b(n) \otimes v_k = h_n(n) \otimes h_b(n) \otimes s(k-n).$$

Так как операция свертки коммутативна, то вместо отдельных массивов коэффициентов ФНЧ и ФВЧ их сверткой может быть определен непосредственно массив коэффициентов полосового фильтра:  $h_n = h_n(n) \otimes h_b(n)$ .

Полосовой режекторный фильтр также может быть получен методом инверсии полосового фильтра. Одночастотные режекторные фильтры обычно выполняются на основе простых рекурсивных цифровых фильтров, более эффективных для данных целей.

Часто к фильтрам предъявляются более сложные требования. Например, фильтр может иметь несколько частотных полос пропускания с разными коэффициентами усиления, а для полос непропускания могут быть заданы разные коэффициенты подавления. Иногда требуемая частотная характеристика фильтра задается вообще произвольной кривой.

**Методика расчетов НЦФ.** Обычно при фильтрации сигналов задается требуемая частотная характеристика фильтра. Задачей является построить фильтр, отвечающий заданным требованиям и провести фильтрацию. Зачастую бывает невозможно построить в точности заданный фильтр, и выполняется фильтр, близкий по характеристикам к заданному.

Существует много способов построения фильтров с заданной частотной характеристикой. Наиболее простой из них – проектирование фильтров с линейной фазой с помощью весовых окон. Этот способ является универсальным и позволяет получить фильтр с любой заданной частотной характеристикой. Отметим, однако, что с помощью других, математически более строгих и совершенных методов, иногда удается построить фильтр меньшей длины, удовлетворяющий тем же требованиям к частотной характеристике.

Наиболее простой является методика расчетов программных двусторонних симметричных фильтров без изменения фазы выходного сигнала относительно входного. В самом общем виде она включает:

1. Задание идеальной амплитудно-частотной характеристики передаточной функции фильтра. Термин идеальности понимается здесь в том смысле, что на характеристике указываются полосы пропускания и подавления частот с коэффициентами передачи 1 и 0 соответственно без переходных зон.

2. Расчет функции импульсного отклика идеального фильтра (обратное преобразование Фурье частотной характеристики фильтра). При наличии скачков функций на границах пропускания/подавления импульсный отклик содержит бесконечно большое количество членов.

3. Ограничение функции отклика до определенного количества членов, при этом на передаточной характеристике фильтра возникает явление Гиббса – осцилляции частотной характеристики с центрами на скачках.

4. Для нейтрализации явления Гиббса производится выбор весовой функции и расчет ее коэффициентов, на которые умножаются коэффициенты функции отклика фильтра. Результатом данной операции являются значения коэффициентов оператора фильтра (рабочий импульсный отклик фильтра). По существу, операции 3 и 4

представляют собой усечение ряда Фурье динамического (временного) представления передаточной функции фильтра определенной весовой функцией (умножение на весовую функцию).

5. С использованием полученных значений коэффициентов оператора фильтра производится построение его частотной характеристики и проверяется ее соответствие поставленной задаче.

При проектировании симметричных нерекурсивных фильтров нет необходимости базироваться на расчете фильтров низких частот с последующим их преобразованием, при необходимости, в фильтры верхних частот или полосовые фильтры. Расчет непосредственно полосового фильтра достаточно прост, а НЧ- и ВЧ-фильтры являются частным случаем полосового фильтра с одной верхней или одной нижней граничной частотой.

**Фильтры с линейной фазовой характеристикой.** Несколько сложнее расчет каузальных (односторонних) частотных фильтров, для которых требуется обеспечить линейность фазово-частотной характеристики для исключения изменения гармонии сочетания частотных составляющих сигнала на его выходе по отношению к входу. Чтобы фильтр имел линейную фазовую характеристику необходимо обеспечить выполнение условия:

$$\varphi(\omega) = \alpha\omega. \quad (7.1.1)$$

Оно выполняется, если импульсная характеристика фильтра имеет положительную симметрию:

$$h(n) = h(N-n-1), \quad n = 0, 1, 2, \dots, (N-1)/2, \quad N - \text{нечетное (тип 1);}$$

$$n = 0, 1, 2, \dots, (N/2)-1, \quad N - \text{четное (тип 2).}$$

При этом фазовая характеристика будет определяться длиной фильтра:

$$\alpha = (N-1)/2.$$

Частотная характеристика фильтра:

$$H(\omega) = |H(\omega)| \exp(j\varphi(\omega)), \quad (7.1.2)$$

где модуль  $|H(\omega)|$  задается аналогично АЧХ симметричных фильтров. Следует также учитывать, что частотную характеристику типа 2 нельзя использовать для проектирования фильтров верхних частот, т.к. она всегда равна нулю на частоте Найквиста.

Собственно методика расчета каузальных фильтров, за исключением использования (7.1.2) для задания частотной характеристики, не отличается от методики расчета симметричных фильтров, включая необходимость использования весовых функций для нейтрализации явления Гиббса. Это позволяет применять чисто практический метод расчетов – вычислить и отработать сначала симметричный фильтр на  $N$ -точек (тип 1), а затем превратить его в каузальный сдвигом вправо на  $(N-1)/2$  точек в область только положительных значений  $n \geq 0$ .

## 7.2. ИДЕАЛЬНЫЕ ЧАСТОТНЫЕ ФИЛЬТРЫ.

Идеальным полосовым фильтром называется фильтр, имеющий единичную амплитудно-частотную характеристику в полосе от определенной нижней частоты  $\omega_n$  до определенной верхней частоты  $\omega_v$ , и нулевой коэффициент передачи за пределами этой полосы (для цифровых фильтров - в главном частотном диапазоне).

**Импульсная реакция фильтра** (коэффициенты оператора) находится обратным преобразованием Фурье заданной передаточной функции  $H(\omega)$ . В общем случае:

$$h(n\Delta t) = (1/2\pi) \int_{-\infty}^{\infty} H(\omega) \exp(j\omega n\Delta t) d\omega.$$

Для получения вещественной функции импульсного отклика фильтра действительная часть передаточной функции должна быть четной, а мнимая - нечетной. Цифровые фильтры задаются в главном частотном диапазоне, границы которого (частота Найквиста  $\pm \omega_N$ ) определяются интервалом дискретизации данных ( $\omega_N = \pi/\Delta t$ ), подлежащих фильтрации, и соответственно определяют интервал дискретизации

оператора фильтра ( $\Delta t = \pi/\omega_N$ ). Для фильтров с нулевым фазовым сдвигом мнимая часть передаточной функции должна быть равна нулю, при этом оператор фильтра определяется косинусным преобразованием Фурье:

$$h(n\Delta t) = (1/\pi) \int_0^{\omega_N} H(\omega) \cos(n\pi\omega/\omega_N) d\omega, \quad n = 0, 1, 2, \dots \quad (7.2.1)$$

Для идеального полосового фильтра  $H(\omega)=1$  в полосе частот от  $\omega_H$  до  $\omega_B$ , и интеграл (7.2.1) вычисляется в этих пределах. Идеальные фильтры низких и высоких частот, как частные случаи идеальных ПФ, интегрируются в диапазоне от 0 до  $\omega_B$  для низкочастотного и от  $\omega_H$  до  $\omega_N$  для высокочастотного фильтра.

При интервале дискретизации данных  $\Delta t$ , условно принимаемым за 1, главный частотный диапазон передаточных функций ограничивается значением частоты Найквиста от  $-\pi$  до  $\pi$ . Если на практике интервал дискретизации данных в физических единицах отличается от 1, то это сказывается только на изменении масштаба частотной шкалы передаточных функций.

**Пример 1.**  $\Delta t = 0.1$  сек.  $f_N = 1/2\Delta t = 5$  Гц.  $\omega_N = \pi/\Delta t = 10 \pi$ .

**Пример 2.**  $\Delta x = 10$  метров.  $f_N = 0.05 \text{ м}^{-1}$ .  $\omega_N = 0.1 \pi$ .

Во всех дальнейших выражениях значение  $\Delta t$ , если это специально не оговорено, будем принимать равным 1.

При  $H(\omega)=A=1$  в полосе пропускания ( $\omega_H, \omega_B$ ), и  $H(\omega)=0$  за ее пределами, для идеальных симметричных полосовых НЦФ из (7.2.1) с границами интегрирования, соответственно, от  $\omega_H$  до  $\omega_B$  в общем виде получаем:

$$h(n) = (A/\pi) [\omega_B \text{sinc}(n\omega_B) - \omega_H \text{sinc}(n\omega_H)], \quad (7.2.2)$$

$$h_0 = (\omega_B - \omega_H)/\pi, \quad h(n) = (\sin n\omega_B - \sin n\omega_H)/(n\pi).$$

где  $\text{sinc}(n\omega) = \sin(n\omega)/(n\omega)$  - функция интегрального синуса (функция отсчетов), бесконечная по координате  $\omega$ .

При инверсии частотной характеристики в заградительный фильтр:

$$h_0 = (1 - (\omega_B - \omega_H))/\pi, \quad h(n) = (\sin n\omega_H - \sin n\omega_B)/(n\pi).$$

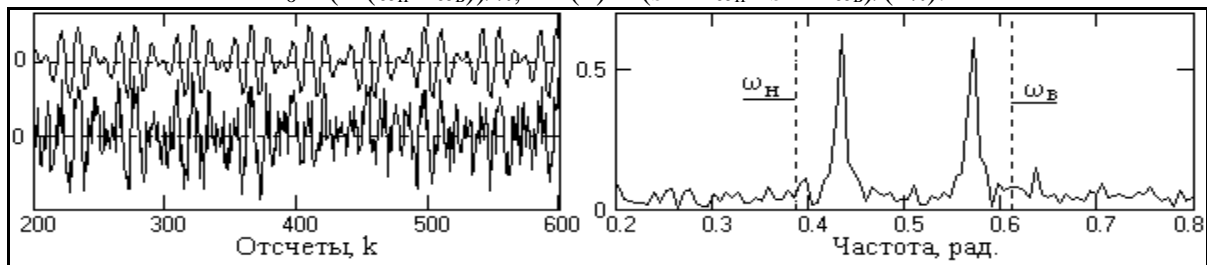


Рис. 7.2.1. Входные сигналы.

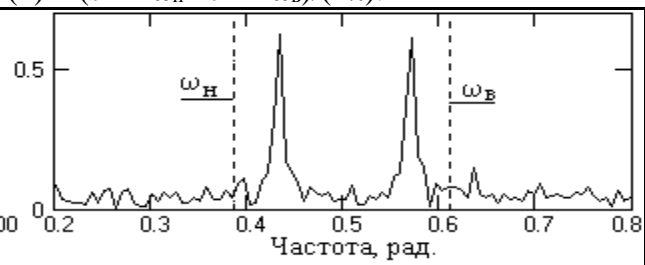
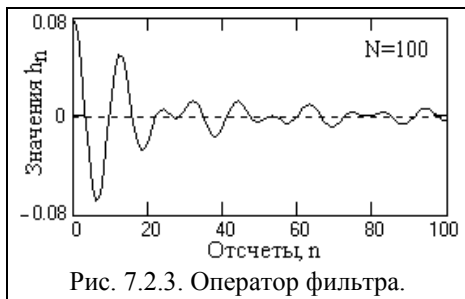


Рис. 7.2.2. Спектр сигнала и границы фильтра.

На рис. 7.2.1 приведен пример сигнала однотональной балансированной амплитудной модуляции (чистого сигнала – сверху, с наложенными шумами внизу, мощность шумов равна мощности сигнала). Если информация заключена в частоте и амплитуде модулирующего сигнала, то полосовой фильтр выделения сигнала из шумов, спектр которого для одной модулирующей частоты приведен на рис. 7.2.2, в идеальном случае должен иметь плоскую частотную характеристику в границах возможных вариаций модулирующей частоты (от  $\omega_H$  до  $\omega_B$ ).

Размер оператора фильтра определяется приблизительно из следующих соображений. Чем больше размер оператора, тем круче будет переходная зона и меньше ее размер, т.е. тем ближе будет фактически реализованная передаточная функция фильтра к идеальной. Обычно сначала стоит попробовать построить фильтр достаточно большого размера, оценить его соответствие заданной частотной характеристике и в дальнейшем попытаться уменьшить. Значение  $N$  для симметричных НЦФ должно быть нечетным числом.





На рис. 7.2.3 приведен оператор полосового фильтра, вычисленный по (7.2.2) для приведенных выше условий, с ограничением по числу коэффициентов оператора до  $N=100$ . Как видно из рисунка, оператор затухает достаточно медленно и явно усечен, что должно сказаться на форме частотной характеристики фильтра. Все дальнейшие вычисления будут проводиться на продолжении данного примера.

### 7.3. КОНЕЧНЫЕ ПРИБЛИЖЕНИЯ ИДЕАЛЬНЫХ ФИЛЬТРОВ

Оператор идеального частотного НЦФ, как это следует из выражения (7.2.2), представляет собой бесконечную затухающую числовую последовательность, реализующую заданную передаточную функцию:

$$H(\omega) = \sum_{n=-\infty}^{\infty} h(n) \cos n\omega. \quad (7.3.1)$$

**Ограничение окна операторов фильтров.** На практике бесконечный ряд (7.3.1) всегда приходится ограничивать определенным числом членов его конечного приближения

$$H'(\omega) = \sum_{n=-N}^N h(n) \cos n\omega,$$

при этом передаточная функция осложняется явлением Гиббса, и появляется переходная зона между полосами пропускания и подавления сигнала (рис. 7.3.1, пунктирная кривая при  $N=100$ ). Явление Гиббса формирует первые выбросы передаточной функции на расстоянии  $\pi/(2(N+1))$  от скачков (разрывов первого рода). Если ширину переходной зоны  $\Delta_p$  в первом приближении принять по расстоянию между первыми выбросами по обе стороны от скачка функции  $H(\omega)$ , то ее значение будет ориентировочно равно  $\pi/(N+1) = \Delta_p$ .

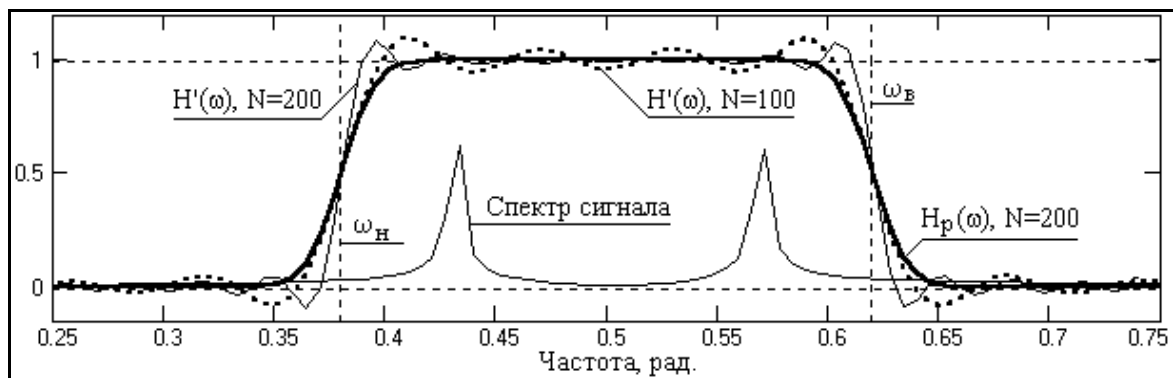


Рис. 7.3.1. Передаточные функции полосового фильтра.

**Применение весовых функций.** Если уровень пульсаций передаточной функции, определяемый явлением Гиббса, не удовлетворяет поставленным задачам фильтрации данных, рекомендуется использование сглаживающих весовых функций. С учетом того, что при применении весовых функций происходит расширение переходных зон примерно в два раза, значение ширины переходной зоны будет равным  $\Delta_p = 2\pi/N$ . Отсюда можно определить минимальное число членов усеченного ряда по заданному размеру переходной зоны:

$$N = 2\pi/\Delta_p. \quad (7.3.2)$$

Для примера на рис. 7.3.1 значение  $N$  принято равным 200, при этом крутизна переходной зоны увеличилась (тонкая кривая  $H'(\omega)$ ,  $N=200$ ), создавая запас на последующее сглаживание весовой функцией.

Выбор весовых функций целесообразно осуществлять по допустимой величине осцилляций усиления сигнала в полосе подавления, т.е. по относительному значению амплитуды первого выброса на передаточных характеристиках весовых функций. Для

выбранной весовой функции (с учетом числа ее членов по (7.3.2)) производится расчет весовых коэффициентов  $p_n$ , после чего устанавливаются окончательные значения оператора фильтра:

$$h_n = p_n h(n). \quad (7.3.3)$$

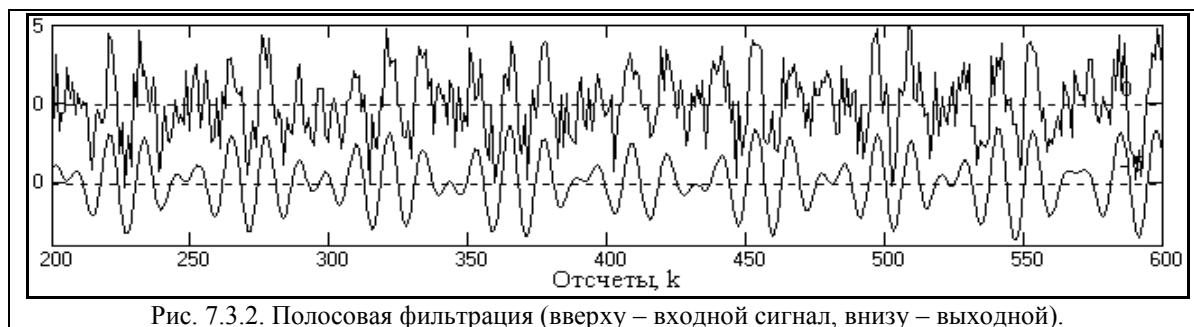


Рис. 7.3.2. Полосовая фильтрация (вверху – входной сигнал, внизу – выходной).

Подстановкой коэффициентов (7.3.3) в (7.3.1) рекомендуется произвести построение полученной передаточной характеристики фильтра и непосредственно по ней оценить пригодность фильтра для поставленных задач. Это наглядно видно на рис. 7.3.1, где для нашего примера была применена весовая функция Гаусса. Передаточная функция  $H_p(\omega)$  имеет практически такую же крутизну, как и функция  $H'(\omega)$  при  $N=100$  и практически плоскую вершину в интервале спектра сигнала. Качество работы фильтра для сигнала, приведенного на рис. 7.2.1, можно видеть на рис. 7.3.2.

При необходимости более точной оценки полученной передаточной функции можно рекомендовать увеличение ее частотного разрешения в 2-4 раза перед выполнением преобразования Фурье, что можно выполнить путем увеличения размеров оператора  $h_n$  дополнением нулями.

**Основные весовые функции.** Ниже в таблицах приведены формулы и основные спектральные характеристики наиболее распространенных весовых окон. Носители весовых функций, в принципе, являются неограниченными и при использовании в качестве весовых окон действуют только в пределах окна и обнуляются за его пределами. Для упрощения записи формулы приводятся в аналитической форме с временным окном  $2\tau$ , симметричным относительно нуля ( $0 \pm \tau$ ). При переходе к дискретной форме окно  $2\tau$  заменяется окном  $2N+1$ , а значения  $t$  – дискретами  $t = n\Delta t$ . Большинство весовых функций на границах окна ( $n = \pm N$ ) принимают нулевые или близкие к нулевым значения. Последнее исключается, если принять  $2\tau = (2N+3)\Delta t$ , при этом близкие к нулю значения перемещаются за границы окна.

*Основные весовые функции.*

Временное окно	Весовая функция	Фурье-образ
Естественное ( $\Pi$ )	$\Pi(t) = 1,  t  \leq \tau; \Pi(t) = 0,  t  > \tau$	$\Pi(\omega) = 2\tau \text{ sinc}[\omega\tau]$
Бартлетта ( $\Delta$ )	$b(t) = 1 -  t /\tau$	$B(\omega) = \tau \text{ sinc}^2(\omega\tau/2)$
Хеннинга, Ганна	$p(t) = 0.5[1 + \cos(\pi t/\tau)]$	$0.5\Pi(\omega) + 0.25\Pi(\omega + \pi/\tau) + 0.25\Pi(\omega - \pi/\tau)$
Хемминга	$p(t) = 0.54 + 0.46 \cos(\pi t/\tau)$	$0.54\Pi(\omega) + 0.23\Pi(\omega + \pi/\tau) + 0.23\Pi(\omega - \pi/\tau)$
Карре (2-е окно)	$p(t) = b(t) \text{ sinc}(\pi t/\tau)$	$\tau \cdot B(\omega) * \Pi(\omega), \Pi(\omega) = 1$ при $ \omega  < \pi/\tau$
Лапласа-Гаусса	$p(t) = \exp[-\beta^2(t/\tau)^2/2]$	$[(\tau/\beta)\sqrt{2\pi} \exp(-\tau^2\omega^2/(2\beta^2))] \otimes \Pi(\omega)$
Кайзера-Бесселя	$p(t) = \frac{J_0[\beta\sqrt{1-(t/\tau)^2}]}{J_0[\beta]}$	Вычисляется преобразованием Фурье. $J_0[x]$ - модифицированная функция Бесселя нулевого порядка
	$J_0[x] = \sum_{k=1}^{\infty} [(x/2)^k/k!]^2$	

*Характеристики спектров весовых функций.*

Параметры	Ед.	П-	Барт-	Лан-	Хен-	Хемм	Кар-	Лапла	Кайзе
-----------	-----	----	-------	------	------	------	------	-------	-------

	изм.	Окно	летт	цош	нинг	инг	ре	с	р
Амплитуда:									
Главный пик	$\tau$	2	1	1.18	1	1.08	0.77	0.83	0.82
1-й выброс(-)	%Гл.п.	0.217	-	0.048	0.027	0.0062	-	0.0016	.00045
2-й выброс(+)	- " -	0.128	0.047	0.020	0.0084	0.0016	-	0.0014	.00028
Ширина Гл. пика	$\omega\tau/2\pi$	0.60	0.89	0.87	1.00	0.91	1.12	1.12	1.15
Положения:									
1-й нуль	$\omega\tau/2\pi$	0.50	1.00	0.82	1.00	1.00	-	1.74	1.52
1-й выброс	$\omega\tau/2\pi$	0.72	-	1.00	1.19	1.09	-	1.91	1.59
2-й нуль	$\omega\tau/2\pi$	1.00	-	1.29	1.50	1.30	-	2.10	1.74
2-й выброс	$\omega\tau/2\pi$	1.22	1.44	1.50	1.72	1.41	-	2.34	1.88

**Весовая функция Кайзера.** Наибольшее распространение при расчетах частотных НЦФ получила весовая функция Кайзера:

$$p(n) = \frac{J_0[\beta\sqrt{1-(n/N)^2}]}{J_0[\beta]} .$$

Это объясняется тем, что параметры функции Кайзера могут устанавливаться непосредственно по техническим требованиям к передаточным функциям проектируемых фильтров – допустимой ширине переходной зоны  $\Delta_p$  и значению коэффициента шума фильтра  $\delta$  (максимальным значениям осцилляций передаточной функции в единицах коэффициента передачи в полосе пропускания).

Кайзером установлено, что для заданного значения  $\delta$  произведение количества членов оператора НЦФ на ширину переходной зоны является величиной постоянной. Оно получило название D-фактора:

$$D = N \cdot \Delta_p / \pi.$$

С другой стороны, установлены следующие эмпирические соотношения между D-фактором и параметром  $\beta$  функции Кайзера:

$$\begin{aligned} D &= (A-7.95)/14.36 && \text{при } A > 21. \\ &= 0.9222 && \text{при } A < 21. \\ \beta &= 0.1102(A-8.7) && \text{при } A > 50. \\ &= 0 && \text{при } A < 21. \\ &= 0.5842(A-21)^{0.4} + 0.07886(A-21), && 21 < A < 50. \end{aligned}$$

где:  $A = -20 \log \delta$  - затухание в децибелах.

Приведенные выражения позволяют по заданному значению коэффициента шума  $\delta$  определить параметр  $\beta$  функции Кайзера, а через D-фактор число членов фильтра:

$$N = \pi D / \Delta_p.$$

При проектировании полосовых фильтров проверка передаточной функции полученного оператора НЦФ исходному заданию по значению коэффициента шума является обязательной. Это объясняется тем, что поскольку полоса пропускания полосового фильтра ограничена двумя скачками, на передаточной характеристике возникают два центра осцилляций, при этом наложение осцилляций может как уменьшить, так и увеличить амплитуду суммарных осцилляций. Если за счет наложения произойдет увеличение амплитуды осцилляций, то расчет НЦФ следует повторить с уменьшением исходного значения  $\delta$ .

#### Пример расчета полосового фильтра.

Произвести расчет ПФ при следующих исходных параметрах:  $\omega_n = 0.3\pi$ ,  $\omega_b = 0.6\pi$ ,  $\Delta_p = 0.1\pi$ ,  $\delta = 0.02$ .

1.  $A = -20 \log \delta$ .  $A = 34$ .
2.  $N = \pi (A-7.95)/(14.36 \Delta_p)$ .  $N = 18$ .
3.  $\beta = 0.5842(A-21)^{0.4} + 0.07886(A-21)$ .  $\beta = 2.62$ .
4.  $h_0 = (\omega_b - \omega_n)/\pi$ .  $h_0 = 0.3$

5.  $h(n) = (\sin n\omega_b - \sin n\omega_n)/(n\pi)$ .  $h(n) = 0.04521, -0.24490, -0.09515, \dots, 0.02721$ .
6.  $p_n = J_0\{\beta \sqrt{1 - (n/N)^2}\} / J_0\{\beta\}$ .  $p_n = 1.00, 0.997, 0.9882, \dots$
7. Оператор фильтра:  $h_n = p_n h(n)$ ,  $n = 0, 1, 2, \dots, N$ .  $h_n = h_n$ .  $h_n = 0.3000, 0.04508, -0.2420, \dots$
8. Проверка по формуле:  $H(\omega) = \sum_{n=-N}^N h_n \cos n\omega$ ,  $0 \leq \omega \leq \pi$ .

Для оценки формы передаточной функции количество точек спектра в интервале  $0-\pi$  достаточно задать равным  $2N$ , т.е. с шагом  $\Delta\omega \leq \pi/36$ .

**Влияние конечной разрядности на цифровые** фильтры должно быть минимальным и не создавать на их частотных характеристиках дополнительных неравномерностей и отклонения от заданной формы. С чисто практической точки зрения ограничение разрядности коэффициентов фильтра в целях повышения производительности вычислений лучше всего (и проще всего) выполнять непосредственно сравнением частотных характеристик с изменением разрядности от большей к меньшей. Следует учитывать, что ограничение разрядности может по-разному сказываться на неравномерности фильтра в полосе пропускания и степени затухания сигналов в полосе подавления.

Ошибки отклонения  $\varepsilon(\omega)$  частотной характеристики относительно заданной при проектировании кроме разрядности коэффициентов  $B$  в битах зависит также от размеров  $N$  оператора фильтра и в первом приближении может оцениваться по формулам:

$$|\varepsilon(\omega)| = N 2^{-B}, \quad (7.3.4)$$

$$|\varepsilon(\omega)| = 2^{-B}(N/3)^{1/2}, \quad (7.3.5)$$

$$|\varepsilon(\omega)| = N 2^{-B}[(N \ln N)/3]^{1/2}. \quad (7.3.6)$$

Выражение (7.3.4) наиболее пессимистично и предполагает наихудшие ситуации вычислений. Два других выражения носят более реальный характер по статистическим данным.

#### 7.4. ГЛАДКИЕ ЧАСТОТНЫЕ ФИЛЬТРЫ

В некоторых случаях (при последовательном соединении фильтров, при выделении сигналов на уровне сильных помех и т.п.) осцилляции на передаточных характеристиках фильтров являются весьма нежелательными даже при их малой остаточной величине. Так, например, двойное последовательное применение фильтров приводит к тому, что ошибки в полосе пропускания приблизительно удваиваются, а в полосе подавления возводятся в квадрат, при этом длина окна эквивалентного фильтра практически удваивается.

**Принцип синтеза фильтров.** Очевидно, что фильтры с гладкой передаточной характеристикой можно получить только в том случае, если возможно разложение передаточной функции в конечный ряд Фурье.

Допустим, мы имеем симметричный НЦФ с передаточной функцией:

$$H(\omega) = h_0 + 2 \sum_{n=1}^N h_n \cos n\omega. \quad (7.4.1)$$

Как известно,  $\cos n\omega$  равен полиному по  $\cos \omega$  степени  $n$ , при этом выражение (7.4.1) можно записать в виде:

$$H(\omega) = \sum_{n=0}^N g_n (\cos \omega)^n = \sum_{n=0}^N g_n x^n, \quad (7.4.2)$$

где переменная  $x = \cos \omega$  изменяется от 1 до -1 (поскольку  $\omega$  изменяется от 0 до  $\pi$ ). Преобразование переменной представляет собой нелинейное растяжение оси абсцисс с поворотом на  $180^\circ$  (по переменной  $x$  передаточные функции ФНЧ похожи на ФВЧ, и наоборот) с выражением функции через степенной полином. Последнее примечательно тем, что синтез гладких функций на базе степенных полиномов затруднений не представляет.

Так, например, для конструирования ФНЧ в качестве исходной может быть принята степенная функция вида:

$$g(x) = (1+x)^z(1-x)^r, \quad (7.4.3)$$

где  $z$  и  $r$  - параметры.

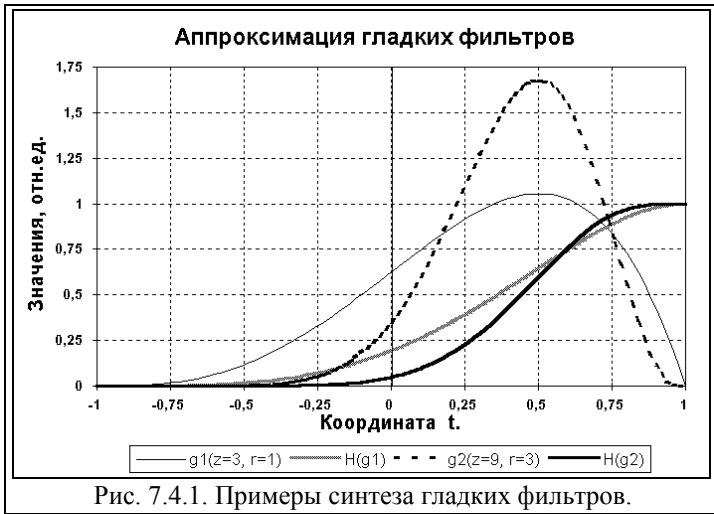


Рис. 7.4.1. Примеры синтеза гладких фильтров.

Функция (7.4.3) имеет нули порядка  $z$  и  $r$  в точках соответственно  $x = -1$  и  $x = 1$  (рис. 7.4.1), при этом значения параметров  $z$  и  $r$  характеризуют степень касания функцией оси абсцисс. Чем больше порядок, тем медленнее функция отходит ("отрывается") от оси абсцисс.

Если выражение функции (7.4.3) проинтегрировать в пределах от  $-1$  до  $x$  и нормировать на значение интеграла от  $-1$  до  $1$ , то будет получена гладкая передаточная характеристика низкочастотного

фильтра. На рисунке 7.4.1 приведены передаточные функции для двух пар параметров  $z$  и  $r$ , вычисленные по формуле:

$$H(x) = \int_{-1}^x g(x)dx / \int_{-1}^1 g(x)dx. \quad (7.4.4)$$

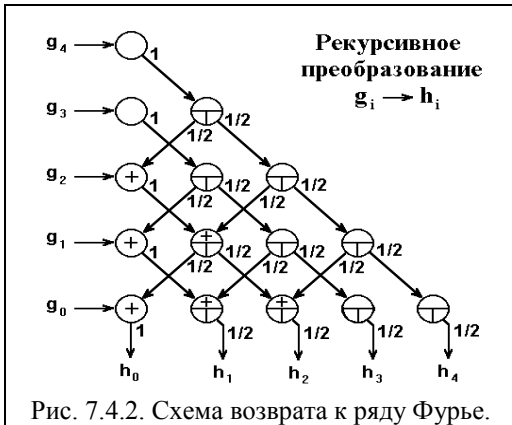


Рис. 7.4.2. Схема возврата к ряду Фурье.

Функция  $H(x)$  имеет перегиб в точке  $(z-r)/(z+r)$  и переходную зону, крутизна которой тем больше, чем больше значения  $z$  и  $r$ . Подстановкой  $x = \cos \omega$  осуществляется возврат к частотной переменной  $\omega$  с сохранением монотонности функции.

В заключение, для определения коэффициентов фильтра  $h_n$  требуется осуществить обратное преобразование от степенной формы (7.4.2) к ряду Фурье (7.4.1). Выполнение данной операции достаточно просто производится рекурсивным способом, показанным на рис. 7.4.2.

Подробное обоснование рекурсии приведено в /24/.

#### Пример расчета гладкого фильтра.

Произвести расчет ФНЧ с гладкой частотной характеристикой с перегибом характеристики в точке  $\pi/3$ . За исходную функцию принять функцию (7.4.3).

1.  $x = \cos(\pi/3) = 0.5 = (z-r)/(z+r)$ . Принято:  $z=3, r=1$ .

Исходный многочлен:  $g(x) = (1-x)(1+x)^3 = 1+2x-2x^3-x^4$ .

2.  $H(x) = \int_{-1}^x g(x)dx = C+x+x^2-0.5x^4-0.2x^5$ . При  $x=-1, H(-1)=0$ , откуда  $C=0.3$ . При  $x=1, H(1)=1.6$ .

Отсюда:  $H(x) = (3+10x+10x^2-5x^4-2x^5)/16$ .  $g_n = \{3/16, 10/16, 10/16, 0, -5/16, -2/16\}$ .

3. Применяя рекурсивное преобразование, получаем:  $h_n = \{98, 70, 20, -5, -5, -1\}/256$ .

Для расчетов гладких фильтров высоких частот в выражении (7.4.4) достаточно поменять местами пределы интегрирования. Гладкие полосовые фильтры получаются комбинацией ФНЧ и ФВЧ с перекрытием частот пропускания.

### 7.5. ДИФФЕРЕНЦИРУЮЩИЕ ЦИФРОВЫЕ ФИЛЬТРЫ.

**Передаточная функция.** Из выражения для производной  $d(\exp(j\omega t))/dt = j\omega \exp(j\omega t)$  следует, что при расчете фильтра производной массива данных необходимо аппроксимировать рядом Фурье передаточную функцию вида  $H(\omega) = j\omega$ . Поскольку

коэффициенты такого фильтра будут обладать нечетной симметрией ( $h_{-n} = -h_n$ ) и выполняется равенство

$$h_n [\exp(j\omega n) - \exp(-j\omega n)] = 2j h_n \sin n\omega,$$

то передаточная характеристика фильтра имеет вид:

$$H(\omega) = 2j(h_1 \sin \omega + h_2 \sin 2\omega + \dots + h_N \sin N\omega),$$

т.е. является мнимой нечетной, а сам фильтр является линейной комбинацией разностей симметрично расположенных относительно  $s_k$  значений функции. Уравнение фильтрации:

$$y_n = \sum_{n=1}^N h_n (s_{k+n} - s_{k-n}).$$

Если дифференцированию подлежит низкочастотный сигнал, а высокие частоты в массиве данных представлены помехами, то для аппроксимации в пределах главного частотного диапазона задается (без индекса мнимости) передаточная функция фильтра вида:

$$H(\omega) = \omega, \quad \omega \leq \omega_b, \quad H(\omega) = 0, \quad \omega_b < \omega \leq \omega_N.$$

Оператор дифференцирующего фильтра:

$$h(n) = (2/\pi) \int_0^{\omega_b} H(\omega) \sin(n\pi\omega/\omega_N) d\omega, \quad n = 0, 1, 2, \dots \quad (7.5.1)$$

Принимая, как обычно,  $\omega_N = \pi$  ( $\Delta t = 1$ ) и решая (7.5.1) при  $H(\omega) = \omega$ , получаем:

$$h_n = (2/\pi) [\sin(n\omega_b)/n^2 - \omega_b \cos(n\omega_b)/n], \quad (7.5.2)$$

$$h_0 = 0, \quad h_{-n} = -h_n.$$

Частотная характеристика:

$$\text{Im}(H(\omega)) = \sum_{n=-\infty}^{\infty} h_n \sin n\omega = 2 \sum_{n=1}^{\infty} h_n \sin n\omega. \quad (7.5.3)$$

**Точность дифференцирования.** На рис. 7.5.1 приведен пример расчета коэффициентов дифференцирующего фильтра на интервал частот  $\{0-0.5\}\pi$  при  $\Delta t=1$  ( $\omega_b = \pi/2$ ). Операторы дифференцирующих фильтров, как правило, затухают очень медленно и, соответственно, достаточно точная реализация функции (7.5.3) весьма затруднительна.

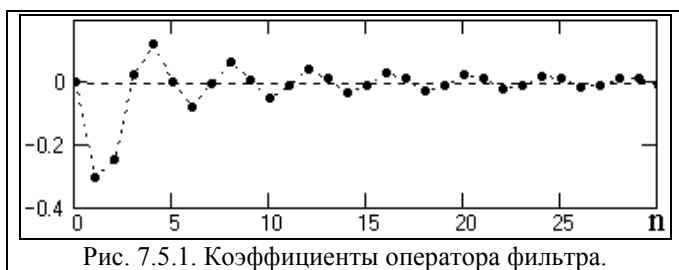


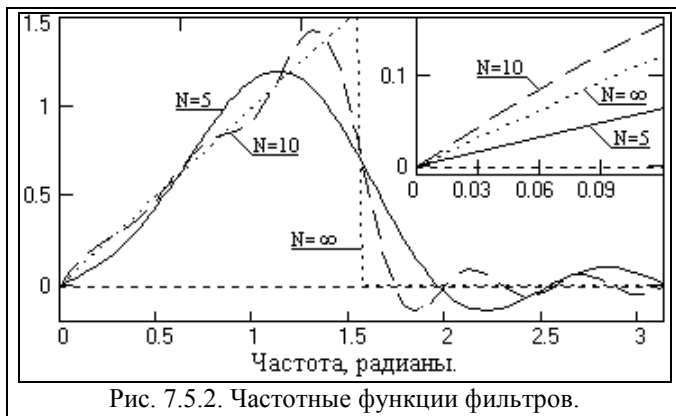
Рис. 7.5.1. Коэффициенты оператора фильтра.

Ряд (7.5.3) усекается до  $N$  членов, и с помощью весовых функций производится нейтрализация явления Гиббса. Явление Гиббса для дифференцирующих фильтров имеет весьма существенное значение, и может приводить к большим погрешностям при обработке информации, если не

произвести его нейтрализацию. Примеры ограничения оператора, приведенного на рис. 7.5.1, и соответствующие передаточные функции  $H'(\omega)$  усеченных операторов показаны на рис. 7.5.2.

Для оценки возможных погрешностей дифференцирования усеченными операторами произведем расчет фильтра при  $\omega_b = \pi/2$ . По формулам (7.5.2) определяем:

$$h_{0-10} = 0, 0.3183, 0.25, -0.0354, -0.125, 0.0127, 0.0833, -0.0065, -0.0625, 0.0039, 0.05.$$



Произведем проверку работы фильтра на простом массиве данных  $s_n = n$ , производная которого постоянна и равна 1. Для массива с постоянной производной фильтр может быть проверен в любой точке массива, в том числе и в точке  $n=0$ , для которой имеем:

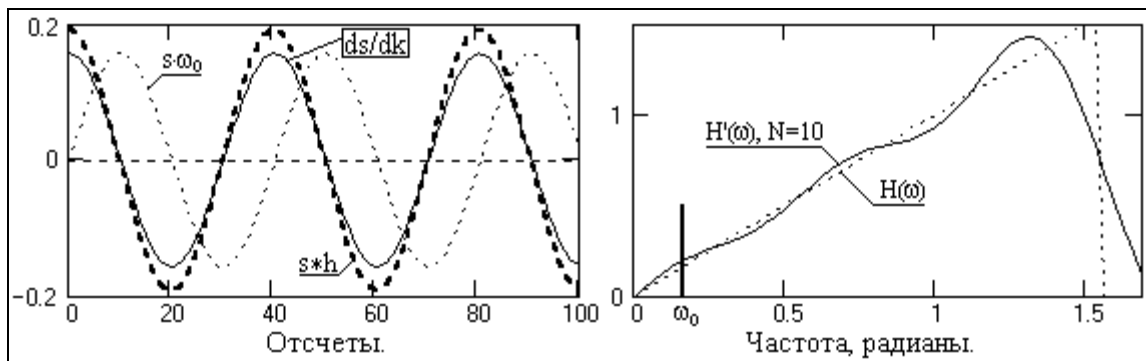
$$y = \sum_{n=-N}^N h_n s_{0-n} = 2 \sum_{n=1}^N n h_n,$$

при этом получаем:  $y=0.5512$  при  $N=5$ ,  $y=1.53$  при  $N=10$ .

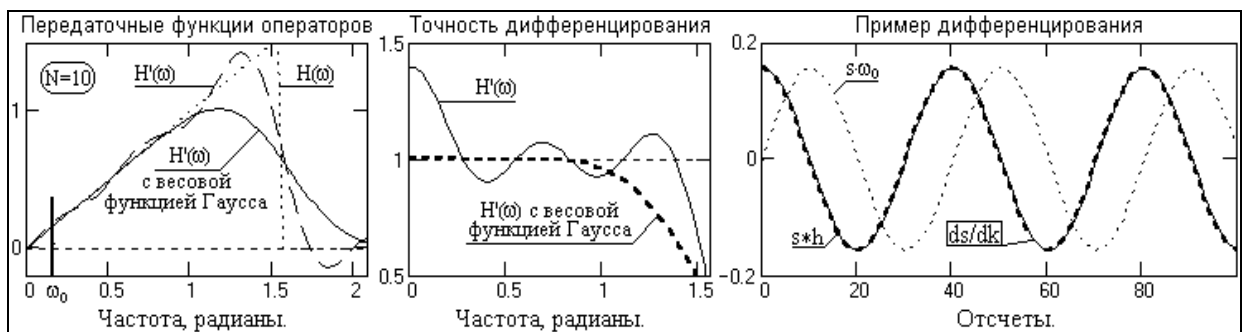
Такое существенное расхождение с действительным значением производной объясняется тем, что при  $\omega=0$  тангенс угла наклона реальных передаточных функций фильтра, как это видно на рисунке 7.5.2, весьма существенно отличается от тангенса угла наклона аппроксимируемой функции  $H(\omega) = \omega$ . На рис. 7.5.3 приведены частотные графики относительной погрешности дифференцирования  $\sigma = H'_N(\omega)/H_N(\omega)$



с вычислением значений на нулевой частоте по пределам функций при  $N \rightarrow \infty$ . На рис. 7.5.4 приведен пример операции дифференцирования гармоник  $s$  с частотой  $\omega_0$  оператором с  $N=10$  в сопоставлении с точным дифференцированием  $ds/dk$ .



**Применение весовых функций.** Применим для нейтрализации явления Гиббса весовую функцию Хемминга. Результат нейтрализации для фильтра с  $N=10$  приведен на рис. 7.5.5. Повторим проверочный расчет дифференцирования на массиве  $s_n = n$  и получим результат  $y=1.041$ , т.е. погрешность дифференцирования уменьшается порядок.



Аналогично производится расчет и полосовых дифференцирующих фильтров с соответствующим изменением пределов интегрирования в (7.5.1) от  $\omega_n$  до  $\omega_v$ . При этом получаем:

$$h_n = (\omega_n \cos n\omega_n - \omega_v \cos n\omega_v) / (n\pi) + (\sin n\omega_v - \sin n\omega_n) / (n2\pi).$$

**Фильтры с линейной групповой задержкой.** Дифференцирующие фильтры, а равно и любые другие фильтр с мнимой частотной характеристикой, например, оператор преобразования Гильберта, могут быть выполнены в каузальном варианте при условии обеспечения линейной групповой задержки сигнала, которое записывается следующим образом:

$$\varphi(\omega) = \beta - \alpha\omega, \quad (7.5.4)$$

где  $\beta$  и  $\alpha$  – константы.

Оно выполняется, если импульсная характеристика фильтра имеет положительную симметрию:

$$h(n) = -h(N-n-1), \quad n = 0, 1, 2, \dots, (N-1)/2, \quad N - \text{нечетное (тип 1);}$$

$$n = 0, 1, 2, \dots, (N/2)-1, \quad N - \text{четное (тип 2).}$$

При этом фазовая характеристика будет определяться длиной фильтра:

$$\alpha = (N-1)/2, \quad \beta = \pi/2.$$

Частотная характеристика фильтра:

$$H(\omega) = |H(\omega)| \exp(j\varphi(\omega)), \quad (7.5.4)$$

где модуль  $|H(\omega)|$  задается нечетным. Оба типа фильтров вводят в выходной сигнал сдвиг фазы на  $90^\circ$ . Кроме того, частотная характеристика фильтра типа 1 всегда равно нулю на частоте Найквиста, что определяется знакопеременностью левой и правой части главного диапазона спектра с учетом периодизации спектра дискретных функций.

### 7.6. АЛЬТЕРНАТИВНЫЕ МЕТОДЫ РАСЧЕТА НЦФ

Метод прямого расчета НЦФ по частотной характеристике понятен и прост для применения. Недостаток метода – отсутствие гибкости. Он не позволяет проектировать фильтры с разной степенью неравномерности частотной характеристики в полосах пропускания и подавления, а степень неравномерности не зависит от количества членов фильтра и не может изменяться. Максимальные осцилляции частотной характеристики всегда наблюдаются в области полосовых границ и уменьшаются при удалении от них, но при близких границах могут наблюдаться явления интерференции осцилляций. Более гибкими в проектировании являются альтернативные методы: оптимизационные,

**Оптимизационные методы** позволяют проектировать экономные по размерам операторы фильтров с оптимальными (по Чебышеву) осцилляциями частотных характеристик. Они основаны на понятии полос равных колебаний.

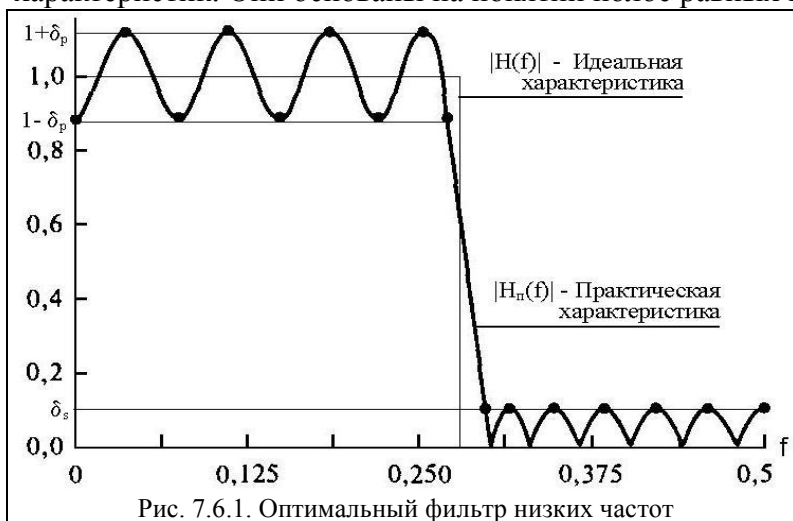


Рис. 7.6.1. Оптимальный фильтр низких частот

Частотная характеристика оптимального фильтра низких частот приведена на рис. 7.6.1. В полосе пропускания реальная характеристика фильтра осциллирует с постоянными амплитудными колебаниями между значениями  $1-\delta_p$  и  $1+\delta_p$ . В полосе подавления осцилляции постоянной амплитуды находятся в интервале  $0-\delta_s$ . Разность между идеальной и практической характеристиками представляет собой функцию ошибок  $E(f)$ . Оптимальный



метод позволяет определить коэффициенты фильтра  $h(n)$ , для которых значение максимальной взвешенной ошибки минимизируется

$$\min[\max(E(f))]$$

в полосе пропускания и в полосе подавления, при этом характеристика фильтра будет иметь равные колебания в пределах полос пропускания и подавления, а количество экстремумов колебаний у фильтров с линейной фазовой характеристикой обычно прямо связано с количеством коэффициентов фильтра  $(N+1)/2$ .

При расчете фильтра ключевым моментом является определение положения частот экстремумов, которое выполняется итерационным алгоритмом Ремеза, после чего по положениям экстремумов задается частотная характеристика фильтра и определяются его коэффициенты. Методика расчета оптимальных фильтров подробно с примерами, в том числе в среде Matlab, рассмотрена в работе /43/.

**Метод частотной выборки** представляет собой вариант метода расчета фильтра по частотной характеристике без применения весовых функций и может применяться для расчетов как частотно-избирательных фильтров, так и фильтров с произвольной частотной характеристикой.

В основе метода лежит непосредственное задание частотной характеристики фильтра в цифровой форме с последующим подбором переходных зон под требуемые характеристики фильтра по величине допустимых осцилляций в полосе пропускания и подавления. Расчет желательно вести в интерактивном режиме, например, в среде Mathcad. В качестве примера приведем расчет низкочастотного фильтра.

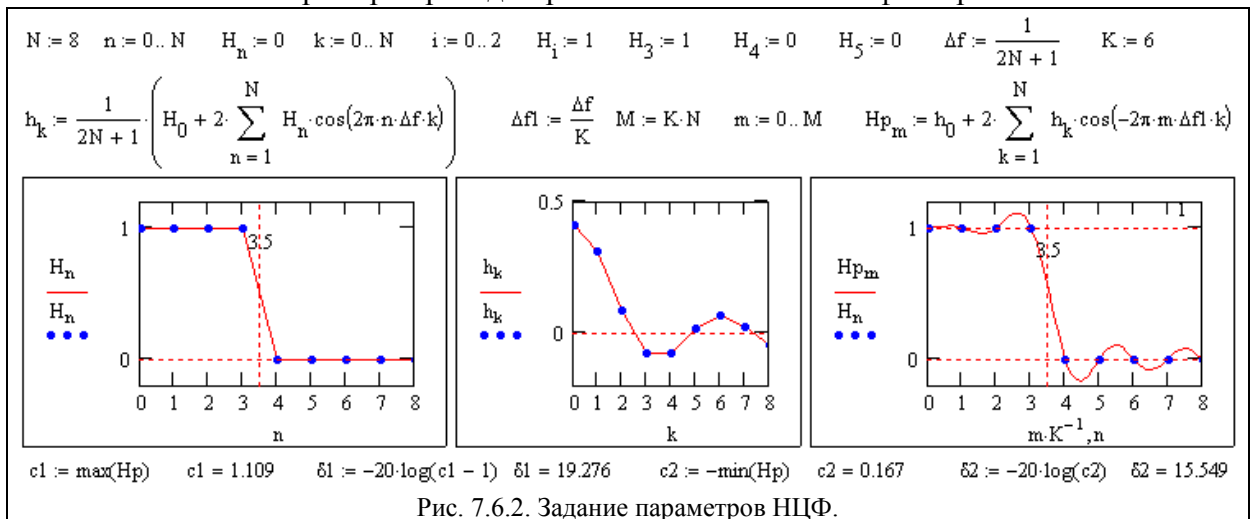
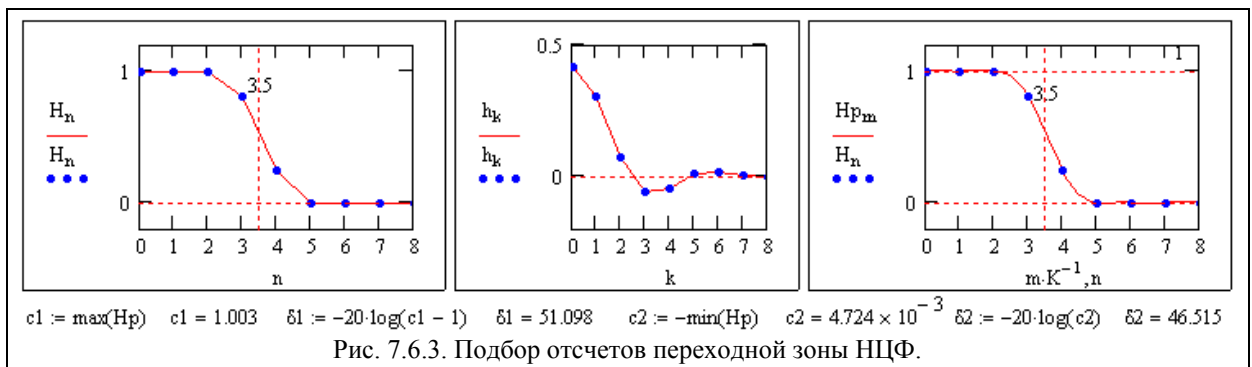
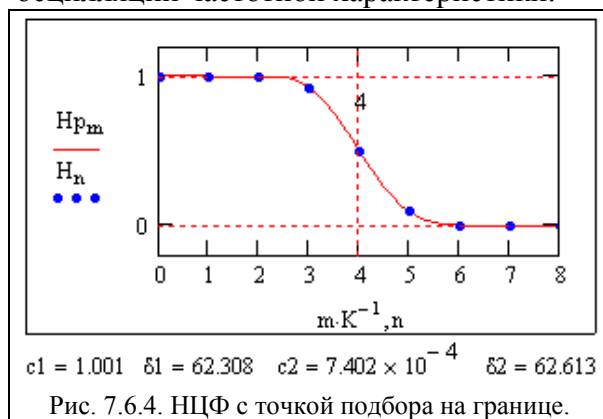


Рис. 7.6.2. Задание параметров НЦФ.

Допустим, нам требуется достаточно простой симметричный низкочастотный фильтр с шириной переходной зоны порядка 0.2 главного частотного диапазона (при  $\Delta k=1$  для фильтра,  $f_N = 0.5$  Гц для спектра и ширина переходной зоны  $0.2 \times 0.5 = 0.1$  Гц). Минимальный размер фильтра при идеальной характеристике для обеспечения такого перехода  $2N+1 = 2(1+1/0.1) = 11$  точек. С учетом расширения переходной зоны при уменьшении осцилляций на границе зон примем для начала  $N=8$ . Частотная характеристика проектируемого фильтра (правая половина) приведена на рис. 7.6.2 с границей раздела зон между 3 и 4 отсчетами спектра. Расчет оператора фильтра проводим обратным преобразованием Фурье, а по полученным отсчетам оператора вычисляем фактическую частотную характеристику этого оператора с уменьшением шага по частоте в 4-6 раз, что позволяет выявить осцилляции и определить погрешность фильтра (по максимумам осцилляций).



На рис. 7.6.3. показан результат подбора частотных значений характеристики фильтра в районе переходной зоны (2 точки), что позволяет более чем в 30 раз снизить осцилляции частотной характеристики.



Попутно заметим, что изменение осцилляций характеристики фильтра может производиться индивидуально для зоны пропускания (левой от границы точкой) и зоны подавления (правой точкой) в зависимости от того, требуется ли более высокая точность пропускания или подавления частот. Особенно эффективно это при использовании трех точек подбора с расположением центральной точки на границе полос пропускания и подавления, как это показано на рис. 7.6.4.

При использовании данного метода может использоваться и комбинированный подход: задание на частотной характеристике избыточного количества точек, отладка параметров фильтра по трем и более точкам в переходных зонах, а затем усечение оператора фильтра с применением весовых функций.

Метод частотных выборок допускает также рекурсивную реализацию фильтров /43/.

## Тема 8. Z-ПРЕОБРАЗОВАНИЕ СИГНАЛОВ И СИСТЕМНЫХ ФУНКЦИЙ

Цифровая обработка сигналов оперирует с дискретными представлениями сигналов. Математика дискретных преобразований зародилась в недрах аналоговой математики еще в 18 веке в рамках теории рядов и их применения для аппроксимации функций, однако ускоренное развитие она получила в 20 веке после появления первых вычислительных машин. В принципе, в своих основных положениях математический аппарат дискретных преобразований подобен преобразованиям аналоговых сигналов и систем. Однако дискретность данных требует учета этого фактора, а его игнорирование может приводить к существенным ошибкам. Кроме того, ряд методов дискретной математики не имеет аналогов в аналитической математике.

Распространенным способом анализа дискретных цифровых последовательностей является z-преобразование (z-transform). Оно играет для дискретных сигналов и систем такую же роль, как для аналоговых – преобразование Лапласа. Большое значение z-преобразование имеет для расчетов рекурсивных цифровых систем обработки сигналов, а потому рассматривается отдельной темой перед началом изучения рекурсивных цифровых фильтров.

### 8.1. Z – ТРАНСФОРМАЦИЯ СИГНАЛОВ

**Определение z-преобразования.** Z-преобразование представляет собой разложение

функций в ряды степенных полиномов по  $z$ . Впервые  $z$ -преобразование введено в употребление П.Лапласом в 1779 и повторно "открыто" В.Гуревичем в 1947 году с изменением символики на  $z^{-k}$ . В настоящее время в технической литературе имеют место оба вида символики. На практическое использование преобразования это не влияет, т.к. смена знака только зеркально изменяет нумерацию членов полинома (относительно  $z^0$ ), числовое пространство которых в общем случае от  $-\infty$  до  $+\infty$ . В дальнейшем в качестве основной будем использовать символику положительных степеней  $z$ , давая пояснения по особенностям отрицательной символики, если таковая имеется.

Произвольной непрерывной функции  $s(t)$ , равномерно дискретизированной и отображенной отсчетами  $s_k = s(k\Delta t)$ , равно как и непосредственно дискретной функции, можно поставить в однозначное соответствие степенной полином по  $z$ , последовательными коэффициентами которого являются значения  $s_k$ :

$$s_k = s(k\Delta t) \Leftrightarrow \text{TZ}[s(k\Delta t)] = \sum_{k=-\infty}^{\infty} s_k z^k = S(z). \quad (8.1.1)$$

где  $z = \sigma + j\omega$  - произвольная комплексная переменная. В показательной форме  $z = r \cdot \exp(-j\varphi)$ , где  $r = |z| = \sqrt{\sigma^2 + \omega^2}$ ,  $\varphi = \arg(z) = \text{argtg}(\omega/\sigma)$ .

Пример 1:  $s_k = \{1, 2, 0, -1, -2, -1, 0, 0\}$ .  
 $S(z) = 1z^0 + 2z^1 + 0z^2 - 1z^3 - 2z^4 - 1z^5 + 0z^6 + 0z^7 = 1 + 2z - z^3 - 2z^4 - z^5$ .

В каузальных системах значения импульсного отклика систем существуют при  $k \geq 0$  и уравнение (8.1.1) действует в одностороннем варианте:

$$H(z) = \sum_{k=0}^{\infty} h_k z^k.$$

В общем случае,  $z$ -преобразование – это степенной ряд с бесконечным количеством членов, поэтому он может сходиться не для всего пространства значений  $z$ . Область  $z$ , в которой  $z$ -преобразование сходится и значения  $S(z)$  конечны, называют областью сходимости.

Пример 2: Последовательность (сигнал) конечной длины, непричинная:  $s_k = \{1, 2, 3, 2, 1\}$ ,  $k = 0, 1, 2, 3, 4$ .

$$S(z) = 1z^0 + 2z^1 + 3z^2 + 2z^3 + 1z^4 = 1 + 2/z + 3/z^2 + 2/z^3 + 1/z^4.$$

Очевидно, что  $S(z) = \infty$  при  $z = 0$ . Область сходимости – все значения  $z$ , за исключением  $z = 0$ .

Пример 3: Последовательность конечной длины, причинная (как импульсный отклик каузальной системы):  $s_k = \{1, 2, 3, 2, 1\}$ ,  $k = 0, 1, 2, 3, 4$ .

$$S(z) = 1 + 2z + 3z^2 + 2z^3 + z^4.$$

$S(z) = \infty$  при  $z = \infty$ . Область сходимости – все значения  $z$ , за исключением  $z = \infty$ .

Пример 4: Последовательность конечной длины, двусторонняя (как импульсный отклик симметричного фильтра):  $s_k = \{1, 2, 3, 2, 1\}$ ,  $k = -2, -1, 0, 1, 2$ .

$$S(z) = 1z^{-2} + 2z^{-1} + 3z^0 + 2z^1 + 1z^2 = 1/z^2 + 2/z + 3 + 2z + z^2.$$

$S(z) = \infty$  при  $z = 0$  и  $z = \infty$ . Область сходимости не включает точки  $z = 0$  и  $z = \infty$ .

Пример 5: Последовательность бесконечной длины, причинная (как импульсный отклик рекурсивного интегрирующего фильтра):  $s_k = 0$  при  $k < 0$ ,  $s_k = 1$  при  $k \geq 0$ .

$$S(z) = z^{-0} + z^1 + z^2 + z^3 + \dots = 1 + z + z^2 + z^3 + \dots = 1/(1-z)$$

Ряд удовлетворяет условию сходимости только при  $|z| < 1$ .

Значения  $z$ , для которых  $S(z) = \infty$ , называются *полюсами*, а для которых  $S(z) = 0$ , называются нулями функции  $S(z)$ . Как видно из примеров, для последовательностей конечной длины  $z$ -преобразование сходится везде кроме точки  $z = \infty$  для имеющих правостороннюю часть ( $k \geq 0$ ), и точки  $z = 0$  для имеющих левостороннюю часть ( $k < 0$ ), в любых их комбинациях. Для бесконечных причинных последовательностей преобразование сходится везде внутри круга единичного радиуса с центром в начале координат.

По заданному или полученному в результате анализа какой-либо системы  $z$ -полиному однозначно восстанавливается соответствующая этому полиному функция

путем идентификации коэффициентов степеней при  $z^k$  с  $k$ -отсчетами функции.

**Пример 6:**  $S(z) = 1+3z^2+8z^3-4z^6-2z^7 = 1z^0+0z^1+3z^2+8z^3+0z^4+0z^5-0z^6-2z^7$ .  
 $s_k = \{1, 0, 3, 8, 0, 0, -4, -2\}$ .

Смысл величины  $z$  в  $z$ -полиноме заключается в том, что она является оператором единичной задержки по координатам функции. Умножение  $z$ -образа сигнала  $s(k)$  на величину  $z^n$  означает задержку сигнала (сдвиг вправо по временной оси) на  $n$  интервалов:  $z^n S(z) \Leftrightarrow s(k-n)$ . Чтобы убедиться в этом, достаточно в приведенном выше примере выполнить умножение многочлена  $S(z)$ , например на  $z^2$ , выполнить обратное преобразование и получить новый сигнал  $s_k = \{0, 0, 1, 0, 3, 8, 0, 0, -4, -2\}$ .

$Z$ -образы с положительными степенями  $z$  соответствуют каузальным (физически реализуемым) процессам и системам, которые работают в реальном масштабе времени с текущими и "прошлыми" значениями сигналов. При обработке информации на ЭВМ каузальность сигналов не относится к числу ограничений и возможно использование отрицательных степеней  $z$ , соответствующих отсчетам сигналов "вперед". Последнее применяется, например, при синтезе симметричных операторов фильтров, что позволяет производить обработку информации без внесения в сигнал фазовых искажений. При использовании символики  $z^{-1}$  "прошлым" значениям соответствуют значения с отрицательными степенями  $z$ , "будущим" – с положительными.

Основное достоинство  $z$ -преобразований заключается в простоте математических операций со степенными полиномами, что имеет немаловажное значение при расчетах цифровых фильтров и в спектральном анализе.

**Связь с преобразованиями Фурье и Лапласа.** Запишем дискретный сигнал  $s_k$  в виде суммы весовых импульсов Кронекера:

$$s_k = s(k\Delta t) = \sum_{n=-\infty}^{\infty} s(n\Delta t) \delta(k\Delta t - n\Delta t).$$

Определим спектр сигнала по теореме запаздывания:

$$S(\omega) = \sum_{k=-\infty}^{\infty} s(k\Delta t) \exp(-j\omega k\Delta t).$$

Выполним замену переменных,  $z = \exp(-j\omega\Delta t)$ , и получим:

$$S(\omega) = \sum_{k=-\infty}^{\infty} s(k\Delta t) \cdot z^k = S(z).$$

Отсюда следует, что дискретное преобразование Фурье является частным случаем  $z$ -преобразования при  $z = \exp(-j\omega\Delta t)$ .

Аналогичной подстановкой  $z = \exp(-p)$  может осуществляться переход к дискретному преобразованию Лапласа. В общем виде:

$$S(\omega) = S(z), \quad z = \exp(-j\omega\Delta t); \quad S(p) = S(z), \quad z = \exp(-p\Delta t). \quad (8.1.2)$$

Обратное преобразование:

$$S(z) = S(\omega), \quad \omega = \ln z / j\Delta t; \quad S(z) = S(p), \quad p = \ln z / \Delta t. \quad (8.1.3)$$

При отрицательной символике  $z$  связь между представлениями осуществляется соответственно подстановками  $z^{-1} = \exp(j\omega\Delta t)$  и  $z^{-1} = \exp(p)$ .

При  $z^k = \exp(-j\omega k\Delta t)$   $z$ -преобразование представляет собой особую форму представления дискретных сигналов, при которой на полином  $S(z)$  можно сослаться как на временную функцию (по значениям коэффициентов  $k\Delta t$ ), так и на функцию частотного спектра сигнала (по значениям аргумента  $\omega$ ).

**Отображение  $z$ -преобразования** выполняются на комплексной  $z$ -плоскости с  $\text{Re } z$  и  $\text{Im } z$  по осям координат (рис. 8.1.1). В частности, спектральной оси частот  $\omega$  на  $z$ -плоскости соответствует окружность радиуса:

$$|z| = |\exp(-j\omega\Delta t)| = \sqrt{\cos^2(\omega\Delta t) + \sin^2(\omega\Delta t)} = 1.$$

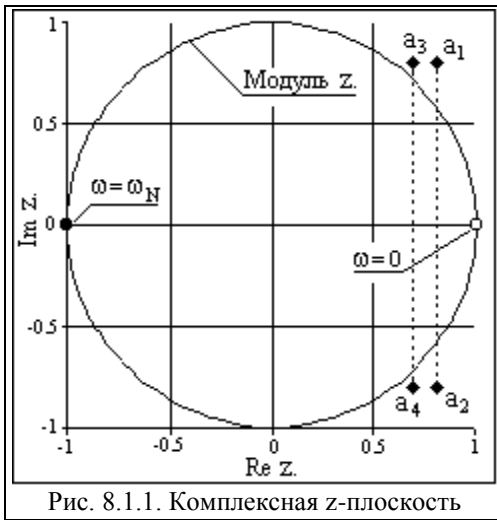


Рис. 8.1.1. Комплексная z-плоскость

Подстановка значения какой-либо частоты  $\omega$  в  $z = \exp(-j\omega\Delta t)$  отображается точкой на окружности. Частоте  $\omega = 0$  соответствует точка  $\text{Re } z = 1$  и  $\text{Im } z = 0$  на правой стороне оси абсцисс. При повышении частоты точка смещается по окружности против часовой стрелки, и занимает крайнее левое положение на частоте Найквиста  $\omega_N = \pi/\Delta t$  ( $\text{Re } z = -1, \text{Im } z = 0$ ). Отрицательные частоты спектра отображаются аналогично по часовой стрелке на нижней полуокружности. Точки  $\pm\omega_N$  совпадают, а при дальнейшем повышении или понижении частоты значения начинают повторяться в полном соответствии с периодичностью спектра дискретной функции. Проход по полной окружности соответствует одному периоду спектра, а любая гармоника спектра сигнала задается на

плоскости двумя точками, симметричными относительно оси абсцисс.

Сигналы и системы непрерывного времени очень часто описываются с помощью преобразования Лапласа. Если  $z = \exp(-s\Delta t)$ , где  $s = \sigma + j\omega$ , то

$$z = \exp(-(\sigma + j\omega)\Delta t) = \exp(-\sigma\Delta t) \exp(-j\omega\Delta t).$$

Следовательно,  $|z| = \exp(-\sigma\Delta t)$ ,  $\arg(z) = \omega\Delta t = 2\pi f\Delta t = 2\pi f/f_\Delta$ , где  $f_\Delta$  - частота дискретизации, при этом ось  $\omega$  отображается на z-плоскости единичной окружностью, правая сторона s-плоскости отображается внутрь окружности, а левая сторона - на внешнюю сторону окружности. При использовании символики  $z^{-1}$  отображение сторон s-плоскости на z-плоскости меняется местами.

## 8.2. ПРОСТРАНСТВО Z-ПОЛИНОМОВ

**Область сходимости.** Полином  $S(z)$  (8.1.1) называют z-образом или z-изображением функции  $s(k\Delta t)$ . Преобразование имеет смысл для области тех значений z, в которой ряд  $S(z)$  сходится, т.е. сумма ряда представляет собой аналитическую функцию переменной z, не имеющую полюсов и особых точек:

$$\sum_{k=-\infty}^{\infty} |s_k||z|^k < \infty$$

В общем случае, множества z, для которых полиномы  $S(z)$  сходится, образуют на z-плоскости определенные области, показанные на рис. 8.2.1.

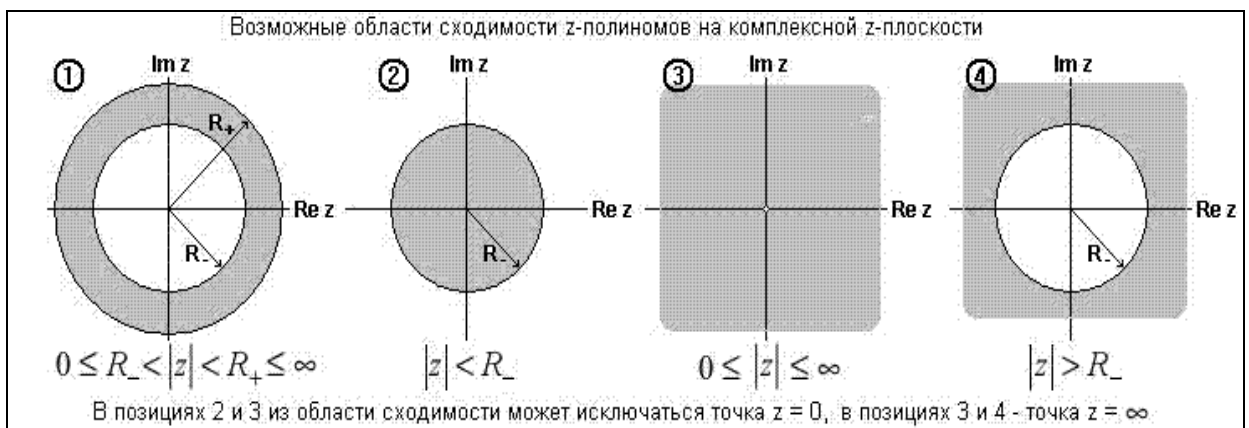


Рис. 8.2.1.

Из приведенной выше связи z-преобразования с преобразованием Фурье следует, что если функция  $s(t)$  имеет спектральное представление  $S(\omega)$ , то единичная окружность  $|z| = |\exp(-j\omega)| = 1$  обязательно должна входить в область сходимости полинома  $S(z)$ . И наоборот, если область сходимости полинома  $S(z)$  включает в себя единичную окружность, то дискретное преобразование Фурье функции  $s(t)$  - прообраз полинома  $S(z)$ ,

должно существовать, а в противном случае – нет. Последнее следует из того, что z-преобразование, являясь более общим случаем преобразования дискретных функций, может существовать и для функций, для которых не существует преобразования Фурье. Примером этого может служить функция единичного скачка:

$$u_n = 1, n \geq 0; \quad u_n = 0, n < 0.$$

Для преобразования Фурье функции  $u(n)$  не выполняется условие абсолютной суммируемости (энергия функции бесконечна). Но для z-преобразования имеем:

$$\sum_{k=-\infty}^{\infty} |u_k| |z|^k = \sum_{k=0}^{\infty} |z|^k < \infty, \text{ при } |z| < 1.$$

**Примеры z-преобразования** часто встречающихся на практике дискретных сигналов.

**Импульсы Кронекера.** В общем случае, для импульса Кронекера в произвольной точке числовой оси:

$$\delta(k-n) = 1 \text{ при } k=n, \quad \delta(k-n) = 0 \text{ при } k \neq n.$$

$$X_{\delta}(z) = \sum_{k=-\infty}^{\infty} \delta(k-n) z^k = z^n.$$

Для импульса Кронекера в нулевой точке соответственно  $X_{\delta}(z) = z^0 = 1$ . Ряд  $X_{\delta}(z)$  сходится на всей z-плоскости.

**Функция Хевисайда** (единичный скачок, причинная последовательность бесконечной длины, например, импульсный отклик рекурсивного интегрирующего фильтра).

$$x(k) = 0 \text{ при } k < 0, \quad x(k) = 1 \text{ при } k \geq 0.$$

$$X(z) = \sum_{k=0}^{\infty} z^k = z^k.$$

Ряд сходится при  $|z| < 1$ , при этом его сумма равна:

$$X(z) = 1/(1-z).$$

Z-преобразование действительно везде внутри круга единичного радиуса с центром в начале координат.

При использовании символики  $z^{-1}$ :

$$X(z) = 1/(1-z^{-1}) = z/(z-1), \quad |z| > 1.$$

На границе области аналитичности функция  $X(z)$  имеет один простой полюс при  $z=1$ .

**Экспоненциальная функция:**

$$x(k) = 0 \text{ при } k < 0, \quad x(k) = a^k \text{ при } k \geq 0.$$

$$X(z) = \sum_{k=-\infty}^{\infty} x(k) z^k = \sum_{k=0}^{\infty} a^k z^k = \sum_{k=0}^{\infty} (az)^k.$$

Как и в предыдущем случае, ряд сходится при  $|az| < 1$ , при этом:

$$X(z) = 1/(1-az), \quad |z| < 1/a.$$

При использовании символики  $z^{-1}$ :

$$X(z) = z/(z-a), \quad |z| > a.$$

**Комплексная экспонента:**

$$x(k) = \exp(j\omega k), \quad k \geq 0; \quad x(k) = 0, \quad k < 0.$$

$$X(z) = \sum_{k=0}^{\infty} \exp(j\omega k) z^k = \sum_{k=0}^{\infty} (z \exp(j\omega))^k = 1/(1-z \exp(j\omega)), \quad |z| < 1.$$

**Аналитическая форма z-образов** существует для z-преобразований, если возможно свертывание степенного ряда в аналитическое выражение. Выше, в примерах z-преобразования, уже приводилось приведение к аналитической форме z-образов функции

Хевисайда и экспоненциальной функции. Ниже в таблице приводится z-трансформация ряда распространенных функций, которые могут использоваться для прямого и обратного преобразования.

Таблица 8.2.1.

Функция $s(k)$ , $k \geq 0$	$z$ - образ $S(z)$	$z^{-1}$ - образ $S(z)$
$\beta$	$\beta / (1-z),  z  < 1$	$\beta z / (z-1),  z  > 1$
$\beta k$	$\beta z / (1-z)^2,  z  < 1$	$\beta z / (z-1)^2,  z  > 1$
$\beta k^2$	$\beta z (1+z) / (1-z)^3,  z  < 1$	$\beta z (z+1) / (z-1)^3,  z  > 1$
$\beta \alpha^k$	$\beta / (1 - z\alpha),  z  < 1/\alpha$	$\beta z / (z - \alpha),  z  > \alpha$
$\beta k \alpha^k$	$\beta \alpha z / (1 - z\alpha)^2,  z  < 1/\alpha$	$\beta \alpha z / (z - \alpha)^2,  z  > \alpha$
$\cos \alpha k$	$(1-z \cos \alpha) / (1-2z \cos \alpha+z^2),  z  < 1$	$z (z-\cos \alpha) / (z^2-2z \cos \alpha+1),  z  > 1$
$\sin \alpha k$	$z \sin \alpha / (1-2z \cos \alpha+z^2),  z  < 1$	$z \sin \alpha / (z^2-2z \cos \alpha+1),  z  > 1$
$\beta \exp(-\alpha k)$	$\beta / (1-z \exp(-\alpha)),  z  < 1/\exp(-\alpha)$	$\beta z / (z-\exp(-\alpha)),  z  > \exp(-\alpha)$
$\beta k \exp(-\alpha k)$	$\beta z \exp(-\alpha) / (1-z \exp(-\alpha))^2,  z  < 1/\exp(-\alpha)$	$\beta z \exp(-\alpha) / (z-\exp(-\alpha))^2,  z  > \exp(-\alpha)$

В таблице приведены преобразования как для символики  $z$ , так и для символики  $z^{-1}$  (по Гуревичу), которая иногда бывает удобней в некоторых математических операциях. Переход из одной символики в другую достаточно прост и выполняется заменой  $z$  в одной символике на  $1/z$  в другой.

### 8.3. СВОЙСТВА Z-ПРЕОБРАЗОВАНИЯ

Важнейшим свойством z-преобразования является свойство его единственности. Любая последовательность  $s(k)$  однозначно определяется z-изображением в области его сходимости, и наоборот, однозначно восстанавливается по z-изображению.

Без углубления в теорию, можно констатировать, что все свойства ДПФ действительны и для z-преобразования. Отметим некоторые из них.

**Линейность:** Если  $s(k) = a \cdot x(k) + b \cdot y(k)$ , то  $S(z) = aX(z) + bY(z)$ . Соответственно, z-преобразование допустимо только для анализа линейных систем и сигналов, удовлетворяющих принципу суперпозиции.

**Задержка** на  $n$  тактов:  $y(k) = x(k-n)$ .

$$Y(z) = \sum_{k=-\infty}^{\infty} y(k) z^k = \sum_{k=-\infty}^{\infty} x(k-n) z^k = z^n \sum_{k=-\infty}^{\infty} x(k-n) z^{k-n} = z^n \sum_{m=-\infty}^{\infty} x(m) z^m = z^n X(z).$$

Соответственно, умножение z-образа сигнала на множитель  $z^n$  вызывает сдвиг сигнала на  $n$  тактов дискретизации.

**Преобразование свертки.** При выполнении нерекурсивной цифровой фильтрации односторонними операторами фильтров:

$$s(k) = \sum_{n=0}^{\infty} h(n) y(k-n), \quad k = 0, 1, 2, \dots$$

Z-преобразование уравнения свертки:

$$\begin{aligned} S(z) &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} h(n) y(k-n) z^k = \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} h(n) z^n y(k-n) z^{k-n} = \\ &= \sum_{n=0}^{\infty} h(n) z^n \sum_{k=0}^{\infty} y(k-n) z^{k-n} = H(z) Y(z). \end{aligned}$$

Таким образом, свертка дискретных функций отображается произведением z-образов этих функций. Аналогично, для z-преобразования могут быть доказаны все известные теоремы о свойствах z-образов, что вполне естественно, т.к. при  $z = \exp(-j\omega)$  эти свойства полностью эквивалентны свойствам спектров функций.

**Разложение сигналов на блоки последовательной свертки.** Z-преобразование позволяет производить разложение сигналов и функций, например передаточных функций фильтров, на короткие составляющие операции свертки, для чего достаточно приравнять z-полином к нулю, найти его корни  $a_i$ , и переписать полином в виде произведения

двучленов:

$$S(z) = a_0(z-a_1)(z-a_2)\dots,$$

где  $a_0$ - последний отсчет сигнала (коэффициент при старшей степени  $z$ ).

Но произведению в  $z$ -области соответствует свертка в координатной области, и при обратном преобразовании двучлены  $(z-a_i)$  превращаются в двухточечные диполи  $\{-a_i, 1\}$ , а сигнал длиной  $N$  представляется сверткой  $(N-1)$  диполей:

$$s_k = a_0 \{-a_1, 1\} * \{-a_2, 1\} * \{-a_3, 1\} * \dots$$

**Пример.**  $s_k = \{1.4464, -2.32, 3.37, -3, 1\}$ .  $S(z) = z^4 - 3z^3 + 3.37z^2 - 2.32z + 1.4464$ .  $a_0 = 1$ .

Корни полинома  $S(z)$ :  $a_1 = 0.8 + 0.8j$ ,  $a_2 = 0.8 - 0.8j$ ,  $a_3 = 0.7 + 0.8j$ ,  $a_4 = 0.7 - 0.8j$ ,

$$S(z) = (z - 0.8 - 0.8j)(z - 0.8 + 0.8j)(z - 0.7 - 0.8j)(z - 0.7 + 0.8j).$$

Корни полинома представлены на  $z$ -плоскости на рис. 8.1.1. Корни полинома комплексные и четыре двучлена в координатной области также будут комплексными. Но они являются сопряженными, и для получения вещественных функций следует перемножить сопряженные двучлены и получить биквадратные блоки:

$$S(z) = (z^2 - 1.4z + 1.13)(z^2 - 1.6z + 1.28).$$

При переходе в координатную область:  $s_k = \{1.13, -1.4, 1\} * \{1.28, -1.6, 1\}$ .

Таким образом, исходный сигнал разложен на свертку двух трехчленных сигналов (функций).

**Дифференцирование.** Если имеем  $s(k) \leftrightarrow S(z)$ , то  $z$ -образ функции  $s'(k)$  можно найти, продифференцировав  $S(z)$ , что бывает полезно для вычисления обратного  $z$ -преобразования функций  $S(z)$  с полюсами высокого порядка:

$$s'(k) \leftrightarrow z \, dS(z)/dz.$$

#### 8.4. ОБРАТНОЕ Z-ПРЕОБРАЗОВАНИЕ

**Методы преобразования.** Обратное  $z$ -преобразование позволяет восстанавливать дискретную функцию по ее  $z$ -образу. Оно широко используется, например, при определении импульсных характеристик рекурсивных цифровых фильтров. В символической форме:

$$x(k) = TZ^{-1}[X(z)].$$

На практике  $X(z)$  в процессе расчетов обычно выражается через отношение двух многочленов от  $z$ :

$$X(z) = (b_0 + b_1z + b_2z^2 + \dots + b_Nz^N) / (a_0 + a_1z + a_2z^2 + \dots + a_Mz^M) = \quad (8.4.1)$$

$$= x(0) + x(1)z + x(2)z^2 + \dots \quad (8.4.1')$$

Самые распространенные методы обратного преобразования из этой формы  $X(z)$ :

- Преобразование интегрированием по контуру (метод вычетов).
- Метод разложения на элементарные дроби.
- Метод разложения в степенной ряд.

Метод разложения в степенной ряд наиболее прост и пригоден для выполнения на компьютерах, но он не дает решения в аналитической форме. При задании большого числа точек обратного преобразования требуется также следить за возможным нарастанием числовых ошибок вследствие рекурсии его алгоритма.

Два первых метода позволяют получать результаты в аналитическом виде, но требуют вычисления полюсов функции  $X(z)$ , что может представлять трудности при высоком порядке функции. При высоких порядках полюсов потребуются также дифференцирование соответствующих порядков.

**Преобразование интегрированием по контуру** относится к числу математически строгих методов. Оно выполняется интегрированием по произвольному замкнутому контуру  $C$ , расположенному в области сходимости и окружающему все особые точки (нули и полюсы)  $z$ -образа. Интегрирование удобнее выполнять над полюсами, расположенными внутри контура, включающего центр системы координат, т.е. в символике  $z^{-1}$ . В этой символике мы и будем рассматривать данный параграф. Контурный интеграл обратного преобразования:



$$s_k = (1/2\pi j) \oint_c S(z) z^{k-1} dz. \quad (8.4.2)$$

Согласно теореме Коши о вычетах, интеграл (8.4.2) равен сумме вычетов (Res) подынтегральной функции относительно всех полюсов этой функции, лежащих внутри контура интегрирования. Каждый вычет связан с определенным полюсом  $p_k$ :

$$\text{Res}[F(z), p_k] = \frac{1}{(m-1)!} \cdot \frac{d^{m-1}}{dz^{m-1}} [(z-p_k) F(z)] \quad \text{при } z=p_k. \quad (8.4.3)$$

где  $F(z) = z^{k-1} S(z)$ ,  $m$  – порядок полюса в точке  $p_k$ . Для простого полюса:

$$\text{Res}[F(z), p_k] = (z-p_k) F(z) = (z-p_k) z^{k-1} S(z) \quad \text{при } z=p_k. \quad (8.4.3')$$

**Пример.** Z-образ функции:  $X(z) = z^2 / (z-0.5)(z-1)^2$ .

$x(k) = \text{Res}[F(z), p_1] + \text{Res}[F(z), p_2]$ .  $F(z) = z^{k-1} X(z) = z^{k+1} / (z-0.5)(z-1)^2$ .

Функция  $F(z)$  имеет простой полюс  $p_1 = 0.5$  и полюс второго порядка  $p_2 = 1$ .

$\text{Res}[F(z), 0.5] = (z-0.5) z^{k+1} / (z-0.5)(z-1)^2 = z^{k+1} / (z-1)^2 |_{z=0.5} = 0.5 (0.5)^k / (0.5)^2 = 2(0.5)^k$ .

$\text{Res}[F(z), 1] = \frac{d}{dz} [(z-1)^2 z^{k+1} / (z-0.5)(z-1)^2] = [(z-0.5)(k+1)z^k - z^{k+1}] / (z-0.5)^2 |_{z=1} = 2(k-1)$ .

**Результат:**  $x(k) = 2[(k-1) + (0.5)^k]$ .

**Преобразование разложением на дроби.** В этом методе  $z$ -образ (8.4.1) раскладывается на рациональные простые дроби с последующим почленным обратным преобразованием с помощью таблицы. Наиболее просто это выполняется, если функция  $S(z)$  может быть разложена по степеням  $z$  в символикe  $z^{-1}$ , т.е. представлена в следующем виде:

$$S(z) = s(0) + s(1) z^{-1} + s(2) z^{-2} + \dots$$

Соответственно, в выражении (8.4.1) отношение многочленов также должно быть в символикe  $z^{-1}$ . Если полюсы  $S(z)$  первого порядка и  $N = M$ , то (8.4.1) можно разложить на следующую сумму:

$$S(z) = B_0 + C_1/(1-p_1 z^{-1}) + C_2/(1-p_2 z^{-2}) + \dots + C_M/(1-p_M z^{-M}) = \\ B_0 + C_1 z/(z-p_1) + C_2 z/(z-p_2) + \dots + C_M z/(z-p_M) = B_0 + \sum_{k=1}^M C_k z/(z-p_k). \quad (8.4.4)$$

$$B_0 = b_N / a_N.$$

где  $C_k$  – коэффициенты элементарных дробей, которые являются вычетами функции  $S(z)$ .

Для вычисления коэффициентов  $C_k$  умножим левую и правую стороны выражения (8.4.4) на  $(z-p_k)/z$  и положим  $z=p_k$ , при этом в правой части за счет множителя  $(z-p_k)=0$  при  $z=p_k$  обнуляются все члены суммы кроме члена с  $C_k$  данного полюса, а в левой остается произведение  $S(z)(z-p_k)/z$ , что позволяет вычислить значения  $C_k$ :

$$C_k = S(z)(z-p_k)/z |_{z=p_k} \quad (8.4.5)$$

Если в (8.4.1)  $N < M$ , то значение  $B_0$  равно нулю. Если функция  $S(z)$  в точке  $z=p_k$  имеет полюс  $m$ -ного порядка, то коэффициент  $C_k$  заменяется суммой коэффициентов:

$$\sum_{i=1}^m D_i / (z-p_k)^i, \quad (8.4.6)$$

$$D_i = \frac{1}{(m-i)!} \cdot \frac{d^{m-i}}{dz^{m-i}} [X(z) (z-p_k)^m / z], \quad \text{при } z=p_k. \quad (8.4.7)$$

**Пример.** Повторим пример преобразования данным способом  $z$ -образа функции  $X(z) = z^2 / (z-0.5)(z-1)^2$ , использованного в предыдущем примере. Функция имеет простой полюс  $p_1 = 0.5$  и полюс второго порядка  $p_2 = 1$ .

$X(z) = Cz/(z-0.5) + D_1 z/(z-1) + D_2 z/(z-1)^2$ .

$C = z/(z-1)^2 = 0.5/(0.5-1)^2 = 2$ .

$D_1 = \frac{d}{dz} [(z-1)^2 X(z)/z] = \frac{d}{dz} [z / (z-0.5)] |_{z=1} = -2$ .

$D_2 = (z-1)^2 X(z)/z = z/(z-0.5) |_{z=1} = 2$ .

$X(z) = 2z/(z-0.5) + D_1 z/(z-1) + D_2 z/(z-1)^2$ .

Обратное преобразование каждой простой дроби выполним по таблице 8.2.1.

**Результат:**  $x(k) = 2(0.5)^k - 2 + 2k = 2[(k-1) + (0.5)^k]$ . Результат аналогичен методу вычетов.

Если z-изображение имеет вид дробно-рациональной функции, то разложение на простые дроби с последующим применением таблицы соответствий обычно труда не представляет. Так, например:

$$S(z) = (b_0 + b_1 z^{-1} + b_2 z^{-2}) / (1 - az^{-1}) = b_0 / (1 - az^{-1}) + b_1 z^{-1} / (1 - az^{-1}) + b_2 z^{-2} / (1 - az^{-1}).$$

По таблице соответствия:

$$X(z) = 1 / (1 - az^{-1}) \rightarrow x(k) = a^k.$$

Отсюда, с учетом линейности преобразования и свойства задержки:

$$x(k) = b_0 a^k + b_1 a^{k-1} + b_2 a^{k-2}.$$

При преобразовании функций со знаменателями более высоких порядков предварительно следует найти полюса функции. Например, для многочлена второго порядка с полюсами  $p_1$  и  $p_2$ :

$$S(z) = 1 / (1 - a_1 z^{-1} + a_2 z^{-2}) = 1 / [(1 - p_1 z^{-1})(1 - p_2 z^{-1})].$$

Представим  $S(z)$  в виде суммы дробей с неизвестными коэффициентами  $b_1$  и  $b_2$ :

$$S(z) = b_1 / (1 - p_1 z^{-1}) + b_2 / (1 - p_2 z^{-1}) = (b_1 - b_1 p_2 z^{-1} + b_2 - b_2 p_1 z^{-1}) / [(1 - p_1 z^{-1})(1 - p_2 z^{-1})].$$

При равенстве знаменателей в этих двух выражениях должны быть равны и числители:

$$(b_1 + b_2) - (b_1 p_2 + b_2 p_1) z^{-1} = 1,$$

а это обеспечивается равенством коэффициентов при одинаковых степенях  $z$ . Отсюда получаем систему уравнений:

$$b_1 + b_2 = 1.$$

$$b_1 p_2 + b_2 p_1 = 0.$$

Решая эту систему уравнений, находим значения коэффициентов  $b_1$  и  $b_2$ , подставляем коэффициенты в  $S(z)$ , выраженное в виде суммы дробей, и по таблице соответствия переводим дроби во временные функции.

**Метод степенных рядов.** Выражение (8.4.1) можно разложить непосредственно в степенной ряд (8.4.1') путем деления в столбик, для чего числитель и знаменатель функции выражаются предварительно через нарастающий или уменьшающийся показатель степени  $z$ . Обратное z-преобразование степенного ряда очевидно.

**Пример нарастающей степени z.**  $X(z) = (1 + 2z + z^2) / (1 - z + 0.4z^2)$ .

$$\begin{array}{r} 1 + 2z + z^2 \quad \quad \quad | 1 - z + 0.4z^2 \\ \underline{1 - z + 0.4z^2} \quad \quad \quad 1 + 3z + 3.6z^2 + 2.4z^3 + 0.96z^4 + \dots \text{Ряд может быть бесконечным.} \\ 3z + 0.6z^2 \\ \underline{3z - 3z^2 + 1.2z^3} \\ 3.6z^2 = 1.2z^3 \\ \underline{3.6z^2 - 3.6z^3 + 1.44z^4} \\ 2.4z^3 = 1.44z^4 \\ \underline{2.4z^3 - 2.4z^4 + 0.96z^5} \\ 0.96z^4 - 0.96z^5 \\ \underline{0.96z^4 - 0.96z^5 + 0.384z^6, \text{ и т.д.}} \end{array}$$

Обратное преобразование выполняется путем идентификации коэффициентов степеней при  $z^k$  с k-отсчетами функции:  $x(k) = \{1, 3, 3.6, 2.4, 0.96, \dots\}$ .

**Пример уменьшающихся номеров степени z.**  $X(z) = (1 + 2z + z^2) / (1 - z + 0.4z^2) \rightarrow$  (деление на  $z^N$  числителя и знаменателя полинома)  $\rightarrow (z^2 + 2z + 1) / (z^2 - z + 0.4)$ .

$$\begin{array}{r} z^2 + 2z + 1 \quad \quad \quad | z^2 - z + 0.4 \\ \underline{z^2 - z^1 + 0.4} \quad \quad \quad 1 + 3z + 3.6z^2 + 2.4z^3 + 0.96z^4 + \dots \text{Результат тот же.} \\ 3z^{-1} + 0.6 \\ \underline{3z^{-1} - 3 + 1.2z} \\ 3.6 = 1.2z \\ \underline{3.6 - 3.6z + 1.44z^2} \\ 2.4z = 1.44z^2 \\ \underline{2.4z - 2.4z^2 + 0.96z^3} \\ 0.96z^2 - 0.96z^3 \\ \underline{0.96z^2 - 0.96z^3 + 0.384z^4, \text{ и т.д.}} \end{array}$$

Метод деления полинома (8.4.1) можно выполнять рекурсивно:

$$x(0) = b_0 / a_0,$$

$$x(1) = (b_1 - x(0) a_1) / a_0,$$

$$x(2) = (b_2 - x(1) a_1 - x(0) a_2) / a_0,$$

...

$$x(n) = (b_n - \sum_{i=1}^n (x(n-i) a_i) / a_0, \quad n = 1, 2, 3, \dots$$

### 8.5. ПРИМЕНЕНИЕ Z – ПРЕОБРАЗОВАНИЯ

**Описание дискретных систем** обработки сигналов с помощью нулей и полюсов - наиболее широкая область использования z-преобразования. Степенной полином передаточной функции системы вида (8.4.1) с нулями  $n_i$  числителя и полюсами  $p_j$  знаменателя всегда может быть представлен в виде произведения сомножителей:

$$H(z) = K \prod_{i=1}^N (z - n_i) / \prod_{j=1}^M (z - p_j), \quad (8.5.1)$$

где  $K$  – коэффициент передачи (усиления) входного сигнала. Полюсы и нули  $H(z)$  могут быть действительными и комплексными, при этом для обеспечения действительных значений коэффициентов  $a_i$  и  $b_j$  в (8.4.1) комплексные коэффициенты должны быть представлены комплексно сопряженными парами.

**Геометрическая оценка АЧХ и ФЧХ системы.** Информацию, содержащуюся в  $H(z)$ , удобно отображать в виде положения нулей (кружками) и полюсов (крестиками) на  $z$ -плоскости. Диаграмма нулей и полюсов наглядно отображает свойства системы и ее устойчивость. Для устойчивых систем все полюсы должны находиться за пределами единичной окружности (внутри окружности при символикe  $z^{-1}$ ) или совпадать с нулями на единичной окружности. На положение нулей ограничений не существует.

По известной диаграмме нулей и полюсов может быть выполнена геометрическая оценка частотной характеристики системы. При  $z = \exp(-j\omega\Delta t)$  единичная окружность  $|z|=1$  отображает частотную ось характеристики главного частотного диапазона от  $\omega = 0$  (при  $z=1$ ) до  $2\pi$  (при  $z=-1$ ). Каждой точке  $z_s = \exp(-j\omega_s\Delta t)$  может быть поставлен в соответствие вектор  $(z_s - n_i)$  на  $i$ -нуль, модуль которого  $U_i = |z_s - n_i|$  отображает расстояние от  $z_s$  до  $i$ -нуля, а аргумент  $\phi_i = \arg(z_s - n_i)$  - фазовый угол из  $z_s$  на  $i$ -нуль, а равно и вектор  $(z_s - p_j)$  на  $j$ -полюс с соответствующим расстоянием  $V_j = |z_s - p_j|$  и фазовым углом  $\phi_j = \arg(z_s - p_j)$ . При этом амплитудная и фазовая характеристики системы могут быть оценены по выражениям при перемещении точки  $\omega_s$  по единичной окружности:

$$|H(\omega)| = \prod_{i=1}^N U_i / \prod_{j=1}^M V_j, \quad (8.5.2)$$

$$\arg(H(\omega)) = \sum_i \phi_i - \sum_j \phi_j. \quad (8.5.3)$$

По (8.5.2) нетрудно сделать заключение, что наибольшее влияние на изменение АЧХ по частоте оказывают нули и полюсы, расположенные ближе к единичной окружности. При расположении нуля непосредственно на окружности гармоника  $\omega_s$  в этой точке полностью обнуляется. И, наоборот, при перемещении  $\omega_s$  к полюсу, близкому к единичной окружности, происходит резкое нарастание коэффициента усиления системы.

**Вычисление частотной характеристики с помощью БПФ.** Так как частотная характеристика дискретной системы – это Фурье образ ее импульсной характеристики, то для систем, описанных в общей форме (8.4.1), сначала производится разложение  $H(z)$  в степенной ряд (8.4.1'), над коэффициентами которого и выполняется БПФ. Гладкость (разрешение по частоте  $\Delta f = 1/(N\Delta t)$ ) будет определяться количеством коэффициентов степенного ряда и при необходимости может увеличиваться дополнением ряда нулями.

Альтернативный способ – вычисление БПФ непосредственно коэффициентов  $b_n$  числителя и  $a_m$  знаменателя выражения (8.4.1) с последующим алгебраическим делением

$B(k)/A(k)$  результатов БПФ. Количество коэффициентов  $b_n$  и  $a_n$  в (8.4.1) обычно невелико и для получения достаточно гладких частотных характеристик их продлевают нулями до необходимого значения  $N = 1/(\Delta t \Delta f)$ .

**Анализ устойчивости систем** выполняется для рекурсивных систем с бесконечной импульсной характеристикой (БИХ-систем). Такие системы описываются либо непосредственно в виде разностного уравнения, либо передаточной функцией в виде  $z$ -образа импульсной характеристики или разностного уравнения. Общее условие устойчивости импульсной характеристики системы:

$$\sum_{k=-\infty}^{\infty} |h(k)| < \infty.$$

Для рекурсивных систем начальный индекс суммирования равен нулю. Практически это означает, что любой ограниченный входной сигнал в устойчивой системе порождает ограниченный выходной сигнал.

В устойчивой системе все полюсы передаточной функции  $H(z)$  должны находиться за границами единичной окружности  $z = \exp(-j\omega\Delta t)$  (внутри окружности при символике  $z^{-1}$ ). Система с полюсом на единичной окружности считается потенциально неустойчивой, даже если во входном сигнале нет гармоник с частотой, соответствующей положению данного полюса на окружности. Это определяется тем, что в соответствии с (8.5.1) коэффициент усиления системы в точке полюса равен бесконечности и любой бесконечно малый сигнал на этой частоте даст бесконечно большой сигнал на выходе. Для практических систем понятия бесконечности не существует и можно пытаться принять определенные меры для исключения таких критических частот. Так, например, в интегрирующих системах полюс находится на нулевой частоте и из входного сигнала можно исключить постоянную составляющую, но при этом изменяется и характер интегрирования (интегрируются только динамические составляющие входного сигнала). Следует также учитывать, что во входных сигналах обычно всегда присутствует статистический шум, наблюдаются скачки, присутствует шум квантования и т.п. эффекты с непрерывным частотным спектром, которые могут приводить к огромным ошибкам при обработке данных в потенциально неустойчивых системах. Практически осуществимый способ повышения устойчивости систем – компенсировать полюсы на окружности нулями в этих же точках, но это может приводить к существенному изменению частотной характеристики системы.

Оценку устойчивости рекурсивной системы можно проводить и по виду ее импульсной характеристики (вычислением обратного  $z$ -преобразования или подачей импульса Кронекера на вход системы). Если значения коэффициентов увеличиваются по мере роста номеров – система неустойчива. Если они очень медленно уменьшаются – система устойчива минимально, имеет большое время установления рабочего режима, и при определенных условиях может давать большие погрешности в обрабатываемых данных.

**Связь разностных уравнений и передаточных функций** рекурсивных систем. Стандартная запись разностного уравнения системы (связи входного воздействия  $x(k)$  и выходного сигнала  $y(k)$  при известных постоянных параметрах нерекурсивной  $b_n$  и рекурсивной  $a_m$  трансформации сигналов):

$$y(k) = \sum_{n=0}^N b_n x(k-n) - \sum_{m=1}^M a_m y(k-m). \quad (8.5.4)$$

От разностного уравнения с использованием свойства задержки  $z$ -преобразования

$$b_n x(k) \leftrightarrow b_n X(z),$$

$$b_n x(k-n) \leftrightarrow b_n z^n X(z),$$

нетрудно перейти к  $z$ -образу разностного уравнения системы:

$$Y(z) = \sum_{n=0}^N b_n X(z) z^n - \sum_{m=1}^M a_m Y(z) z^m. \quad (8.5.5)$$

Отсюда, передаточная функция системы:

$$Y(z) \left(1 + \sum_{m=1}^M a_m z^m\right) = \sum_{n=0}^N b_n X(z) z^n.$$

$$H(z) = Y(z) / X(z) = \sum_{n=0}^N b_n z^n / \left(1 + \sum_{m=1}^M a_m z^m\right). \quad (8.5.6)$$

И, наоборот, при приведении выражения (8.4.1) к виду (8.5.6) (нормировкой на  $a_0$ ) можно без дальнейших преобразований переходить к выражению (8.5.4).

**Пример.** Передаточная функция:  $H(z) = 2(1-z) / (2+z)$ . Определить алгоритм вычислений.

$$H(z) = Y(z)/X(z) = (1-z) / (1+0.5z).$$

$$Y(z) + 0.5 z Y(z) = X(z) - z X(z).$$

$$y(k) + 0.5 y(k-1) = x(k) - x(k-1)$$

**Результат:**  $y(k) = x(k) - x(k-1) - 0.5 y(k-1)$

## Тема 9. РЕКУРСИВНЫЕ ЦИФРОВЫЕ ФИЛЬТРЫ

Высококачественные частотные нерекурсивные цифровые фильтры (НЦФ) имеют, как правило, большую ширину окна (многочленный оператор фильтра). Чем меньше допустимая ширина переходной зоны частотной характеристики фильтра между полосами пропускания и подавления, тем больше окно фильтра. Альтернативное решение - применение рекурсивных цифровых фильтров (РЦФ), для которых количество коэффициентов фильтра может быть сокращено на несколько порядков по сравнению с НЦФ.

Рекурсивные фильтры имеют определенную "память" по значениям предыдущих отсчетов, которая, в пределе, может быть бесконечной. С учетом этого фактора рекурсивные фильтры получили название фильтров с бесконечной импульсной характеристикой (БИХ-фильтров), в отличие от нерекурсивных фильтров, всегда имеющих конечную импульсную характеристику (КИХ-фильтры). Реакция рекурсивного фильтра на сигнал с учетом "памяти" исключает возможность создания фильтров с четным импульсным откликом, и частотные характеристики рекурсивных фильтров всегда являются комплексными. Проектирование рекурсивных частотных фильтров с заданными частотными характеристиками осуществляется с использованием  $z$ -преобразований.

### 9.1. ПРИНЦИПЫ РЕКУРСИВНОЙ ФИЛЬТРАЦИИ.

**Конструкция РЦФ** отображается в  $z$ -образе передаточной функции фильтра в виде отношения двух многочленов:

$$H(z) = H_0 + H_1 z + H_2 z^2 + \dots = B(z) / [1 + A(z)], \quad (9.1.1)$$

где:  $B(z) = B_0 + B_1 z + B_2 z^2 + \dots + B_N z^N$ ,  $A(z) = A_1 z + A_2 z^2 + \dots + A_M z^M$ .

Естественно, что переход на РЦФ имеет смысл только в том случае, если степень многочленов  $A(z)$  и  $B(z)$  во много раз меньше степени многочлена  $H(z)$  прямого  $z$ -преобразования импульсной реакции фильтра. При  $z$ -образе входных данных  $X(z)$ , на выходе РЦФ имеем:

$$Y(z) = H(z)X(z) = X(z)B(z) / [1 + A(z)],$$

$$Y(z)[1 + A(z)] = Y(z) + Y(z)A(z) = X(z)B(z),$$

$$Y(z) = X(z)B(z) - Y(z)A(z). \quad (9.1.2)$$

При обратном  $z$ -преобразовании выражения (9.1.2) получаем уравнение рекурсивной цифровой фильтрации:

$$y_k = \sum_{n=0}^N b_n x_{k-n} - \sum_{m=1}^M a_m y_{k-m}. \quad (9.1.3)$$

Рекурсивная фильтрация требует задания начальных условий как по  $x_k$ , так и по  $y_k$

при  $k < 0$ . Схема рекурсивной фильтрации приведена на рис. 9.1.1.

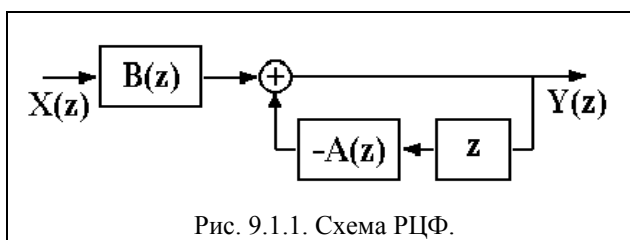


Рис. 9.1.1. Схема РЦФ.

Как следует из выражения (9.1.3), при вычислении значения  $y_k$  текущей точки используются предыдущие вычисленные значения  $y_{k-m}$ , ( $m > 0$ ), что и определяет принцип рекурсии - фильтрации с обратной связью. Другой особенностью РЦФ является их односторонность и физическая

реализуемость в реальном масштабе времени. При машинной обработке данных многочлен  $B(z)$  передаточной функции фильтра может реализоваться и в двухстороннем варианте.

Одно из важнейших свойств рекурсивных фильтров - возможность получения узких переходных зон при конструировании частотных фильтров, так как функция  $H(z)$  фильтра может резко изменяться при приближении к нулю (но не нулевого) многочлена в знаменателе (9.1.1).

Рекурсивная фильтрация требует более высокой точности вычислений по сравнению с нерекурсивной, т.к. использование предыдущих выходных отсчетов для текущих вычислений может приводить к накоплению ошибок. Особое значение это имеет для фильтров с передаточными функциями высоких порядков ( $M > 3$ ), которые чувствительны к эффектам конечной разрядности. Такие фильтры, как правило, разбиваются на фрагменты - звенья второго и/или первого порядка, и реализуются в каскадной или в параллельной форме.

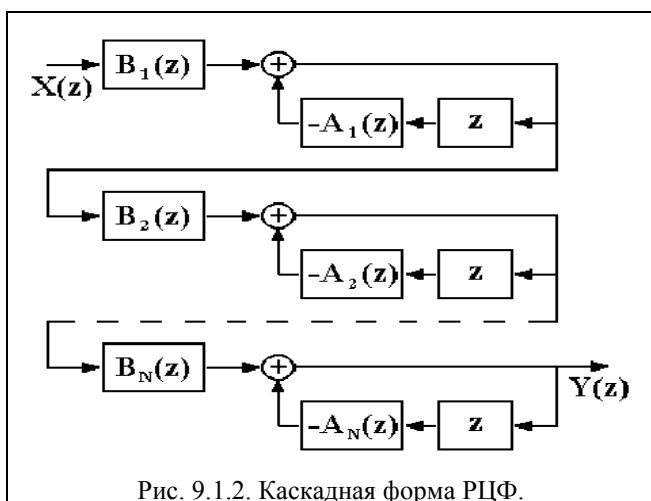


Рис. 9.1.2. Каскадная форма РЦФ.

**Каскадная форма.** Находятся корни многочленов  $A(z)$ ,  $B(z)$  и производится разложение  $H(z)$ :

$$H(z) = G \frac{B_1(z)}{A_1(z)} \frac{B_2(z)}{A_2(z)} \dots \frac{B_N(z)}{A_N(z)}, \quad (9.1.4)$$

где  $G$  - масштабный множитель. Это позволяет применять каскадное построение фильтров, показанное на рис. 9.1.2, в котором:

$$H(z) = G H_1(z) H_2(z) \dots H_N(z),$$

$$H_n(z) = B_n(z)/A_n(z). \quad (9.1.5)$$

Функции  $A_n(z)$  и  $B_n(z)$  обычно

представляются в виде биквадратных блоков (фильтров второго порядка):

$$B_n(z) = b_{n,0} + b_{n,1} z + b_{n,2} z^2, \quad A_n(z) = 1 + a_{n,1} z + a_{n,2} z^2.$$

В принципе, порядок расположения блоков в каскадной форме, равно как и порядок множителей  $B(z)$  и  $A(z)$  в числителе и знаменателе функции (9.1.4), значения не имеет. Однако следует учитывать, что полюса знаменателя, близкие к единичной окружности на  $z$ -плоскости (близкие по модулю к 1), формируют большие коэффициенты усиления на соответствующих частотах (в блоках, в которых они находятся), и при обработке сигналов могут вызывать переполнение разрядов числовых ячеек этих блоков, если их разрядность ограничена. С учетом этого при формировании каскадов желательно объединять в пары  $B_i(z)/A_i(z)$  нули и полюса, близкие по модулю к 1, и располагать их в концевые блоки каскадной схемы. Такое комбинирование полезно также с позиций наилучшего отношения сигнал/шум в выходном сигнале.

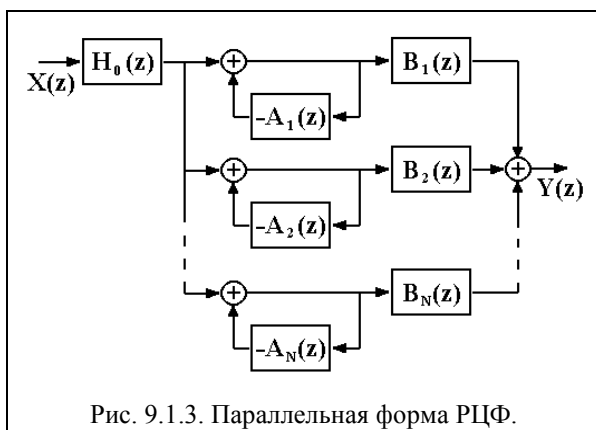


Рис. 9.1.3. Параллельная форма РЦФ.

**Параллельная форма.** Функция  $H(z)$  разлагается на элементарные дроби:

$$H(z) = H_0(z) \sum_{n=0}^N B_n(z) / [1 + A_n(z)],$$

что дает параллельную форму фильтра, показанную на рис. 9.1.3. Параллельная конструкция фильтра применяется реже каскадной, хотя это может объясняться и тем, что в аналоговых фильтрах, исторически предшествовавших цифровым фильтрам, теоретическая база анализа и синтеза каскадных рекурсивных фильтров получила

детальное развитие.

**Стандартные блоки рекурсивных фильтров** обычно реализуются биквадратными звеньями в канонической форме, которая имеет минимальное количество элементов задержки. Уравнения звена:

$$v(k) = x(k) - \sum_{n=1}^2 a(n) v(k-n), \quad y(k) = \sum_{n=0}^2 b(n) v(k-n). \quad (9.1.6)$$

Функциональная схема реализации звена приведена на рис. 9.1.4.

Вторая форма реализации – по уравнению (9.1.5) в прямой форме, приведенная на рис. 9.1.5:

$$y(k) = \sum_{n=0}^2 b(n)x(k-n) - \sum_{n=1}^2 a(n)y(k-n). \quad (9.1.7)$$

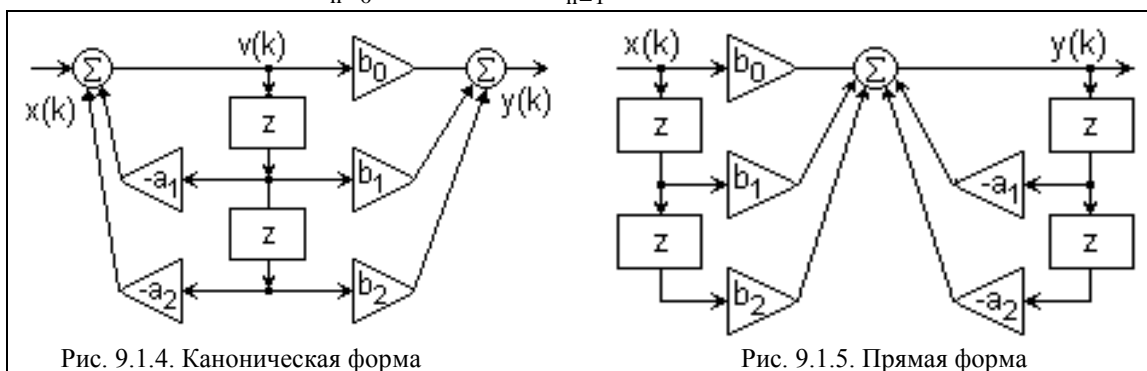


Рис. 9.1.4. Каноническая форма

Рис. 9.1.5. Прямая форма

При определенных условиях прямая форма лучше канонической с точки зрения шумовых характеристик.

При нулевых значениях коэффициентов  $a_2$  и  $b_2$  звенья второго порядка превращаются в звенья первого порядка.

**Устранение сдвига фазы.** Рекурсивные фильтры являются фазосдвигающими фильтрами. Если требуется обеспечить нулевой фазовый сдвиг, то операция фильтрации производится дважды, в прямом и обратном направлении числовой последовательности массива данных, при этом амплитудно-частотная характеристика (АЧХ) фильтрации будет равна  $|H(\omega)|^2$  фильтра, что необходимо учитывать при конструировании фильтра.

## 9.2. РАЗРАБОТКА РЕКУРСИВНЫХ ЦИФРОВЫХ ФИЛЬТРОВ

Процесс проектирования рекурсивного фильтра заключается в задании необходимой передаточной характеристики фильтра в частотной области и ее аппроксимации с определенной точностью какой-либо непрерывной передаточной функцией, с последующим  $z$ -преобразованием для перехода в  $z$ -область. Первые две операции хорошо отработаны в теории аналоговой фильтрации сигналов, что позволяет использовать для проектирования цифровых фильтров большой справочный материал по аналоговым фильтрам. Последняя операция является специфичной для цифровых

фильтров.

Для алгебраического преобразования непрерывной передаточной функции в многочлен по  $z$  используется билинейное преобразование, известное в теории комплексных переменных под названием дробно-линейного преобразования.

**Этапы разработки рекурсивных фильтров** включают:

1. Задание частотной характеристики или передаточной функции фильтра.
2. Аппроксимация и расчет коэффициентов  $b(n)$  и  $a(m)$  передаточной функции фильтра (9.1.3). Этот этап может выполняться четырьмя методами:
  - Метод размещения нулей и полюсов на комплексной  $z$ -плоскости.
  - Метод инвариантного преобразования импульсной характеристики.
  - Согласованное  $z$ -преобразование.
  - Билинейное  $z$ -преобразование.
3. Выбор структуры реализации фильтра – параллельная или каскадная, блоками второго и/или первого порядка.
4. Программное или аппаратное обеспечение реализации фильтра.

**Метод размещения нулей и полюсов** применяется при разработке простых фильтров с ограниченным количеством нулей и полюсов, если параметры фильтра не обязательно задавать точно. Амплитудная характеристики системы может быть оценена при перемещении точки  $\omega_s$  по единичной окружности  $\exp(-j\omega_s\Delta t)$  по выражению:

$$|H(\omega)| = \prod_{i=1}^N U_i / \prod_{j=1}^M V_j, \quad (9.2.1)$$

Каждой точке  $z_s = \exp(-j\omega_s\Delta t)$  может быть поставлен в соответствие вектор  $(z_s - n_i)$  на  $n_i$ -нуль, модуль которого  $U_i = |(z_s - n_i)|$  отображает расстояние от  $z_s$  до  $i$ -нуля, а равно и вектор  $(z_s - p_j)$  на  $p_j$ -полюс с соответствующим расстоянием  $V_j = |(z_s - p_j)|$ . Наибольшее влияние на изменение АЧХ по частоте оказывают нули и полюсы, расположенные ближе к единичной окружности. При расположении нуля непосредственно на окружности гармоника  $\omega_s$  в этой точке полностью обнуляется (коэффициент передачи фильтра равен нулю). И, наоборот, при перемещении  $\omega_s$  к полюсу, близкому к единичной окружности, происходит резкое нарастание коэффициента усиления системы.

В качестве иллюстрации метода выполним расчет фильтра со следующими параметрами:

1. Полная режекция сигнала на частотах 0 и 250 Гц.
2. Полоса пропускания с центром на  $f_p = 125$  Гц с шириной полосы по уровню 3 дБ  $\Delta f = 10$  Гц.
3. Частота дискретизации данных  $f_\Delta = 500$  Гц.

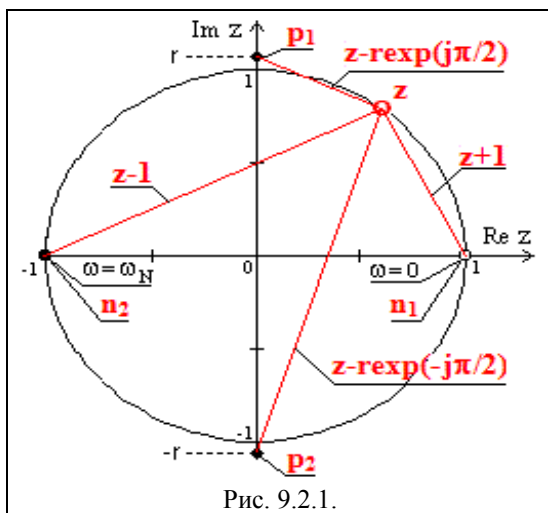


Рис. 9.2.1.

При частоте дискретизации 500 Гц интервал временной дискретизации  $\Delta t = 1/f_\Delta$ , а частота Найквиста  $f_N = 1/2\Delta t = 250$  Гц. Соответственно, нули передаточной функции располагаются в точках  $n_1 = \exp(-j2\pi \cdot 0 \cdot \Delta t) = 1$  и  $n_2 = \exp(-j2\pi \cdot 250 \cdot \Delta t) = -1$  (рис. 9.2.1). Угол из начала координат  $z$ -плоскости на полюс  $p$  с учетом его сопряженности для получения действительных коэффициентов  $\pm 180^\circ \cdot f_p/f_N = \pm 90^\circ$ . Значение радиуса  $r$  до полюса определяет ширину полосы пропускания и в первом приближении (при  $0 < r < 1.1$ ) оценивается по выражению:  $r \approx 1 + (\Delta f/f_\Delta)\pi$ .  $r \approx 1.063$ .

Передаточная функция:

$$H(z) = (z-1)(z+1) / [(z-r \exp(j\pi/2))(z-r \exp(-j\pi/2))] =$$



$$= (z^2-1)/(z^2+r^2) = Y(z)/X(z).$$

$$z^2Y(z)+r^2Y(z) = z^2X(z)-X(z). \quad Y(z) = [z^2X(z)-X(z)] / r^2.$$

Алгоритм фильтра:

$$y(k) = [x(k-2) - x(k) - y(k-2)] / r^2.$$

При использовании символики  $z^{-1}$  полюс располагается внутри единичной окружности на том же радиусе со значением (при  $r < 1$ ):

$$r \approx 1 - (\Delta p / f_{\Delta})\pi. \quad r \approx 0.937.$$

Передаточная функция и алгоритм фильтра:

$$H(z) = (z-1)(z+1)/[(z-r \exp(j\pi/2))(z-r \exp(-j\pi/2))] = (z^2-1)/(z^2+r^2) = (1-z^{-2})/(1+r^2 z^{-2}).$$

$$y(k) = x(k) - x(k-2) - r^2 y(k-2).$$

Характеристики фильтров приведены на рис. 9.2.2. Индексы  $h1$ ,  $H1$ ,  $y1$  относятся к первому фильтру с полюсом за пределами единичной окружности, индексами  $h2$ ,  $H2$ ,  $y2$  – внутри окружности (символика  $z^{-1}$ ). Импульсные отклики фильтров получены подачей на их входы импульса Кронекера, частотные характеристики вычислены по импульсным откликам. Значение  $r$  первого фильтра подобрано по АЧХ под равный коэффициент усиления гармонике  $f_p$  со вторым фильтром, после чего коэффициенты фильтров нормированы по коэффициенту усиления к 1 на частоте  $f_p$ .

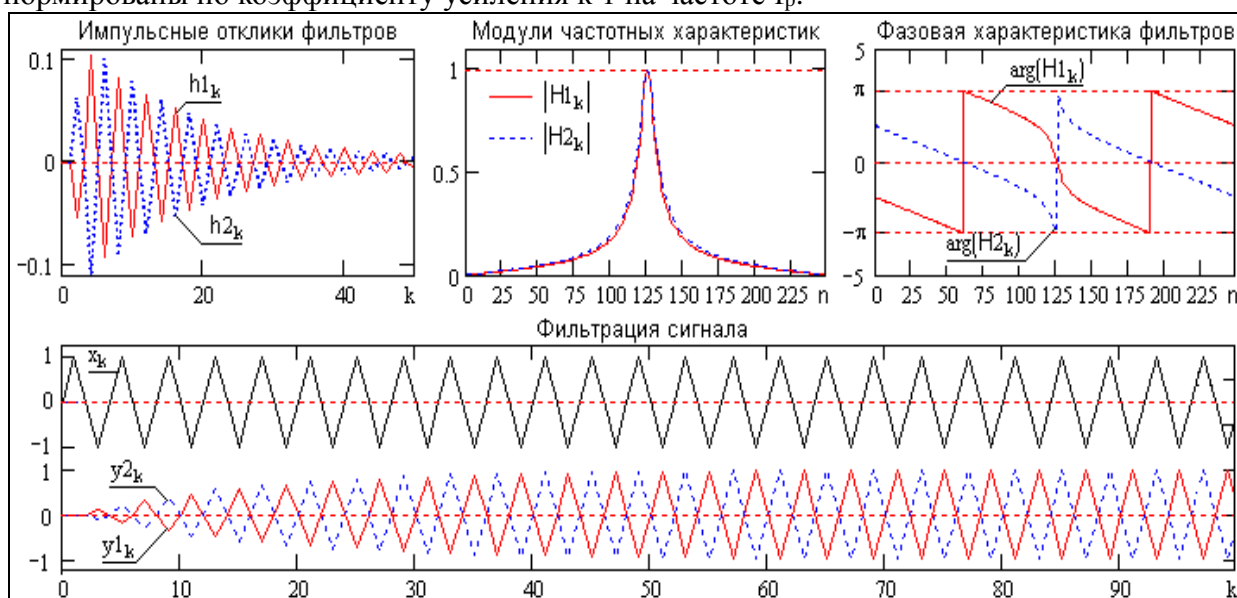


Рис. 9.2.2.

Как следует из рисунков, изменение многочлена по степеням  $z$  на  $1/z$  хотя и изменило коэффициенты разностного уравнения фильтра, но практически не повлияло на его амплитудно-частотную характеристику. Однако при этом произошло изменение области сходимости фильтра с соответствующим изменением фазовых углов направления на полюсы из всех точек единичной окружности, что отразилось на фазово-частотной характеристике и отсчетах импульсного отклика фильтра изменением фазы на  $\pi$ .

Синтез систем непосредственно в  $z$ -области применяется, в основном, только для режекторных и селекторных фильтров и более детально рассматривается ниже.

**Метод инвариантного преобразования** импульсной характеристики применяется для получения из подходящей аналоговой передаточной функции  $H(s)$  с помощью преобразования Лапласа импульсной характеристики  $h(t)$ , которая затем дискретизируется и подвергается  $z$ -преобразованию,

Допустим, имеется простая аналоговая система с передаточной функцией:

$$H(s) = C / (s-p).$$

Выполняем обратное преобразование Лапласа функции  $H(s)$  и дискретизируем результат преобразования с определенной постоянной времени  $\Delta t$ :

$$h(t) = \mathcal{TL}^{-1}[H(s)] = C \exp(pt) \rightarrow C \exp(pn\Delta t).$$

Выполняем z-преобразование и формируем передаточную функцию H(z):

$$H(z) = \sum_{n=0}^{\infty} h(n\Delta t) z^n = \sum_{n=0}^{\infty} C \exp(pn\Delta t) z^n = C / (1 - z \exp(p\Delta t)).$$

При преобразовании фильтров более высоких порядков функции H(s) раскладываются на простые дроби, для каждой из которых находится соответствующий блок H<sub>i</sub>(z), а система в целом реализуется в параллельной форме.

**Согласованное z-преобразование** применяется для преобразования аналоговых фильтров в эквивалентные цифровые непосредственным переводом всех полюсов и нулей с s-плоскости в z-плоскость:

$$(s - \alpha) \rightarrow z \exp(\alpha \Delta t).$$

Большинство полюсов и нулей являются комплексно сопряженными и реализуются фильтрами второго порядка:

$$(s - \alpha)(s - \alpha^*) \rightarrow 1 - 2z \exp(\operatorname{Re}(\alpha)\Delta t) \cos(\operatorname{Im}(\alpha)\Delta t) + z^2 \exp(\operatorname{Re}(\alpha)\Delta t).$$

Следует учитывать, что полоса частот аналоговых фильтров от нуля до бесконечности, а цифровых фильтров – от нуля до частоты Найквиста. При преобразовании происходит нелинейное сжатие бесконечной полосы частот в конечную с соответствующим искажением частотных характеристик фильтров.

**Билинейное z-преобразование** является основным методом получения коэффициентов рекурсивных БИХ-фильтров и использует следующую замену:

$$p = \gamma (1 - z) / (1 + z), \quad \gamma = 1 \text{ или } 2/\Delta t,$$

при этом ось jω p-плоскости отображается в единичную окружность z-плоскости, правая половина p-плоскости – внутрь единичной окружности, а левая половина с полюсами устойчивых аналоговых фильтров – снаружи единичной окружности. Аналогичная замена при отрицательной символике z<sup>-1</sup> с соответствующей сменой отображения:

$$p = \gamma (z - 1) / (z + 1).$$

Для фильтров верхних и нижних частот порядок фильтра H(z) равен порядку фильтра H(p). Для полосовых и заградительных фильтров порядок H(z) вдвое больше порядка H(p). Для сохранения частотных характеристик фильтра при нелинейном сжатии частотной шкалы аналогового фильтра (переход от ∞ к ω<sub>N</sub>) предварительно выполняется деформация частотной шкалы аналогового фильтра. Более подробно эти вопросы рассмотрены ниже.

### 9.3. РЕЖЕКТОРНЫЕ И СЕЛЕКТОРНЫЕ ФИЛЬТРЫ.

Режекторный фильтр (фильтр-пробка) подавляет определенную частоту во входном сигнале. Он может быть спроектирован непосредственно по z-диаграмме.

**Комплексная z-плоскость.** Простейший режекторный фильтр типа НЦФ имеет один нуль на единичной окружности в z-плоскости в точке с частотой, которую необходимо подавить. Так, например, если из входного сигнала требуется исключить постоянную составляющую (нулевая частота), то импульсная реакция фильтра НЦФ имеет вид:

$$H(z) = 1 - z. \tag{9.3.1}$$

Нуль функции (9.3.1) равен n<sub>1</sub>=1. Как можно видеть на рис. 9.3.1, коэффициент передачи сигнала H(ω) на любой частоте ω<sub>i</sub> от 0 до ω<sub>N</sub>=π/Δt - частоты Найквиста, определяемый выражением (9.3.1), будет равен длине вектора V<sub>n<sub>1</sub></sub>, проведенного из нуля функции H(z) - точка n<sub>1</sub> на действительной оси, до соответствующей частоты ω<sub>i</sub> - точки z(ω<sub>i</sub>) на единичной окружности. На частоте ω = 0 длина этого вектора равна нулю. Амплитудно-частотная характеристика фильтра, приведенная пунктиром на рисунке 9.3.2 для передаточной функции (9.3.1), далека от идеальной для фильтр-пробки.

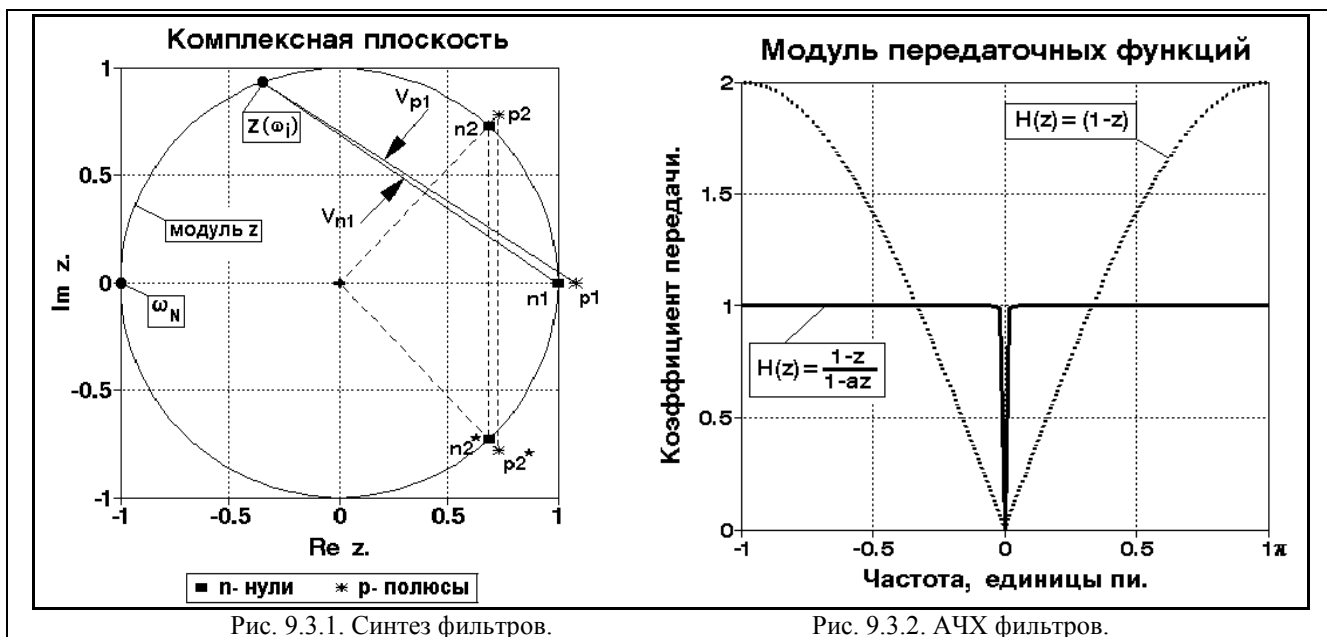


Рис. 9.3.1. Синтез фильтров.

Рис. 9.3.2. АЧХ фильтров.

**Режекторный фильтр постоянной составляющей сигнала.** Сконструируем простейший РЦФ, добавив к оператору (9.3.1) один полюс вне единичной окружности на малом расстоянии от нуля:

$$H_n(z) = G(1-z)/(1-az), \quad z_p = 1/a. \quad (9.3.2)$$

Допустим, что полюс помещен в точке  $p_1 = 1.01$ , при этом,  $a=0,99$ . Масштабный коэффициент  $G$  получим нормировкой  $H(z)$  к 1 на частоте Найквиста. Для приведенных условий  $G=0,995$ . Отсюда, при  $\Delta t=1$ :

$$H_n(z) = 0,995(1-z)/(1-0,99z),$$

$$y(k) = 0,995[x(k) - x(k-1)] + 0,99y(k-1).$$

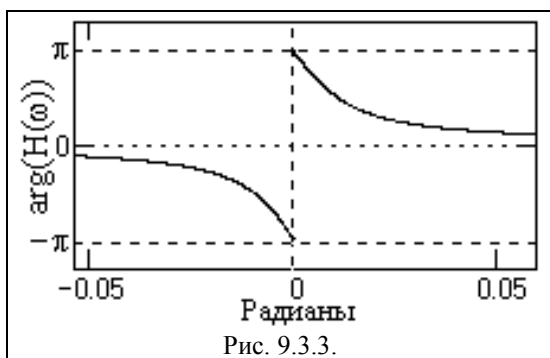


Рис. 9.3.3.

Отображение нуля  $n_1$  и полюса  $p_1$  на  $z$ -плоскости и АЧХ фильтра для исключения постоянной составляющей приведены на рис.9.3.1. Коэффициент передачи сигнала на произвольной частоте  $\omega_i$  равен отношению длин векторов  $V_{n1}(z)$  и  $V_{p1}(z)$ :

$$|H_n(z)| = G V_{n1}(z)/V_{p1}(z),$$

соответственно из нуля и полюса до точки  $z(\omega_i)$  на единичной окружности и близок к единице

для всех частот, за исключением нулевой.

Фазочастотная характеристика фильтра приведена на рис. 9.3.3 и определяется разностью фазовых углов векторов  $V_{n1}(z)$  и  $V_{p1}(z)$ :

$$\varphi_n(\omega) = \varphi_{n1} - \varphi_{p1}.$$

**Режекторный фильтр произвольной частоты.** При проектировании на подавление любой другой частоты  $\omega_v$  нули и полюсы располагаются на соответствующем радиусе  $z$ -плоскости. Радиальный угол направления на нуль и полюс определяются выражением:

$$\varphi_v = \pm \pi \cdot \omega_v / \omega_N. \quad (9.3.3)$$

Наличие двух знаков в выражении (9.3.3) отражает тот факт, что для получения вещественной функции фильтра нули и полюсы должны быть комплексно-сопряженными парами (их произведение дает вещественную функцию), т.е.:

$$H_v(z) = G(z-z_n)(z-z_n^*)/[(z-z_p)(z-z_p^*)]. \quad (9.3.4)$$

Нули фильтра располагаются на единичной окружности:

$$z_n = \cos \varphi_v + j \sin \varphi_v = \operatorname{Re} z_n + j \operatorname{Im} z_n. \quad (9.3.5)$$

Полюсы - на полярном радиусе R:

$$z_p = R \cos \varphi_v + j R \sin \varphi_v = \operatorname{Re} z_p + j \operatorname{Im} z_p. \quad (9.3.6)$$

Пример положения нулей ( $n_2$  и  $n_2^*$ ) и полюсов ( $p_2$  и  $p_2^*$ ) приведен на рис.9.3.1. Подставляя (9.3.5-9.3.6) в (9.3.4), получаем:

$$H_v(z) = \frac{G(z^2 - 2z \operatorname{Re} z_n + 1)}{1 + (z^2 - 2z \operatorname{Re} z_p)/R^2}, \quad (9.3.7)$$

$$G = [1 + (1 + 2\operatorname{Re} z_p)/R^2] / (2 + 2\operatorname{Re} z_n). \quad (9.3.8)$$

При приведении уравнения (9.3.7) в типовую форму:

$$H_v(z) = \frac{G(b_0 + b_1 z + b_2 z^2)}{1 + a_1 z + a_2 z^2}, \quad (9.3.7')$$

$$b_0 = 1, \quad b_1 = -2 \cdot \operatorname{Re} z_n, \quad b_2 = 1. \quad (9.3.9)$$

$$a_1 = - (2 \cdot \operatorname{Re} z_p)/R^2, \quad a_2 = 1/R^2.$$

Соответственно, алгоритм вычислений:

$$y(k) = G \cdot [x(k) + b_1 \cdot x(k-1) + x(k-2)] - a_1 \cdot y(k-1) - a_2 \cdot y(k-2). \quad (9.3.10)$$

**Пример.** Проведем расчет режекторного фильтра на частоту питания приборов  $f_s = 50$  Гц, которая очень часто попадает в измеренные данные. Шаг дискретизации данных  $\Delta t = 0.001$  сек.

Частота Найквиста:  $f_N = 1/2\Delta t = 500$  Гц.

Радиальный угол на нули и полюса фильтра в  $z$ -плоскости:  $\varphi = \pi \cdot f_s / f_N = 0.1\pi$ .

Радиус полюса фильтра примем равным  $R = 1.01$ . Значения нуля и полюса:

$$z_n = \cos \varphi + j \sin \varphi = 0.951 + 0.309 j, \quad z_p = R \cdot \cos \varphi_v + j R \cdot \sin \varphi_v = 0.961 + 0.312 j.$$

Значение масштабного множителя  $G$  по (9.3.8):  $G = 0.99$ .

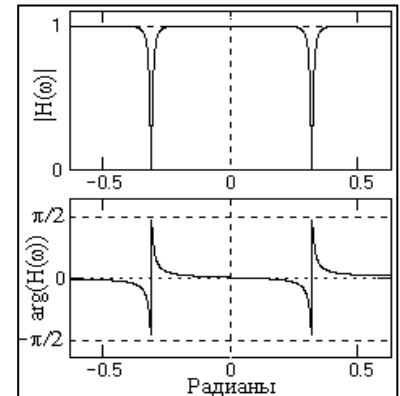
Значения коэффициентов передаточной функции:  $b_1 = -2 \cdot \operatorname{Re} z_n = -1.902$ ,

$$a_1 = - (2 \cdot \operatorname{Re} z_p)/R^2 = -1.883, \quad a_2 = 1/R^2 = 0.98.$$

Частотная передаточная функция фильтра при подстановке коэффициентов в уравнение (9.3.7') и замене  $z = \exp(-j\omega)$ :

$$H(\omega) = 0.99[1 - 1.902 \cdot \exp(-j\omega) + \exp(-2j\omega)] / [1 - 1.883 \cdot \exp(-j\omega) + 0.98 \cdot \exp(-2j\omega)].$$

Алгоритм фильтра:  $y(k) = 0.99[x(k) - 1.902 x(k-1) + x(k-2)] + 1.883 y(k-1) - 0.98 y(k-2)$ .



Для проверки вычисленного в примере фильтра на рис. 9.3.4 приведен модельный входной сигнал, состоящий из суммы двух равных по амплитуде гармоник с частотой 50 и 53 Гц, и сигнал на выходе фильтра (смещен вверх). Справа на рисунке приведены спектры входного и выходного сигналов. Спектр выходного сигнала зарегистрирован после интервала установления реакции фильтра, который хорошо заметен на начальной части графика выходного сигнала. После установления сигнал на выходе фильтра практически полностью освобожден от гармоники 50 Гц.

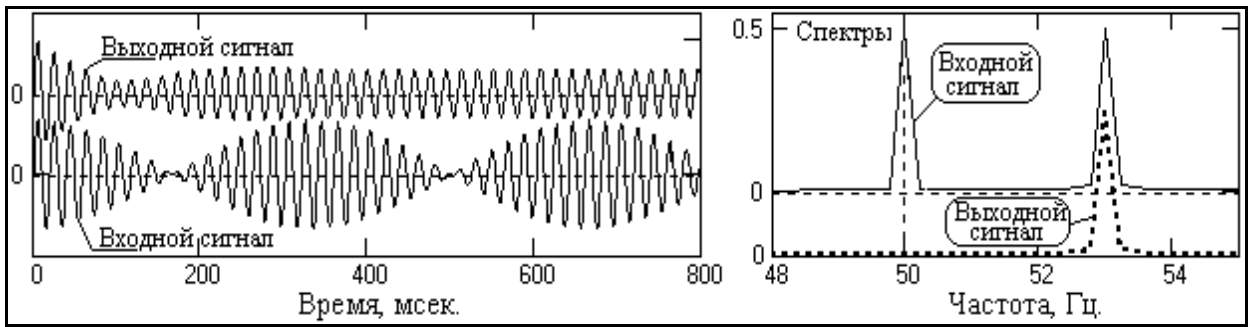


Рис. 9.3.4.

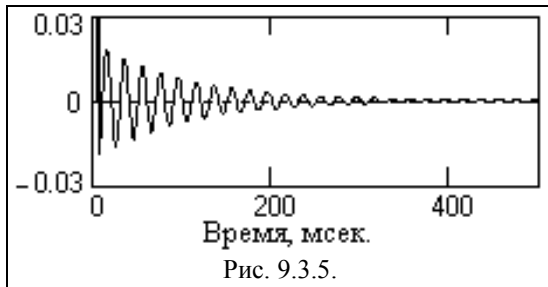


Рис. 9.3.5.

При  $R \rightarrow 1$  ширина полосы подавления фильтра становится все более узкой, но при этом увеличивается длительность импульсной реакции фильтра и, соответственно, время установления фильтра при изменении спектра входного сигнала. В первом приближении значимая часть импульсной реакции режекторных фильтров равна  $(4 \div 5)/(R-1)$ . Пример импульсной реакции для фильтра,

вычисленного выше, приведен на рис. 9.3.5. Отклик фильтра получен при подаче на вход РЦФ импульса Кронекера. На графике не показан начальный пик отклика (отсчет на нулевой точке), амплитуда которого равна значению  $G$ .

**Селекторный фильтр.** Если в уравнении (9.3.4) опустить нули, то получим селекторный фильтр, выделяющий сигналы одной частоты  $\omega_s$  – частоты селекции, с передаточной функцией:

$$H_s(z) = G / [(z - z_p)(z - z_p^*)] = G1 / (1 + a_1 z + a_2 z^2). \quad (9.3.11)$$

Характер передаточной функции (9.3.11) можно представить непосредственно по  $z$ -плоскости (рис. 9.3.1). При расположении полюсов фильтра за пределами единичного круга (например, в точках  $p_2$  и  $p_2^*$ ) значение коэффициента передачи фильтра на произвольной частоте  $\omega$  на единичной окружности будет обратно пропорционально величине векторов из этих точек окружности на полюса фильтра. При изменении  $\omega$  от нуля до  $\pm\pi$  (движение по единичной окружности на  $z$ -плоскости по или против часовой стрелки) один из векторов (на полюс противоположной полуплоскости) изменяется в

достаточно небольших пределах (не превышая значения 2), в то время как второй из векторов (на полюс в своей полуплоскости) будут сначала уменьшаться, достигает минимума при расположении  $\omega$  на полярном радиусе полюса (на частоте селекции  $\omega_s$ ), а затем снова начинает увеличиваться.

Соответственно, значение  $H_s(\omega)$  максимально на частоте селекции  $\pm\omega_s$  и при  $R \rightarrow 1$  может быть очень высоким. Пример передаточной функции

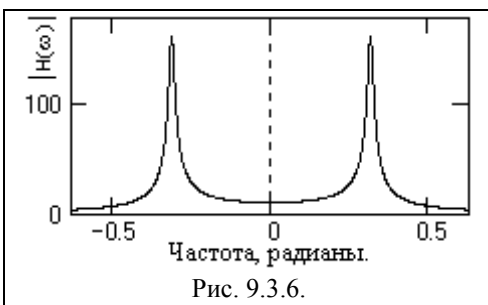


Рис. 9.3.6.

приведен на рис. 9.3.6.

При необходимости фильтр может быть пронормирован к 1 на частоте селекции определением значения  $G1$  по условию  $H_s(\omega) = 1$  при  $\omega = \omega_s$ , т.е.:

$$G1 = 1 + a_1 z(\omega_s) + a_2 z(\omega_s)^2.$$

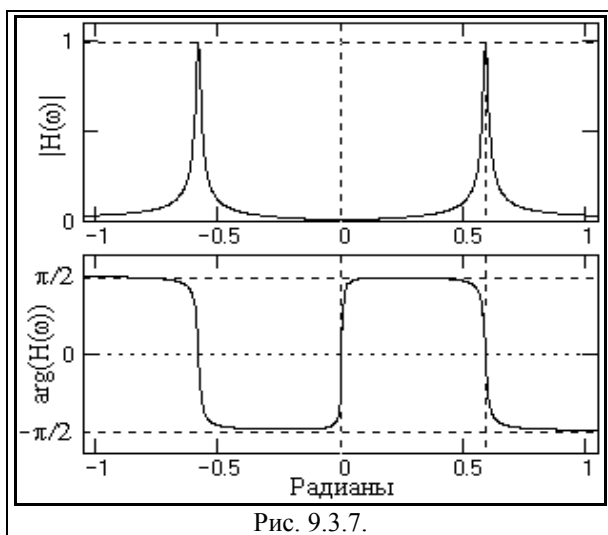


Рис. 9.3.7.

Фильтр (9.3.11) в принципе не может иметь нулевого коэффициента передачи на других частотах главного диапазона. Если последнее является обязательным, то фильтр выполняется методом обращения режекторного фильтра  $H_v(z)$ :

$$H_s(z) = 1 - H_v(z).$$

$$H_s(z) = \frac{c_0 + c_1 z + c_2 z^2}{1 + a_1 z + a_2 z^2}. \quad (9.3.12)$$

$$c_0 = 1 - G, \quad c_1 = a_1 - G b_1, \quad c_2 = a_2 - G.$$

Пример передаточной функции фильтра приведен на рис. 9.3.7. Пример применения фильтра для выделения гармонического сигнала на уровне шумов, мощность которых больше мощности

сигнала, приведен на рис. 9.3.8.

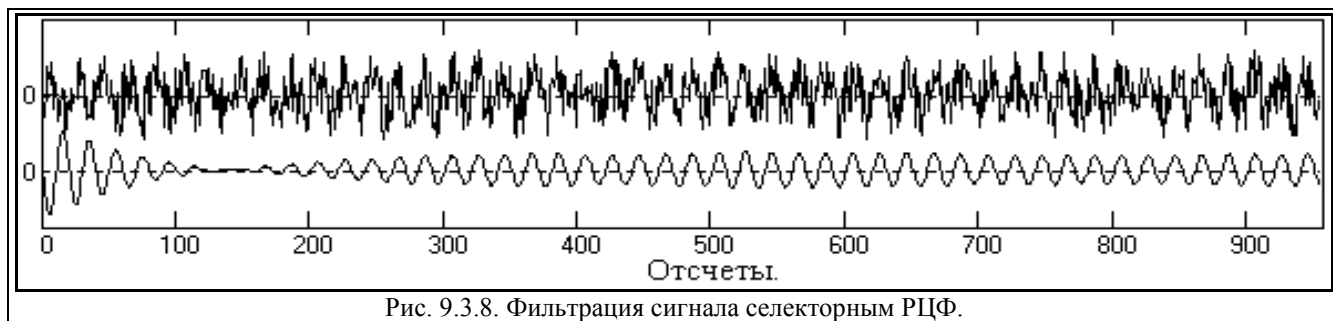


Рис. 9.3.8. Фильтрация сигнала селекторным РЦФ.

#### 9.4. БИЛИНЕЙНОЕ Z-ПРЕОБРАЗОВАНИЕ.

**Принцип преобразования.** При стандартном z-преобразовании передаточной функции используется замена переменной вида:

$$z = \exp(-p\Delta t), \quad (9.4.1)$$

где  $\Delta t$  - шаг дискретизации данных,  $p$  - комплексная переменная,  $p = \sigma + j\omega$ .

Уравнение (9.4.1) можно записать в виде  $\ln z = -p\Delta t$  и разложить  $\ln z$  в ряд:

$$\ln z = -2[(1-z)/(1+z) + (1-z)^3/(3(1-z)^3) + \dots], \quad z > 0.$$

Первый член этого разложения и представляет собой билинейное z-преобразование:

$$p = (2/\Delta t)(1-z)/(1+z). \quad (9.4.2)$$

По сути, оно представляет собой отображение точек комплексной p-плоскости в точки комплексной z-плоскости, и наоборот. В общем виде:

$$p = \gamma(1-z)/(1+z), \quad (9.4.3)$$

$$z = (\gamma-p)/(\gamma+p). \quad (9.4.4)$$

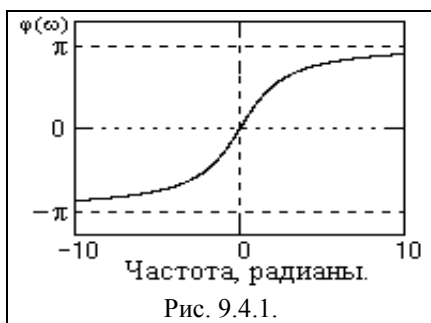
Значение множителя  $\gamma$  не меняет формы преобразования, в связи с чем обычно принимают  $\gamma = 1$ . Подставим  $p = j\omega$  в (9.4.4) и выразим  $z$  в показательной форме:

$$z = r \exp(j\varphi(\omega)), \quad r = |z| = 1.$$

$$\varphi(\omega) = 2 \operatorname{arctg}(\omega/\gamma),$$

$$\omega = 0, \quad z = \exp(j0) = 1,$$

$$\omega = \pm \infty, \quad z = \exp(\pm j\pi) = -1$$



При изменении  $\omega$  от  $-\infty$  до  $\infty$  фазовый угол  $\varphi(\omega)$  монотонно изменяется от  $-\pi$  до  $\pi$  (см. рис. 9.4.1), т.е. мнимая ось  $p$ -плоскости ( $p = j\omega$ ,  $-\infty < \omega < \infty$ ) отображается в единичную окружность  $z$ -плоскости, правая половина  $s$ -плоскости – внутрь единичной окружности, а левая половина с полюсами устойчивых аналоговых фильтров – снаружи единичной окружности.

#### Деформация частотной шкалы.

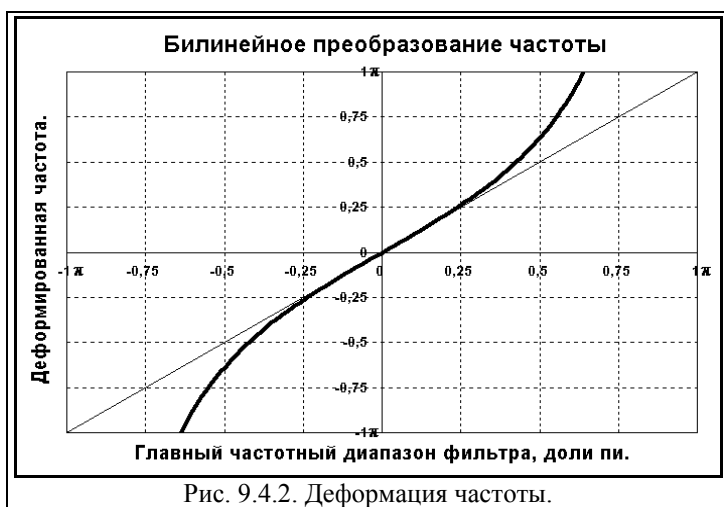
Реальное отображение передаточных функций фильтров является непрерывным (в силу своей физической сущности) и для упрощения дальнейших расчетов обычно задается в аналитической форме в комплексной  $p$ -плоскости по частотному аргументу  $\omega$  от  $-\infty$  до  $+\infty$ . При билинейном  $z$ -преобразовании происходит нелинейное искажение шкалы частот: полный частотный диапазон от  $-\infty$  до  $\infty$  непрерывных функций в  $p$ -плоскости сжимается до главного частотного диапазона от  $-\pi/\Delta t$  до  $\pi/\Delta t$  дискретных функций в  $z$ -плоскости. При задании уравнений непрерывных передаточных функций в частотной области это должно сопровождаться соответствующей обратной деформацией частотной шкалы, которая будет скомпенсирована при билинейном  $z$ -преобразовании. Подставляя в (9.4.2)  $z = \exp(-j\omega\Delta t)$  и умножая числитель и знаменатель правой части полученного уравнения на  $\exp(j\omega\Delta t/2)$ , получим:

$$p = (2/\Delta t)[\exp(j\omega\Delta t/2) - \exp(-j\omega\Delta t/2)] / [\exp(j\omega\Delta t/2) + \exp(-j\omega\Delta t/2)],$$

$$p = (2/\Delta t) \operatorname{th}(j\omega\Delta t/2). \quad (9.4.5)$$

Обозначим новую шкалу частот в  $p$ -области через индекс  $\omega_d$  (деформированная) и, полагая  $p = j\omega_d$ , с учетом тождества  $\operatorname{th}(x) = -j \operatorname{tg}(jx)$ , получаем:

$$\omega_d = (2/\Delta t) \operatorname{tg}(\omega\Delta t/2) = \gamma \operatorname{tg}(\omega\Delta t/2), \quad -\pi/\Delta t < \omega < \pi/\Delta t. \quad (9.4.6)$$



Выражение (9.4.6) позволяет осуществлять переход от фактических частот  $\omega$  главного частотного диапазона, которым должен соответствовать оператор РЦФ, к деформированным частотам  $\omega_d$  комплексной  $p$ -плоскости, на которой можно задавать требуемую форму передаточной функции проектируемого фильтра, при этом аппроксимация передаточных функций, учитывая область существования  $\omega$  от  $-\infty$  до  $\infty$ , может производиться многочленами и

рациональными функциями. Связь частот приведена на рис. 9.4.2 (в начальной части  $\pm \pi$  пространства деформированных частот).

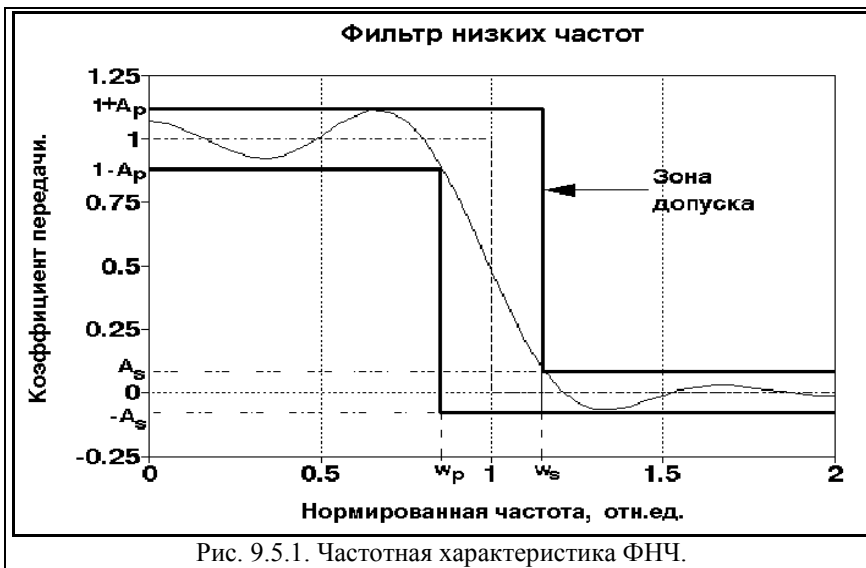
### 9.5. ТИПЫ РЕКУРСИВНЫХ ЧАСТОТНЫХ ФИЛЬТРОВ.

Рекурсивные цифровые фильтры, как и нерекурсивные, не могут обеспечить реализацию идеальной частотной характеристики со скачкообразными переходами от полосы пропускания к полосе подавления. Поэтому на этапе решения аппроксимационной задачи необходимо определить передаточную функцию  $H(\omega)$  фильтра, которая обеспечивает воспроизведение необходимой амплитудно-частотной характеристики (АЧХ) с требуемой точностью. Требования к фазочастотной характеристике (ФЧХ) частотных фильтров, как правило, не задаются, т. к. это приводит к резкому усложнению решения задачи. Специальные требования к форме ФЧХ обычно реализуются после расчета фильтров с заданной АЧХ путем контроля полученной ФЧХ и разработкой, при

необходимости, дополнительных корректоров ФЧХ.

Синтез рекурсивных фильтров, как и НЦФ, выполняется на базе фильтров низких частот (ФНЧ). Другие типы фильтров (ФВЧ - высоких частот и ПФ - полосовые) образуются на основе ФНЧ путем частотного преобразования.

**Аппроксимационная задача** низкочастотного фильтра. В качестве основных исходных данных для решения аппроксимационных задач принимаются граничные частоты  $\omega_p$  - полосы пропускания, и  $\omega_s$  - начала полосы подавления сигнала. Как правило, задаются также допуски  $A_p$  - на максимальное значение неравномерности в полосе пропускания, и  $A_s$  - на максимальное отклонение АЧХ от нуля в полосе подавления (уровень шума фильтра). Разность между граничными частотами  $\omega_p$  и  $\omega_s$  будет определять ширину переходной зоны. Типичный пример задания формы АЧХ приведен на рис. 9.5.1. В допустимой зоне передаточной функции условно показана возможная форма АЧХ, удовлетворяющая заданным условиям.



Кроме основных частотных параметров могут задаваться и требования к форме АЧХ (монотонность в полосе пропускания или подавления, характер пульсаций и т.п.), которые определяют выбор функции аппроксимации.

**Передаточная функция.** При решении аппроксимационной задачи амплитудно-частотная характеристика

фильтра обычно задается в действительной аналитической форме - виде квадрата передаточной функции, нормированной по амплитуде и граничной частоте передачи:

$$|H(W)|^2 = H(W) \cdot H^*(W) = 1/(1+A_n(W)), \quad (9.5.1)$$

где  $A_n(W)$  - многочлен  $n$ -го порядка,  $W$  - нормированная частота (например,  $W = \omega/\omega_p$ ). Вид многочлена  $A_n(W)$  выбирается таким образом, чтобы выполнялось условие:  $A_n(W) \ll 1$  при  $0 < W < 1$ , что обеспечивает  $|H(W)|^2 \rightarrow 1$ , и  $A_n(W) \gg 1$  при  $W > 1$ , соответственно  $|H(W)|^2 \rightarrow 0$ . Крутизна переходных зон фильтра устанавливается величиной порядка фильтра (чем больше значение  $n$ , тем больше крутизна переходных зон).

По знаменателю правой части выражения (9.5.1) достаточно просто могут быть определены комплексные полюса передаточной функции в  $p$ -области преобразования Лапласа и соответствующим комбинированием и объединением комплексно-сопряженных полюсов получены передаточные функции в виде биквадратных блоков при четном порядке, и с одним линейным блоком при нечетном порядке:

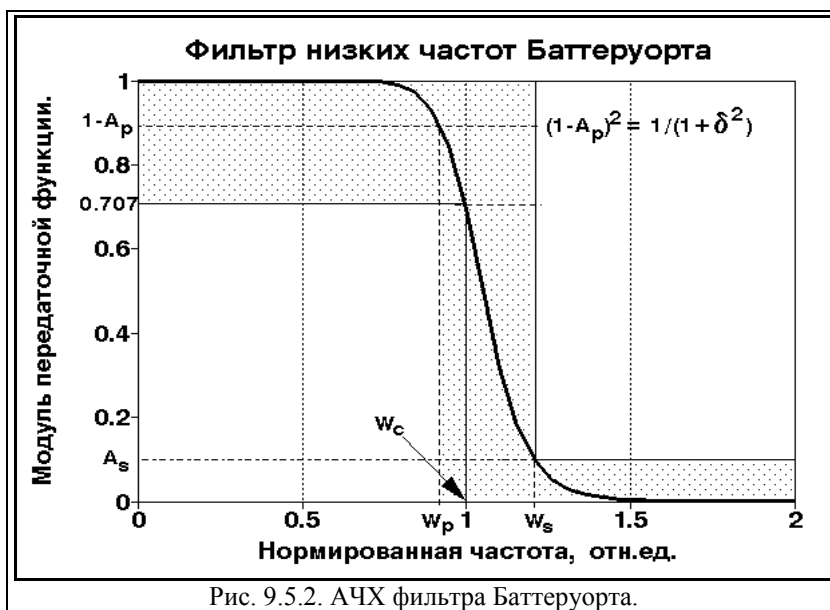
$$H(p) = G \prod_{n=1}^{N/2} B_n(p), \quad n\text{-четное}, \quad (9.5.2)$$

$$H(p) = \frac{G}{p-p_0} \prod_{n=1}^{(N-1)/2} B_n(p), \quad n\text{-нечетное}, \quad (9.5.3)$$

где  $B_n(p)$  выражается в форме:

$$B_n(p) = 1/[(p-p_n)(p-p_n^*)] = 1/(p^2 - 2 a_n p + b_n). \quad (9.5.4)$$





**Виды фильтров.** В настоящее время существует достаточно большое количество видов рекурсивных частотных фильтров и их модификаций. Наиболее известный из них - фильтр Баттеруорта (рис.9.5.2). Он имеет монотонную гладкую АЧХ во всем частотном диапазоне. При том же порядке многочленов фильтров (равном количестве полюсов) большую крутизну обеспечивают фильтры

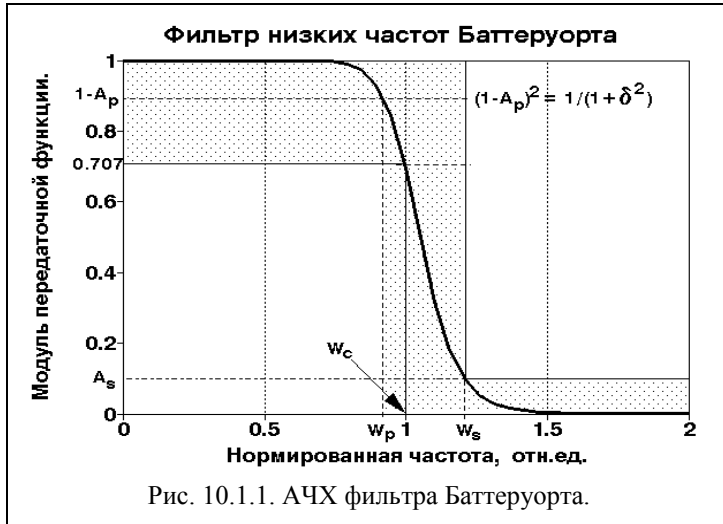
Чебышева – прямой и инверсный, однако при этом в полосе пропускания (для инверсного – в полосе подавления) у фильтров Чебышева появляются равноволновые пульсации (с одинаковой амплитудой пульсаций). Еще более крутые срезы характеристик (при равноволновых пульсациях как в полосах пропускания, так и в полосе подавления) реализуются с использованием эллиптических функций.

## Тема 10. РЕКУРСИВНЫЕ ЧАСТОТНЫЕ ЦИФРОВЫЕ ФИЛЬТРЫ

Процесс проектирования рекурсивного частотного фильтра обычно заключается в задании необходимой передаточной характеристики фильтра в частотной области и ее аппроксимации с определенной точностью какой-либо непрерывной передаточной функцией, с последующим Z-преобразованием для перехода в Z-область. Первые две операции хорошо отработаны в теории аналоговой фильтрации сигналов, что позволяет использовать для проектирования цифровых фильтров большой справочный материал по аналоговым фильтрам. Последняя операция является специфичной для цифровых фильтров.

Для алгебраического преобразования непрерывной передаточной функции в многочлен по Z используется билинейное преобразование, известное в теории комплексных переменных под названием дробно-линейного преобразования.

## 10.1. НИЗКОЧАСТОТНЫЙ ФИЛЬТР БАТТЕРУОРТА



### Передаточная функция.

Гладкий вид амплитудно-частотной характеристики фильтра Баттеруорта (рис. 10.1.1) задают квадратом передаточной функции вида:

$$|H(W)|^2 = H(W)H^*(W) = 1/(1+W^{2N}),$$

где  $W = \omega/\omega_c$  - нормированная частота,  $\omega_c$  - частота среза АЧХ фильтра, на которой  $|H(\omega)|^2 = 1/2$  (соответственно  $H(\omega) = 0.707$ , или 3 дБ),  $N$  - порядок фильтра, определяющий крутизну среза АЧХ. Функция  $|H(W)|^2$  - представляет собой энергетический спектр

сигнала (спектральную плотность мощности) и не имеет фазовой характеристики, т.е. является четной вещественной, образованной произведением двух комплексно сопряженных функций  $H(W)$  и  $H^*(W)$ . При  $W \rightarrow 0$  коэффициент передачи фильтра стремится к 1. Учитывая, что результаты вычислений будут относиться к цифровым фильтрам и при  $z$ -преобразовании с переходом в главный частотный диапазон произойдет искажение частот, до начала расчетов фактические значения задаваемых частотных характеристик (значения  $\omega_c$ ,  $\omega_p$  и  $\omega_s$ ) следует перевести в значения деформированных частот по выражению:

$$\omega_d = (2/\Delta t) \operatorname{tg}(\omega \Delta t / 2), \quad -\pi/\Delta t < \omega < \pi/\Delta t. \quad (10.1.1)$$

**Крутизна среза.** Наклон частотной характеристики фильтра при переходе от области пропускания к области подавления можно характеризовать коэффициентом крутизны среза фильтра  $K$  в децибелах на октаву:

$$K = 20 \log|H(\omega_2)/H(\omega_1)|, \quad (10.1.2)$$

где  $\omega_1$  и  $\omega_2$  - частоты с интервалом в одну октаву, т.е.  $\omega_2 = 2\omega_1$ .

Длительность импульсной реакции фильтра в пределах ее значимой части также зависит от крутизны среза: чем больше крутизна, тем больше длительность импульсного отклика фильтра.

**Порядок фильтра.** Принимая  $\omega_1 = W_c$ ,  $\omega_2 = W_s$  и подставляя в (10.1.2) значения  $H(W)$  с приведенными данными, получим приближенное выражение для определения порядка фильтра по заданному значению  $K$ :

$$N = K/6. \quad (10.1.6')$$

Так, для гарантированного ослабления сигнала в полосе подавления в 100 раз (40 децибел) порядок фильтра  $N = 7$ . В среднем, при изменении  $N$  на единицу коэффициент подавления сигнала изменяется на 6 децибел.

Исходные требования к передаточной функции фильтра обычно задаются в виде значений  $\omega_p$ ,  $\omega_s$  и коэффициентов неравномерности (пульсаций)  $A_p$  и  $A_s$  (см. рис. 10.1.1). Для определения частоты среза  $\omega_c$  по уровню 0.707 и порядка фильтра введем параметр  $\delta$ , связанный с коэффициентом  $A_p$  следующим соотношением:

$$(1 - A_p)^2 = 1/(1 + \delta^2),$$

$$\delta = [1/(1 - A_p)] \sqrt{1 - (1 - A_p)^2} = A_p \sqrt{2/A_p - 1/(1 - A_p)}. \quad (10.1.3)$$

Для учета деформации частотной шкалы в процессе билинейного преобразования при переходе в дальнейшем к полиномам по  $Z$ , выполняем расчет деформированных частот  $\omega_{dp}$  и  $\omega_{ds}$  по формулам:

$$\omega_{dp} = 2 \operatorname{tg}(\omega_p \Delta t / 2) / \Delta t, \quad (10.1.4)$$

$$\omega_{ds} = 2 \operatorname{tg}(\omega_s \Delta t / 2) / \Delta t.$$

При нормированной частоте  $W = \omega / \omega_{dc}$ , где  $\omega_{dc}$  соответственно также деформированная частота, на границах переходной зоны выполняются равенства:

$$1 / (1 + \delta^2) = 1 / [1 + (\omega_{dp} / \omega_{dc})^{2N}], \quad (10.1.5)$$

$$A_s^2 = 1 / [1 + (\omega_{ds} / \omega_{dc})^{2N}].$$

Отсюда:

$$\delta^2 = (\omega_{dp} / \omega_{dc})^{2N}, \quad 1 / A_s^2 - 1 = (\omega_{ds} / \omega_{dc})^{2N}.$$

Решая эти два уравнения совместно, находим:

$$N = \ln [\delta / \sqrt{1 / A_s^2 - 1}] / \ln(\omega_{dp} / \omega_{ds}), \quad (10.1.6)$$

$$\omega_{dc} = \omega_{dp} / \delta^{1/N}. \quad (10.1.7)$$

### Пример расчета фильтра низких частот Баттеруорта.

Начиная с этого параграфа, будем сопровождать рассмотрение теории последовательным расчетом фильтра низких частот с применением приводимых формул. Для расчета примем следующие исходные параметры фильтра:

- Шаг дискретизации данных  $\Delta t = 0.0005$  сек. Частота Найквиста  $f_N = 1 / 2\Delta t = 1000$  Гц,  $\omega_N = 6.283 \cdot 10^3$  рад.
- Граничная частота пропускания:  $f_p = 300$  Гц,  $\omega_p = 1.885 \cdot 10^3$  рад.
- Граничная частота подавления:  $f_s = 500$  Гц,  $\omega_s = 3.142 \cdot 10^3$  рад.
- Коэффициенты неравномерности:  $A_p = A_s = 0.1$ .

Расчет дополнительных параметров:

1. Значение  $\delta$  по формуле (10.1.3):  $\delta = 0.484$ .
2. Деформированные частоты по формуле (10.1.4):  $\omega_{dp} = 2.038 \cdot 10^3$  рад,  $\omega_{ds} = 4 \cdot 10^3$  рад.
3. Порядок фильтра по формуле (10.1.6):  $N = 4.483$ .

Для пояснения порядка расчетов при четном и нечетном порядке фильтра, принимаем  $N1=4, N2=5$ .

4. Частота среза по формуле (10.1.7):  $\omega_{dc}(N1) = 2.443 \cdot 10^3$  рад (389 Гц),  $\omega_{dc}(N2) = 2.356 \cdot 10^3$  рад (375 Гц).
5. По формуле  $H(w) = [1 / (1 + w^{2N})]^{1/2}$ ,  $w = \omega / \omega_{dc}$ , строим графики передаточных функций (рис.10.1.2).

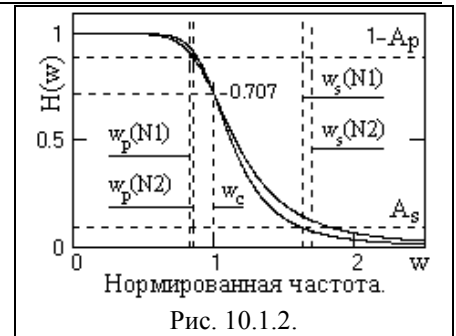


Рис. 10.1.2.

**Преобразование Лапласа.** Переводим функцию  $|H(W)|^2$  на координатную ось пространства преобразования Лапласа при  $p = jW$ , для чего достаточно подставить  $W = p/j$ :

$$|H(p)|^2 = 1 / [1 + (p/j)^{2N}]. \quad (10.1.8)$$

Полюсы функции находятся в точках нулевых значений знаменателя:

$$1 + (p/j)^{2N} = 0, \quad p = j \sqrt[2N]{-1}. \quad (10.1.9)$$

Отсюда следует, что полюсы располагаются на единичной окружности в  $p$ -плоскости, а их местоположение определяется корнями уравнения (10.1.9). В полярных координатах:

$$p_n = j \exp(j\pi(2n-1)/2N), \quad n = 1, 2, \dots, 2N. \quad (10.1.10)$$

$$p_n = j \cos[\pi(2n-1)/2N] - \sin[\pi(2k-1)/2N]. \quad (10.1.10')$$

**Продолжение примера.**

6. Вычисляем значения полюсов фильтра по формуле (10.1.10). Значения полюсов и их расположение на  $p$ -плоскости приведены на рис. 10.1.2. Положение первого полюса отмечено. Нумерация полюсов идет против часовой стрелки.

Как следует из формулы (10.1.10) и наглядно видно на рис. 10.1.2, все полюса с  $n \geq N$  являются комплексно сопряженными с полюсами  $n < N$ . Устойчивую минимально-фазовую передаточную функцию фильтра образуют полюса левой половины  $p$ -плоскости:

$$H(p) = G/V(p), \quad (10.1.11)$$

где  $G$  - масштабный множитель,  $V(p)$  - полином Баттеруорта:

$$V(p) = V_1(p) V_2(p) \dots V_N(p), \quad (10.1.12)$$

$$V_n(p) = p - p_n. \quad (10.1.13)$$

Практическая реализация фильтра

Баттеруорта при четном значении  $N$  производится в виде последовательной каскадной схемы биквадратными блоками, т.е. составными фильтрами второго порядка. Для этого множители  $V(p)$  в (10.1.12) объединяются попарно с обоих концов ряда по  $n$  (от 1 до  $N$ ) по комплексно сопряженным полюсам, при этом для каждой пары получаем вещественные квадратичные множители:

$$\begin{aligned} V_m(p) &= V_n(p) \cdot V_{N+1-n}(p) = \\ &= [p - j \exp(j\pi(2n-1)/2N)][p + j \exp(j\pi(2(N+1)-2n-1)/2N)] = \\ &= [p + j \exp(j\pi(2n-1)/2N)][p - j \exp(j\pi(2n-1)/2N)] = \\ &= p^2 + 2p \sin(\pi(2m-1)/2N) + 1, \quad n = 1, 2, \dots, N/2; \quad m = n. \end{aligned} \quad (10.1.14)$$

Общее количество секций фильтра  $M=N/2$ . При нечетном  $N$  к членам (10.1.14) добавляется один линейный множитель с вещественным полюсом  $p_{(N+1)/2} = -1$ , пример положения которого на  $p$ -плоскости можно видеть на рисунке 10.1.2 для  $N=5$ :

$$V_{(N+1)/2}(p) = p + 1. \quad (10.1.15)$$

Машинное время фильтрации на один оператор фильтра первого или второго порядка практически не отличаются, поэтому использование операторов первого порядка можно не рекомендовать и при установлении порядка фильтра по формуле (10.1.6) округлять расчетное значение  $N$  в сторону большего четного числа, что создает определенный запас по крутизне среза частотной характеристики.

Таким образом, передаточная функция ФНЧ Баттеруорта в  $p$ -области при четном  $N$ :

$$H(p) = G \prod_{m=1}^M 1/V_m(p) = G \prod_{m=1}^M 1/(p^2 + a_m p + 1), \quad (10.1.16)$$

$$a_m = 2 \sin(\pi(2m-1)/2N), \quad m = 1, 2, \dots, N/2. \quad (10.1.17)$$

При нечетном  $N$ :

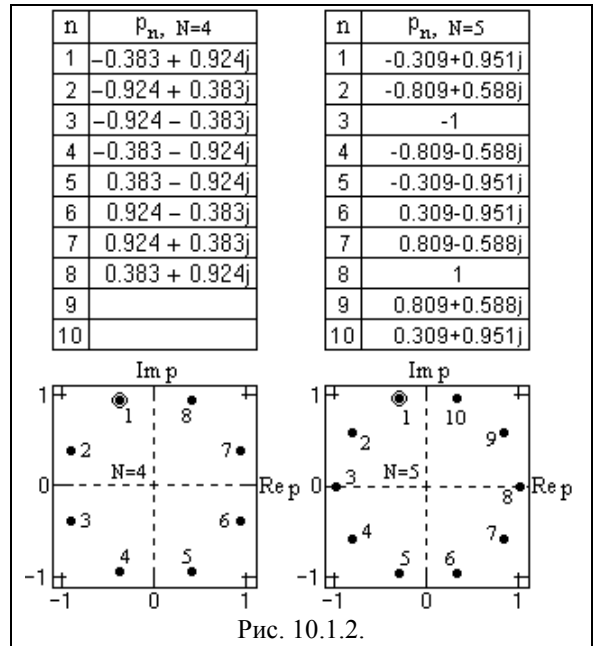
$$H(p) = (G/p + 1) \prod_{m=1}^{(N-1)/2} 1/(p^2 + a_m p + 1), \quad (10.1.16')$$

**Продолжение примера.**

7. Вычисляем значения коэффициентов  $a_m$  по формуле (10.1.17):

-  $N=4$ :  $a_1 = 0.765, \quad a_2 = 1.848.$

-  $N=5$ :  $a_1 = 0.618, \quad a_2 = 1.618.$



**Билинейное преобразование.** Для перевода передаточной функции фильтра в  $z$ -область производится билинейное преобразование, для чего в выражение (10.1.16)

подставляется параметр  $p$ :

$$p = \gamma(1-z)/(1+z). \quad (10.1.18)$$

С учетом автоматического возврата к нормальной шкале частот в главном частотном диапазоне  $z$ -преобразования значение коэффициента  $\gamma$ :

$$\gamma = 2/(\Delta t \cdot \omega_{dc}). \quad (10.1.19)$$

После перехода в  $z$ -область и приведения уравнения передаточной функции в типовую форму, для четного  $N$  получаем передаточную функцию из  $M=N/2$  биквадратных блоков:

$$H(z) = G \prod_{m=1}^M G_m (1+z)^2 / (1-b_m z + c_m z^2). \quad (10.1.20)$$

$$G_m = 1/(\gamma^2 + a_m \gamma + 1). \quad (10.1.21)$$

$$b_m = 2 \cdot G_m (\gamma^2 - 1). \quad (10.1.22)$$

$$c_m = G_m (\gamma^2 - a_m \gamma + 1). \quad (10.1.23)$$

При нечетном  $N$  добавляется один линейный блок первого порядка, который можно считать нулевым блоком фильтра ( $m=0$ ):

$$H(z) = G \frac{(1+z)/(\gamma+1)}{1-z(\gamma-1)/(\gamma+1)} \prod_{m=1}^{(N-1)/2} G_m (1+z)^2 / (1-b_m z + c_m z^2), \quad (10.1.24)$$

при этом, естественно, в выражении (10.1.24) используются значения коэффициентов  $G_m$ ,  $b_m$  и  $c_m$ , вычисленные по (10.1.21-10.1.23) для нечетного значения  $N$ .

Значение множителя  $G$  в общем случае находится нормировкой к 1 коэффициента передачи фильтра при  $\omega = 0$ . Для ФНЧ при использовании вышеприведенных формул значение  $G$  равно 1.

При  $z = \exp(-j\omega)$  главный диапазон функций  $H(z)$  от  $-\pi$  до  $\pi$ . Для получения передаточной функции в шкале физических частот достаточно вместо  $z$  в выражения (10.1.20, 10.1.24) подставить значение  $z = \exp(-j\omega \Delta t)$ , где  $\Delta t$  – физический интервал дискретизации данных, и проверить соответствие расчетной передаточной функции заданным условиям.

**Продолжение примера.**

8. Вычисляем значения коэффициентов  $G_m$ ,  $b_m$  и  $c_m$ :

-  $N=4$ :  $\gamma = 1.637$ ,  $G_1 = 0.203$ ,  $G_2 = 0.149$ ,  $b_1 = 0.681$ ,  $b_2 = 0.501$ ,  $c_1 = 0.492$ ,  $c_2 = 0.098$ .

-  $N=5$ :  $\gamma = 1.698$ ,  $G_1 = 0.203$ ,  $G_2 = 0.151$ ,  $b_1 = 0.763$ ,  $b_2 = 0.568$ ,  $c_1 = 0.574$ ,  $c_2 = 0.171$ .

9. Подставляем вычисленные коэффициенты в выражения (10.1.20, 10.1.24) и вычисляем значения передаточных функций при  $z = \exp(-j\omega \Delta t)$ . Графики полученных функций приведены на рис. 10.1.3. На рис. 10.1.4

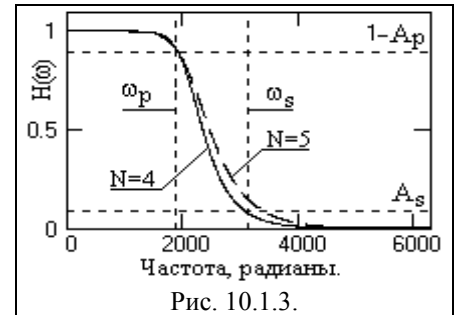


Рис. 10.1.3.

Во временной области фильтрация выполняется последовательной сверткой входного сигнала с операторами ячеек фильтра:

$$y_k = x_k \otimes \{h_0(i)\} \otimes h_1(i) \otimes \dots \otimes h_M(i), \quad i = 0, 1, 2.$$

Уравнение рекурсивной фильтрации для  $m$ -го оператора фильтра:

$$y_k = G_m (x_k + 2x_{k-1} + x_{k-2}) + b_m y_{k-1} - c_m y_{k-2}. \quad (10.1.25)$$

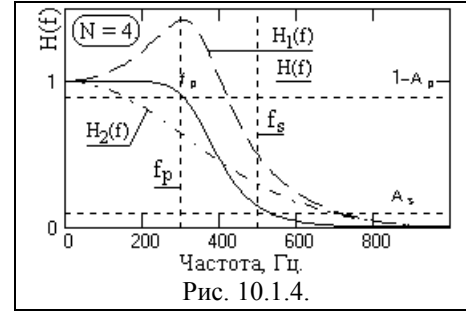
Уравнение рекурсивной фильтрации для дополнительного  $h_0(i)$  линейного оператора фильтра при нечетном  $N$ :

$$y_0 = (x_k + x_{k-1})/(\gamma+1) + y_{k-1} \cdot (\gamma-1)/(\gamma+1) \quad (10.1.26)$$

**Продолжение примера.**

10. Каждый оператор фильтра имеет определенную передаточную функцию, что можно видеть на рис. 10.1.4. Порядок последовательной свертки сигнала с операторами фильтра значения не имеет, но с учетом разрядности ячеек памяти звено  $H_1(f)$  целесообразно реализовать за  $H_2(f)$ .

11. Для оценки длительности импульсной реакции фильтра подаем на вход фильтра импульс Кронекера на отсчете  $k = 3$ , и начинаем фильтрацию со второго отсчета (что обеспечивает начальные условия фильтрации на точках  $k=0$  и  $k=1$ ). Коэффициент усиления дисперсии шумов (сумма квадратов значений импульсного отклика) равен 0.341 при  $N=5$ , и 0.278 при  $N=4$ .



## 10.2. ВЫСОКОЧАСТОТНЫЙ ФИЛЬТР БАТТЕРУОРТА

**Синтез фильтров методом частотного преобразования.** Высокочастотные и полосовые фильтры конструируются путем частотной трансформации передаточных функций фильтров низких частот. Если обозначить аргумент передаточных функций ФНЧ через  $p=jW$ , а функций ФВЧ и ПФ через  $s=jw$ , то всегда можно найти такую функцию частотного преобразования  $p=F(s)$ , которая превращает один тип фильтров в другой. Для преобразования ФНЧ  $\rightarrow$  ФВЧ функция частотного преобразования имеет вид:

$$p = 1/s, \quad (10.2.1)$$

В этом нетрудно убедиться сравнением двух видов преобразования. Как известно, передаточная функция ФВЧ может быть получена из ФНЧ разностью между широкополосным фильтром ( $H(\omega)=1$ ) и ФНЧ. Применяя этот метод для функции Баттеруорта, получаем:

$$|H(w)|^2 = 1 - |H(W)|^2 = 1 - 1/(1+W^{2N}) = W^{2N}/(1+W^{2N}). \quad (10.2.2)$$

С другой стороны, при  $W = p/j$ :  $|H(p)|^2 = 1/(1-p^{2N})$ . Выполняя подстановку (10.2.1) в это выражение, получаем:

$$|H(s)|^2 = s^{2N}/(s^{2N}-1).$$

Возвратимся из последнего выражения к аргументу  $w$  с учетом принятого равенства  $s=jw$ :

$$|H(s)|^2 = (jw)^{2N}/((jw)^{2N}-1) = (w)^{2N}/(1+(w)^{2N}),$$

что полностью повторяет (10.2.2) при  $w=W$ .

Подставляя  $p=1/s$  непосредственно в выражение  $H(p)$  (10.1.16) для четного значения  $N$ , получаем:

$$H(s) = G \prod_{m=1}^{N/2} s^2/(s^2+a_m s+1). \quad (10.2.3)$$

Для нечетного  $N$ :

$$H(s) = [G \cdot s/(s+1)] \prod_{m=1}^{N/2} s^2/(s^2+a_m s+1). \quad (10.2.4)$$

После билинейного  $z$ -преобразования выражения с подстановкой  $s=\gamma(1-z)/(1+z)$ , для четного и нечетного значений  $N$  соответственно:

$$H(z) = G \prod_{m=1}^{N/2} \gamma^2 \cdot G_m \cdot (1-z)^2/(1-b_m z+c_m z^2). \quad (10.2.5)$$

$$H(z) = G \frac{\gamma(1-z)/(\gamma+1)}{1-z(\gamma-1)/(\gamma+1)} \prod_{m=1}^{N/2} \gamma^2 \cdot G_m \cdot (1-z)^2/(1-b_m z+c_m z^2). \quad (10.2.6)$$

$$G_m = 1/(\gamma^2 + a_m \gamma + 1). \quad (10.2.7)$$

$$b_m = 2 \cdot G_m (\gamma^2 - 1).$$

$$c_m = G_m (\gamma^2 - a_m \gamma + 1).$$

Значения коэффициентов  $G_m$ ,  $b_m$ ,  $c_m$  остаются без изменения (сравнить с (10.1.21-

10.1.23)). При задании частотных параметров ФВЧ в том же виде, что и для ФНЧ, формула расчетов  $N$  и  $\omega_{dc}$  получается аналогично ФНЧ, при этом в знаменателе выражения (10.1.6) отношение  $\omega_{dp}/\omega_{ds}$  заменяется на  $\omega_{ds}/\omega_{dp}$ :

$$N = \ln [\delta / \sqrt{1/A_s^2 - 1}] / \ln(\omega_{ds}/\omega_{dp}), \quad (10.2.8)$$

а в (10.1.7) деление членов правой части меняется на умножение:

$$\omega_{dc} = \omega_{dp} \cdot \delta^{1/N}. \quad (10.2.9)$$

Уравнение рекурсивной фильтрации для  $m$ -го оператора фильтра:

$$y_k = \gamma^2 \cdot G_m (x_k - 2x_{k-1} + x_{k-2}) + b_m y_{k-1} - c_m y_{k-2}. \quad (10.2.10)$$

Уравнение рекурсивной фильтрации для дополнительного  $h_0(i)$  линейного оператора фильтра при нечетном  $N$ :

$$y_0 = \gamma \cdot (x_k - x_{k-1}) / (\gamma + 1) + y_{k-1} \cdot (\gamma - 1) / (\gamma + 1). \quad (10.2.11)$$

### Пример расчета фильтра высоких частот Баттеруорта.

Техническое задание:

- Шаг дискретизации данных  $\Delta t = 0.0005$  сек. Частота Найквиста  $f_N = 1/2\Delta t = 1000$  Гц,  $\omega_N = 6.283 \cdot 10^3$  рад.

- Граничная частота полосы пропускания:  $f_p = 700$  Гц,  $\omega_p = 4.398 \cdot 10^3$  рад.

- Граничная частота полосы подавления:  $f_s = 500$  Гц,  $\omega_s = 3.142 \cdot 10^3$  рад.

- Коэффициенты неравномерности:  $A_p = A_s = 0.1$ .

Расчет дополнительных параметров:

1.  $\delta = A_p \sqrt{2/A_p - 1} / (1 - A_p)$ :  $\delta = 0.484$ .

2. Деформированные частоты по формуле (10.1.4):  $\omega_{dp} = 7.85 \cdot 10^3$  рад.  $\omega_{ds} = 4 \cdot 10^3$  рад.

3. Порядок фильтра по формуле (10.2.8):  $N = 4.483$ . Для расчетов принимаем  $N=4$ .

4. Частота среза фильтра по формуле (10.2.9):

$$\omega_{dc} = 6.549 \cdot 10^3 \text{ рад (1042 Гц)},$$

5. Строим график функции  $H(w)$ ,  $w = \omega/\omega_{dc}$ , (рис.10.2.1).

6. Полюса  $p_n$  фильтра полностью повторяют полюса ФНЧ (рис. 10.1.2), а, соответственно, повторяются и значения коэффициентов  $a_m$ . Остальные коэффициенты:  $\gamma = 0.611$ ,  $G_1 = 0.543$ ,  $G_2 = 0.4$ ,  $b_1 = -0.681$ ,  $b_2 = -0.501$ ,  $c_1 = 0.492$ ,  $c_2 = 0.098$ .



Рис. 10.2.1.

При сравнении коэффициентов  $b_m$ ,  $c_m$  и коэффициентов в числителе передаточных функций ФВЧ с соответствующими коэффициентами ФНЧ предыдущего примера можно заметить, что в данном фильтре относительно ФНЧ произошла только смена знаков коэффициентов при нечетных степенях  $z$ . Это объясняется тем, что заданные в данном примере параметры ФВЧ по частоте соответствуют частотному реверсу ФНЧ:  $\omega' = \pi - \omega$ , что приводит к частотному реверсу передаточной функции низкочастотного фильтра и превращению его в высокочастотный фильтр. Этот способ обращения ФНЧ также может использоваться для расчетов ФВЧ.

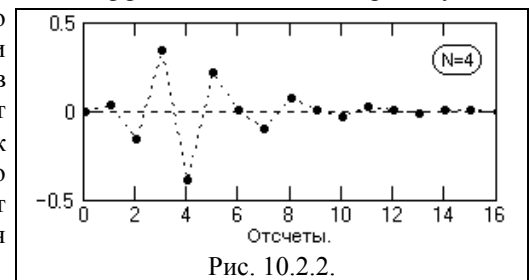


Рис. 10.2.2.

7. Импульсная реакция фильтра, вычисленная по (10.2.10) при подаче на вход фильтра импульса Кронекера приведена на рис. 10.2.2.

### 10.3. ПОЛОСОВОЙ ФИЛЬТР БАТТЕРУОРТА

Как известно, полосовой фильтр можно получить непосредственной комбинацией низкочастотного и высокочастотного фильтра при перекрытии полосы пропускания фильтров. Аналогичный эффект достигается и частотным преобразованием ФНЧ, которое в этом случае имеет вид:

$$p = s + 1/s. \quad (10.3.1)$$

Подставив в (10.3.1) значения  $p = jW$  и  $s = jw$ , получим:

$$W = [w^2 - 1]/w, \quad w^2 - Ww - 1 = 0. \quad (10.3.2)$$

Корни уравнения (10.3.2):

$$(w)_{1,2} = W/2 \pm \sqrt{(W/2)^2 + 1}. \quad (10.3.3)$$

**Расщепление спектра.** При  $W=0$  имеем  $w = \pm 1$ , т.е. центр полосы пропускания ФНЧ (от  $-W_c$  до  $+W_c$ ) расщепляется на два (как и положено, для полосовых фильтров) и смещается в точки  $w = \pm 1$ . Подставив в (10.3.3) граничную частоту  $W_c=1$  нормированного ФНЧ, определяем граничные частоты нормированного полосового фильтра в виде пары сопряженных частот:

$$w_1 = \pm 0.618, \quad w_2 = \pm 1.618$$

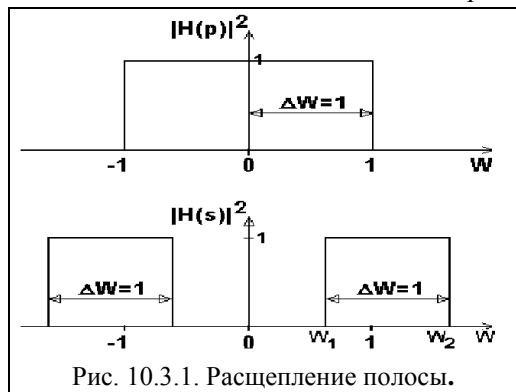


Рис. 10.3.1. Расщепление полосы.

Сущность произведенного преобразования наглядно видна на рис. 10.3.1. Ширина полосы пропускания нормированного ПФ равна 1.

Полученное преобразование можно распространить на полосовой фильтр с ненормированными частотами  $\omega_H$  и  $\omega_B$ .

Введем понятие геометрической средней частоты фильтра  $\omega_0$ :

$$\omega_0 = \sqrt{\omega_H \omega_B}. \quad (10.3.4)$$

Ширина полосы пропускания ПФ связана (см. рис.10.3.1) с граничной частотой ФНЧ соотношением:

$$\Delta\omega = \omega_B - \omega_H = \omega_c = \omega_H.$$

В долях средней геометрической частоты:

$$W_H = (\omega_B - \omega_H) / \omega_0 = W_c. \quad (10.3.5)$$

Заменяя в (10.3.4-10.3.5) значение  $\omega_B$  на произвольную частоту  $\omega$  и подставляя в (10.3.5) значение  $\omega_H = \omega \cdot \omega_0^2$  из (10.3.4), получаем произвольную частоту  $W$ :

$$W = (\omega - \omega_H) / \omega_0 = \omega / \omega_0 - \omega_0 / \omega. \quad (10.3.6)$$

Отсюда, в выражении (10.1.1) вместо нормированной частоты  $W = \omega / \omega_c$  можно применить функцию частоты полосового фильтра  $w(\omega)$ :

$$w(\omega) = (\omega^2 - \omega_0^2) / [\omega(\omega_B - \omega_H)],$$

или, подставляя (10.3.4) вместо  $\omega_0$ :

$$w(\omega) = (\omega^2 - \omega_H \omega_B) / [\omega(\omega_B - \omega_H)]. \quad (10.3.7)$$

Тем самым передаточная функция ФНЧ выражается в единицах, которые позволяют после применения преобразования (10.3.1) использовать для задания необходимые граничные частоты  $\omega_H$  и  $\omega_B$  полосового фильтра.

### Пример расчета полосового фильтра Баттеруорта.

Техническое задание:

- Шаг дискретизации данных  $\Delta t = 0.0005$  сек. Частота Найквиста  $f_N = 1/2\Delta t = 1000$  Гц,  $\omega_N = 6.283 \cdot 10^3$  рад.

- Нижняя граничная частота полосы пропускания:  $f_H = 340$  Гц,  $\omega_H = 2.136 \cdot 10^3$  рад.

- Верхняя граничная частота полосы пропускания:  $f_B = 470$  Гц,  $\omega_B = 2.953 \cdot 10^3$  рад.

- Крутизна срезов в децибелах на октаву:  $K_p = 45$ .

Расчет параметров:

1. Порядок фильтра по формуле (10.1.6'):  $N = K_p/6 = 45/6 = 7.5$ .

Для расчетов принимаем  $N=8$ .

2. Строим график функции  $H(\omega) = \sqrt{1/(1+w(\omega)^{2N})}$  с

использованием выражения (10.3.7). Передаточная характеристика фильтра приведена на рис. 10.3.2.

3. Деформированные частоты по формуле (10.1.4):

$$\omega_{dH} = 2.366 \cdot 10^3 \text{ рад.} \quad \omega_{dB} = 3.64 \cdot 10^3 \text{ рад.} \quad \omega_{d0} = 2.934 \cdot 10^3.$$

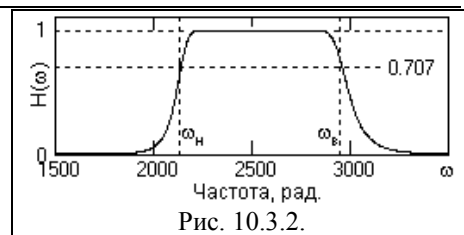


Рис. 10.3.2.

**Полосовой фильтр на s-плоскости.** С учетом деформации частот, принимаем  $p = j\omega = j(\omega^2 - \omega_{dH}\omega_{dB}) / [\omega(\omega_{dB} - \omega_{dH})]$ ,  $s = j\omega$  и заменяем  $\omega = s/j$  в выражении p:



$$p = (s^2 + \omega_{dH}\omega_{dB}) / [s(\omega_{dB} - \omega_{dH})],$$

$$s^2 - p(\omega_{dB} - \omega_{dH})s + \omega_{dH}\omega_{dB} = 0. \quad (10.3.8)$$

Корни уравнения (10.3.8) определяют местоположение полюсов ПФ:

$$s = s^* = p(\omega_{dB} - \omega_{dH})/2 \pm \sqrt{[p(\omega_{dB} - \omega_{dH})/2]^2 - \omega_{dB}\omega_{dH}}. \quad (10.3.9)$$

Уравнение (10.3.9) показывает расщепление каждого  $p$ -полюса, определяемых выражением (10.1.14), на два комплексно сопряженных полюса  $s$ -плоскости, произведение которых будет давать вещественные биквадратные блоки в  $s$ -плоскости. При этом следует учесть то обстоятельство, что устойчивому рекурсивному фильтру на  $z$ -плоскости должны соответствовать полюса только одной (левой) половины  $p$ - и  $s$ -плоскостей.

**Передаточная функция.** При применении преобразования (10.3.1) к передаточной функции в полиномиальной форме (10.1.11), получаем:

$$H(p) = G \prod_{m=1}^N 1/(p-p_m) \Leftrightarrow G \prod_{m=1}^N s/(s^2 - p_m s + 1) = H(s), \quad (10.3.10)$$

Выражение (10.3.10) не требует нахождения полюсов, т.к. они уже известны и определяются выражением (10.3.9). С учетом этого функция  $H(s)$  может быть записана с объединением в биквадратные блоки комплексно сопряженных полюсов с вещественными коэффициентами:

$$H(s) = G \prod_{m=1}^N s / [(s-s_m)(s-s_m^*)] = G \prod_{m=1}^N s / (s^2 + a_m s + g_m), \quad (10.3.11)$$

где значения  $a_m$  и  $g_m$  могут быть определены непосредственно по полюсам (10.3.9):

$$a_m = -2 \operatorname{Re} s_m, \quad g_m = (\operatorname{Re} s_m)^2 + (\operatorname{Im} s_m)^2 = |s_m|^2. \quad (10.3.12)$$

#### Продолжение расчета.

4. Полюса фильтра на единичной окружности в  $p$ -плоскости:

$$p_n = j \cdot \exp[j \cdot \pi(2n-1)/2N], \quad n = 1, 2, \dots, N.$$

Положение полюсов приведено на рис. 10.3.3.

5. Полюса в левой половине  $s$ -плоскости,  $n = 1, 2, \dots, 2N$

(приведены на рис. 10.3.4):

$$s_n := p_{\operatorname{int}(\frac{n+1}{2})} \cdot \frac{(\omega_{dB} - \omega_{dH})}{2} + (-1)^{n-1} \cdot \sqrt{\left[ p_{\operatorname{int}(\frac{n+1}{2})} \cdot \frac{(\omega_{dB} - \omega_{dH})}{2} \right]^2 - \omega_{dB} \cdot \omega_{dH}}$$

6. По полученным значениям полюсов вычисляем коэффициенты  $a_m$  и  $g_m$  (10.3.12),  $m = n$ .

$$a_m = 196.8, 300.4, 581.2, 834.5, 930.5, 1188, 1196, 1304.$$

$$g_m = 5.64 \cdot 10^6, 1.314 \cdot 10^7, 5.997 \cdot 10^6, 1.236 \cdot 10^7, 6.742 \cdot 10^6, 1.1 \cdot 10^7, 7.895 \cdot 10^6, 9.39 \cdot 10^6.$$

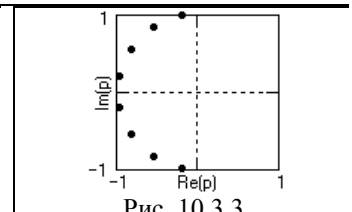


Рис. 10.3.3.

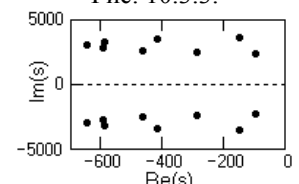


Рис. 10.3.4.

По приведенному примеру можно заметить, что при использовании ненормированных частот  $\omega$ , достаточно существенных по своей величине, значения  $s$ -полюсов и, соответственно, величины коэффициентов  $a_m$  и  $g_m$  имеют большие порядки, что нежелательно для дальнейших расчетов и может приводить к появлению погрешностей при ограничении разрядности. Для исключения этого фактора значения полюсов  $s_n$  рекомендуется нормировать на среднюю геометрическую частоту:

$$s_n = s_n / \omega_0.$$

#### Продолжение расчета.

6'. Значения коэффициентов  $a_m$  и  $g_m$  (10.3.12), вычисленные по нормированным значениям  $s_n$ .

$$a_m = 0.067, 0.102, 0.198, 0.284, 0.317, 0.405, 0.407, 0.444.$$

$$g_m = 0.655, 1.527, 0.697, 1.436, 0.783, 1.277, 0.917, 1.091.$$

Коэффициент  $\gamma$  билинейного преобразования для ненормированных значений  $\omega$  и полюсов  $s_n$  имеет классическую форму:  $\gamma = 2/\Delta t$ . Соответственно, для нормированных значений:  $\gamma = 2/(\Delta t \cdot \omega_0)$ . После билинейного  $z$ -преобразования выражения (10.3.11), получаем:

$$H(z) = G \prod_{m=1}^N G_m (1-z^2)/(1-b_m z+c_m z^2). \quad (10.3.13)$$

$$G_m = 1/(\gamma+a_m+g_m\gamma^{-1}). \quad (10.3.14)$$

$$b_m = 2G_m(\gamma-g_m\gamma^{-1}). \quad (10.3.15)$$

$$c_m = G_m(\gamma-a_m+g_m\gamma^{-1}). \quad (10.3.16)$$

**Продолжение расчета** (по нормированным полюсам  $s_n$ ).

7. Значения коэффициента  $\gamma$ :  $\gamma = 1.363$ .

8. Значения  $G_m$  по (10.3.14):  $G_m = 0.523, 0.387, 0.483, 0.37, 0.444, 0.37, 0.409, 0.384$ .

9. Значения  $b_m$  по (10.3.15):  $b_m = 0.924, 0.188, 0.823, 0.23, 0.7, 0.315, 0.565, 0.432$ .

10. Значения  $c_m$  по (10.3.16):  $c_m = 0.93, 0.921, 0.809, 0.789, 0.719, 0.701, 0.666, 0.659$ .

11. Общий нормировочный множитель  $G$ :  $G = 1.264 \cdot 10^{-3}$ .

12. Заключительная передаточная функция:

$$H(\omega) = G \prod_{m=1}^N \frac{G_m [1 - (\exp(-j \cdot \omega \cdot \Delta t))^2]}{1 - b_m \cdot \exp(-j \cdot \omega \cdot \Delta t) + c_m \cdot \exp(-j \cdot \omega \cdot \Delta t)^2}$$

При построении графика данной функции можно убедиться, что она полностью соответствует рис. 10.3.2.

13. Уравнение одной секции фильтра:

$$y_{m,k} = G_m (y_{m-1,k} - y_{m-1,k-2}) + b_m y_{m,k-1} - c_m y_{m,k-2}.$$

Нормировкой  $H(z)$  к 1 на геометрической средней частоте фильтра определяют общий множитель  $G$ :

$$G = 1/H(\exp(-j\Delta t\omega_0)). \quad (10.3.17)$$

Если применить обратное частотное преобразование  $p = s(\omega_B - \omega_H)/(s^2 + \omega_B \omega_H)$ , то в результате будет получен полосовой заградительный фильтр.

#### 10.4. ФИЛЬТРЫ ЧЕБЫШЕВА

**Фильтры первого рода.** Фильтры Чебышева с пульсациями передаточной функции в полосе пропускания и гладким затуханием в полосе подавления называют фильтрами Чебышева первого рода, в отличие от инверсных фильтров Чебышева (второго рода). Аппроксимационная формула фильтров Чебышева первого рода определяется выражением:

$$|H(W)|^2 = 1/[1 + \delta_N^2 T_N^2(W)], \quad (10.4.1)$$

где  $T_N(W)$  - многочлен Чебышева  $N$ -го порядка:

$$T_n(W) = \cos(n \arccos(W)), \quad W \leq 1. \quad (10.4.2)$$

$$= \operatorname{ch}(n \operatorname{arcch}(W)), \quad W > 1. \quad n = 1, 2, \dots$$

**Критерий приближения Чебышева**, который широко используется не только в теории фильтров - минимум максимальной ошибки приближения (минимаксное приближение). В соответствии с этим приближением параметры передаточной функции подбираются таким образом, чтобы в полосе передачи АЧХ наблюдались равноволновые пульсации коэффициента передачи, которые являются "платой" за повышение крутизны среза фильтра.

Полиномы Чебышева вычисляются по рекуррентной формуле:

$$T_n(W) = 2W T_{n-1}(W) - T_{n-2}(W), \quad (10.4.3)$$

$$T_1(W) = W, \quad T_0(W) = 1.$$

Для ФНЧ при  $W = \omega/\omega_p$  имеет место  $T_n(1) = 1$ ,  $|H(W)|^2 = 1/(1+\delta^2)$  и значением  $\delta$  задается коэффициент пульсаций в полосе передачи. При задании полосы по уровню  $A_p$  значение  $\delta$  рассчитывается аналогично фильтру Баттеруорта.

Соответственно, при задании  $A_s$  на границе полосы подавления, имеем:

$$1/(1+\delta^2 T_N^2(\omega_s/\omega_p)) = A_s^2. \quad (10.4.4)$$

$$N = \operatorname{arcch}[\sqrt{1/A_s^2 - 1}/\delta] / \operatorname{arcch}(\omega_s/\omega_p). \quad (10.4.5)$$

Дальнейшие расчеты идентичны расчетам фильтров Баттеруорта, равно как и частотные преобразования фильтров ФНЧ в ФВЧ и ПФ.

**Фильтры второго рода.** Для фильтров Чебышева второго рода, с гладкой передаточной характеристикой в зоне пропускания и равноволновыми пульсациями в зоне подавления, используется функция:

$$|H(W)|^2 = 1/[1 + \delta^2(T_N^2(W_s)/T_N^2(W_s/W))], \quad (10.4.6)$$

где  $W = \omega/\omega_p$ ,  $W_s = \omega_s/\omega_p$ . Условие задания параметра  $\delta$  остается без изменений. На границе полосы подавления при  $\omega = \omega_s$ :  $1 + \delta^2 T_N^2(\omega_s/\omega_p) = 1/A_s^2$ , откуда значение  $N$  также определяется аналогично фильтру первого рода. Дальнейший порядок расчетов фильтров Чебышева второго рода не отличается от фильтров первого рода.

#### 10.5. ДОПОЛНИТЕЛЬНЫЕ СВЕДЕНИЯ.

При использовании РЦФ очень часто упускается вопрос длительности фактического затухания переходного процесса. Между тем, для эффективного запуска РЦФ необходим поток входных данных  $x_n$  и множество начальных значений  $y_n$ . Если начальные значения  $y_n$  неизвестны и принимаются равными нулю, начальный переходной процесс включения неизбежен. При этом существует четкая тенденция - чем больше крутизна фильтра, тем дольше затухает переходной процесс. Поэтому РЦФ применяют, в основном, при обработке достаточно протяженных массивов. При обработке коротких массивов, длина которых соизмерима с длительностью переходного процесса РЦФ, необходим предварительный подбор начальных значений  $y_n$ . Проводится он, как правило, чисто эмпирически, с использованием различных наборов начальных данных.

Второй фактор, который следует учитывать - сдвиг фазы. Если при обработке данных сдвиг фазы входных сигналов недопустим, то следует применять либо дополнительный компенсирующий фильтр, восстанавливающий фазу обработанных сигналов, либо применять последовательную двойную фильтрацию однотипным рекурсивным фильтром с прямым и обратным проходом обрабатываемых данных.

### Тема 11. АДАПТИВНАЯ ФИЛЬТРАЦИЯ ЦИФРОВЫХ ДАННЫХ

В традиционных методах обработки данных информация извлекается из входных сигналов линейными системами с постоянными параметрами алгоритмов преобразования данных. Системы могут иметь как конечную, так и бесконечную импульсную характеристику, но передаточная функция систем не зависит от параметров входных сигналов и их изменения во времени.

Адаптивные устройства обработки данных отличаются наличием определенной связи параметров передаточной функции с параметрами входных, выходных, ожидаемых, прогнозируемых и прочих дополнительных сигналов или с параметрами их статистических соотношений, что позволяет самонастраиваться на оптимальную обработку сигналов. В простейшем случае, адаптивное устройство содержит программируемый фильтр обработки данных и блок (алгоритм) адаптации, который на основании определенной программы анализа входных, выходных и прочих дополнительных данных вырабатывает сигнал управления параметрами программируемого фильтра. Импульсная характеристика адаптивных систем также может иметь как конечный, так и бесконечный характер.

Как правило, адаптивные устройства выполняются узкоцелевого функционального назначения под определенные типы сигналов. Внутренняя структура адаптивных систем и алгоритм адаптации практически полностью регламентируются функциональным назначением и определенным минимальным объемом исходной априорной информации о характере входных данных и их статистических и информационных параметрах. Это порождает многообразие подходов при разработке систем, существенно затрудняет их классификацию и разработку общих теоретических положений [38]. Но можно отметить, что наибольшее применение при разработке систем для адаптивной обработки сигналов находят два подхода: на основе схемы наименьших квадратов (СНК) и рекурсивной схемы наименьших квадратов (РСНК).

### 11.1. ОБЩИЕ СВЕДЕНИЯ ОБ АДАПТИВНОЙ ЦИФРОВОЙ ФИЛЬТРАЦИИ

**Основные области применения** адаптивной фильтрации – очистка данных от нестабильных мешающих сигналов и шумов, перекрывающихся по спектру со спектром полезных сигналов, или когда полоса мешающих частот неизвестна, переменна и не может быть задана априорно для расчета параметрических фильтров. Так, например, в цифровой связи сильная активная помеха может интерферировать с полезным сигналом, а при передаче цифровой информации по каналам с плохими частотными характеристиками может наблюдаться межсимвольная интерференция цифровых кодов. Эффективное решение этих проблем возможно только адаптивными фильтрами.

Частотная характеристика адаптивных фильтров автоматически регулируется или модифицируется в соответствии с определенным критерием, позволяющем фильтру адаптироваться к изменениям характеристик входного сигнала. Они достаточно широко используются в радио- и гидролокации, в системах навигации, в выделении биомедицинских сигналов, и многих других отраслях техники. В качестве примера рассмотрим наиболее распространенные схемы адаптивной фильтрации сигналов.

**Адаптивный шумоподавитель.** Блок-схема фильтра приведена на рис. 11.1.1.

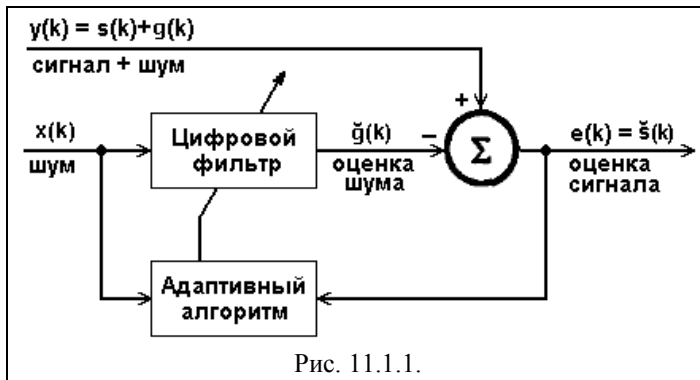


Рис. 11.1.1.

Фильтр состоит из блока цифрового фильтра с регулируемыми коэффициентами и адаптивного алгоритма для настройки и изменения коэффициентов фильтра. На фильтр одновременно подаются входные сигналы  $y(k)$  и  $x(k)$ . Сигнал  $y(k)$  содержит полезный сигнал  $s(k)$  и некоррелированный с ним загрязняющий сигнал  $g(k)$ . Сигнал  $x(k)$  какого-либо источника шума,

коррелированный с  $g(k)$ , который используется для формирования оценки сигнала  $\hat{g}(k)$ . Полезный сигнал оценивается по разности:

$$\hat{s}(k) = y(k) - \hat{g}(k) = s(k) + g(k) - \hat{g}(k). \quad (11.1.1)$$

Возводим уравнение в квадрат и получаем:

$$\hat{s}^2(k) = s^2(k) + (g(k) - \hat{g}(k))^2 + 2s(k)(g(k) - \hat{g}(k)). \quad (11.1.2)$$

Вычислим математическое ожидание левой и правой части этого уравнения:

$$M[\hat{s}^2(k)] = M[s^2(k)] + M[(g(k) - \hat{g}(k))^2] + 2M[s(k)(g(k) - \hat{g}(k))]. \quad (11.1.3)$$

Последнее слагаемое в выражении равно нулю, поскольку сигнал  $s(k)$  не коррелирует с сигналами  $g(k)$  и  $\hat{g}(k)$ .

$$M[\hat{s}^2(k)] = M[s^2(k)] + M[(g(k) - \hat{g}(k))^2]. \quad (11.1.4)$$

В этом выражении  $M[s^2(k)] = W(s(k))$  – мощность сигнала  $s(k)$ ,  $M[\hat{s}^2(k)] = W(\hat{s}(k))$  – оценка мощности сигнала  $s(k)$  и общая выходная мощность,  $M[(g(k) - \hat{g}(k))^2] = W(\varepsilon_g)$  – остаточная мощность шума, который может содержаться в выходном сигнале. При настройке адаптивного фильтра к оптимальному положению минимизируется мощность остаточного шума, а, следовательно, и мощность выходного сигнала:

$$\min W(\hat{s}(k)) = W(s(k)) + \min W(\varepsilon_g). \quad (11.1.5)$$

На мощность полезного сигнала настройка не влияет, поскольку сигнал не коррелирован с шумом. Эффект минимизации общей выходной мощности будет выражаться в максимизации выходного отношения сигнал/шум. Если настройка фильтра обеспечивает равенство  $\hat{g}(k) = g(k)$ , то при этом  $\hat{s}(k) = s(k)$ . Если сигнал не содержит шума, адаптивный алгоритм должен устанавливать нулевые значения всем коэффициентам цифрового фильтра.

**Адаптивный фильтр Винера.**

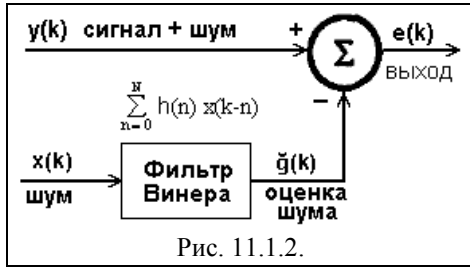


Рис. 11.1.2.

Входной сигнал  $y(k)$  фильтра, приведенного на рис. 11.1.2, включает компоненту, коррелированную с вторым сигналом  $x(k)$ , и полезную компоненту, некоррелированную с  $x(k)$ . Фильтр формирует из  $x(t)$  сигнал  $\hat{g}(k)$  - оптимальную оценку той части  $y(k)$ , которая коррелирована с  $x(k)$ , и вычитает ее из сигнала  $y(k)$ . Выходной сигнал:

$$e(k) = y(k) - \hat{g}(k) = y(k) - \mathbf{H}^T \mathbf{X}_k = y(k) - \sum_{n=0}^N h(n) x(k-n),$$

где  $\mathbf{H}^T$  и  $\mathbf{X}_k$  – векторы весовых коэффициентов фильтра и его входного сигнала.

Аналогично предыдущему методу, возводим в квадрат левую и правую части уравнения, находим математические ожидания обеих частей и получаем уравнение оптимизации  $\varepsilon$  выходного сигнала:

$$\varepsilon = \sigma^2 + 2\mathbf{P}^T \mathbf{H} + \mathbf{H}^T \mathbf{R} \mathbf{H}, \quad (11.1.6)$$

где  $\sigma^2 = M[y^2(k)]$  – дисперсия  $y(k)$ ,  $\mathbf{P} = M[y(k)\mathbf{X}_k]$  – вектор взаимной корреляции,  $\mathbf{R} = M[\mathbf{X}_k \mathbf{X}_k^T]$  – автокорреляционная матрица.

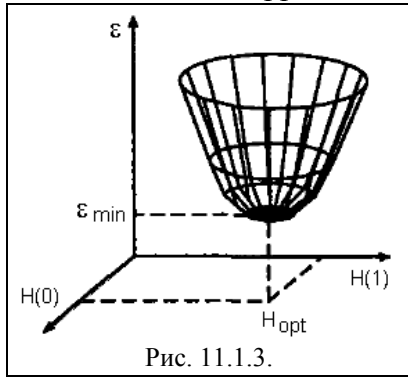


Рис. 11.1.3.

В стационарной среде график зависимости  $\varepsilon$  от коэффициентов  $\mathbf{H}$  представляет собой чашеобразную поверхность адаптации (рис. 11.1.3). Градиент поверхности:

$$\Delta = d\varepsilon / d\mathbf{H} = -2\mathbf{P} + 2\mathbf{R}\mathbf{H}.$$

Каждому набору коэффициентов  $h(n)$  на этой поверхности соответствует определенная точка. В точке минимума градиент равен нулю и вектор весовых коэффициентов фильтра является оптимальным:

$$\mathbf{H}_{opt} = \mathbf{R}^{-1} \mathbf{P}. \quad (11.1.7)$$

Эта формула называется уравнением Винера-Хопфа.

Задачей алгоритма автоматической настройки является подбор таких весовых коэффициентов фильтра, которые обеспечивают работу в оптимальной точке поверхности адаптации.

Однако практическое применение фильтра затрудняется использованием корреляционных матриц  $\mathbf{R}$  и  $\mathbf{P}$ , априори неизвестных, и которые могут изменяться со временем для нестационарных сигналов.

**Адаптивный алгоритм наименьших квадратов Уидроу-Хопфа.** По существу, это модификация фильтра Винера, в которой вместо вычисления коэффициентов (11.1.7) за один шаг используется алгоритм последовательного спуска в оптимальную точку при обработке каждой выборки:

$$\mathbf{H}_{k+1} = \mathbf{H}_k - \mu e_k \mathbf{X}_k, \quad (11.1.8)$$

$$e_k = y_k - \mathbf{H}^T \mathbf{X}_k. \quad (11.1.9)$$

Условие сходимости к оптимуму:

$$0 < \mu < 1/\lambda_{max}, \quad (11.1.10)$$

где  $\mu$  - параметр скорости спуска,  $\lambda_{max}$  – максимальное собственное значение ковариационной матрицы данных. Блок-схема алгоритма приведена на рис. 11.1.4.

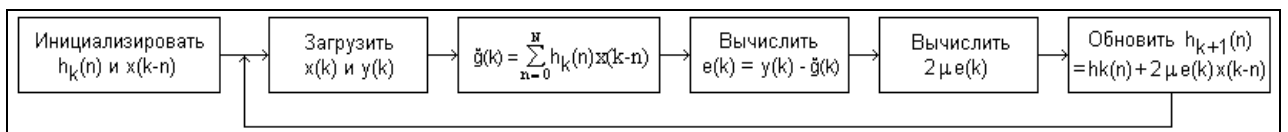


Рис. 11.1.4. Алгоритм адаптации методом наименьших квадратов.

На практике точка максимальной оптимальности флуктуирует около теоретически возможной. Если входной сигнал нестационарный, то изменение статистик сигнала

должно происходить достаточно медленно, чтобы коэффициенты фильтра успевали следить за этими изменениями.

**Рекурсивные схемы наименьших квадратов** отличаются тем, что вычисление каждой последующей выборки коэффициентов  $h(n)$  производится не только по коэффициентам только одной предыдущей выборки, но и с определенной длиной постепенно затухающей памяти по предшествующим выборкам, что позволяет снижать флюктуации оценок при обработке стационарных сигналов.

## 11.2. ОСНОВЫ СТАТИСТИЧЕСКОЙ ГРУППИРОВКИ ИНФОРМАЦИИ.

При построении систем адаптивной фильтрации данных большое значение имеют статистические характеристики обрабатываемых сигналов и шумов, их стационарность, и наличие какой-либо дополнительной информации, коррелированной с основной. Возможность использования дополнительной информации при построении адаптивных систем рассмотрим на конкретном примере – системе адаптивной фильтрации данных непрерывных ядерногеофизических измерений.

**Предпосылки метода.** Физической величиной, регистрируемой в процессе ядерно-физических измерений в геофизике, обычно является частота импульсных сигналов на выходе детекторов ионизирующего излучения в интегральном или дифференциальном режиме амплитудной селекции. Значения измеряемой величины, как статистически распределенной по своей природе, могут быть определены только путем усреднения числа актов регистрации ионизирующих частиц по интервалам времени. Зарегистрированное количество импульсов определяет статистическую погрешность единичного измерения, а временной интервал усреднения, обеспечивающий нормативную погрешность – их производительность. Для методов с непрерывной регистрацией информации во времени (или в пространстве) временное окно измерений определяет также временную (или пространственную, с учетом скорости перемещения детектора) разрешающую способность интерпретации результатов измерений, при этом эффективность регистрации информации обычно ограничена условиями измерений и/или техническими средствами их исполнения. Типичный пример - каротаж скважин, где возможности увеличения интенсивности потоков информации ограничены параметрами эффективности регистрации и чувствительности детекторов излучения, которые зависят от их типа и размеров. Размеры детекторов, естественно, существенно зависят от размеров скважинных приборов, которые, в свою очередь, ограничены диаметрами скважин.

Ниже рассматривается возможность повышения точности и производительности непрерывных ядерно-физических измерений, для наглядности, применительно к условиям измерений в варианте скважинного гамма-опробования, хотя в такой же мере она может быть использована в авто- и аэрогаммасъемке, при радиометрическом обогащении руд, в рентгенорадиометрии и других методах ядерной геофизики. Предполагается, что регистрация данных производится в цифровой форме с накоплением отсчета по постоянным интервалам дискретизации данных (по времени и по пространству, при условии постоянной скорости перемещения детектора).

В общем случае полезная (целевая) информация может присутствовать в нескольких энергетических интервалах спектра излучения. Рабочими интервалами измерений обычно считаются участки спектра, где полезная информация присутствует в "чистом" виде либо в смеси с помехами (фоном), значение которых может быть учтено при обработке результатов измерений. Так, например, при гамма-опробовании пород на содержание естественных радионуклидов (ЕРН) регистрируется излучение с энергией более 250-300 кэВ, представленное в основном первичными и однократно рассеянными квантами, плотность потока которых пропорциональна массовой доле ЕРН в породах. Плотность потока излучения в низкоэнергетическом интервале спектра (20-250 кэВ, в основном многократно рассеянное излучение) также зависит от массовой доли ЕРН, но эта зависимость является параметрически связанной с эффективным атомным номером излучающе-поглощающей среды в области детектора, вариации которого по стволу

скважины могут приводить к большой погрешности интерпретации результатов измерений. Между тем плотность потока информации (относительно массовой доли ЕРН) в интервале 20-250 кэВ много выше, чем в интервале более 250 кэВ, особенно при регистрации излучения сцинтилляционными детекторами малых объемов, которые имеют повышенную чувствительность именно к низкоэнергетической части спектра излучения.

**Задача статистической группировки** информации в потоках сигналов в общей и наиболее простой форме может быть сформулирована следующим образом. Полезная информация присутствует в двух статистически независимых потоках сигналов (в двух неперекрывающихся интервалах спектра излучения). В первом потоке сигналов, условно-основном, полезная информация присутствует в "чистом" виде: плотность потока сигналов пропорциональна определяемой физической величине. Во втором потоке, условно-дополнительном, на полезную информацию наложено влияние дестабилизирующих факторов, значение которых неизвестно. При отсутствии дестабилизирующих факторов коэффициент корреляции средних значений плотностей потоков в этих двух потоках сигналов постоянен и близок к 1. Для снижения статистической погрешности измерений требуется осуществить извлечение полезной информации из дополнительного потока сигналов и ее суммирование с основным потоком.

Обозначим потоки, а равно и частоты основного и дополнительного потоков сигналов индексами  $n$  и  $m$  (импульсов в секунду), связь потоков по частотам индексом  $x = m/n$ . Определению подлежит частота потока  $n$ . Значение  $x$  может изменяться за счет влияния дестабилизирующих факторов на поток  $m$  и в общем случае представляет собой случайную величину, распределенную по определенному закону с плотностью вероятностей  $P(x)$ , математическим ожиданием  $\bar{x}$ , и дисперсией  $D_x$ .

На основе теоремы Байеса, плотность вероятностей распределения частоты  $n$  по измеренному за единичный интервал  $t$  числу отсчетов сигнала  $N$  определяется выражением:

$$P_N(n) = P(n) P_n(N) / P(N), \quad (11.2.1)$$

$$P_n(N) = (nT)^N e^{-nT} / N!, \quad (11.2.2)$$

$$P(N) = \int_n P_n(N) P(n) dn, \quad (11.2.3)$$

где:  $P(n)$  - априорная плотность вероятностей частоты  $n$ ,  $P_n(N)$  - апостериорное распределение вероятностей числовых отсчетов  $N$  (закон Пуассона). Принимая в дальнейшем в качестве искомой величины значения отсчетов  $z = nT$  по интервалам  $T$  (экспозиция цифровых отсчетов или скользящее временное окно аналоговых данных) и подставляя (11.2.2, 11.2.3) в (11.2.1), получаем:

$$P_N(z) = P(z) z^N e^{-z} / \int_z P(z) z^N e^{-z} dz. \quad (11.2.4)$$

При неизвестном распределении значений  $z$  априорная плотность распределения  $P(z)$  принимается равномерной от 0 до  $\infty$ , при этом из выражения (11.2.4) следуют общеизвестные выражения:

$$z = D_z = N+1 \cong N, \quad (11.2.5)$$

$$\delta_z^2 = D_z / z^2 = 1 / (N+1) \cong 1/N. \quad (11.2.6)$$

Значениями единиц в выражениях пренебрегаем, что не только корректно в условиях "хорошей" статистики, но и необходимо в режиме последовательных непрерывных измерений для исключения смещения средних значений.

Как следует из теории гамма-каротажа (ГК) и достаточно хорошо подтверждено практикой гамма-опробования, пространственная разрешающая способность гамма-каротажных измерений при интерпретации результатов ГК на содержание естественных радиоактивных элементов в породах по стволу скважин в среднем составляет 10 см, а в скважинах малого диаметра может даже повышаться до 5-7 см. Однако реализация такой разрешающей способности возможна только в условиях достаточно "хорошей" статистики. Коэффициент усиления дисперсии помех цифровых фильтров деконволюции, которые используются при интерпретации ГК, в среднем порядка 12 и изменяется от 4 до

25 в зависимости от плотности пород, диаметра скважин, диаметра скважинных приборов и пр. Отсюда следует, что для достижения разрешающей способности в 10 см при нормативной погрешности дифференциальной интерпретации не более 10-20 % статистическая погрешность измерений не должна превышать 3-7 %. А это, в свою очередь, определяет объем отсчета за единичную экспозицию не менее 200-1000 импульсов. При гамма-каротаже последнее возможно только для пород с относительно высоким содержанием ЕРН (более 0.001 % эквивалентного урана), при использовании детекторов больших размеров (с эффективностью регистрации более 10 имп/сек на 1 мкР/час) и при низкой скорости каротажа (не более 100-300 м/час). В той или иной мере эта проблема характерна для всех методов ядерной геофизики, и особенно остро стоит в спектротрических модификациях измерений.

Вместе с тем следует отметить, что процесс непрерывных измерений имеет определенную физическую базу как для применения методов регуляризации результатов интерпретации данных, так и для регуляризации непосредственно самих статистических данных (массивов отсчетов  $N$ ) при их обработке.

Простейшим способом подготовки цифровых данных для интерпретации является их низкочастотная фильтрация методом наименьших квадратов (МНК) или весовыми функциями (Лапласа-Гаусса, Кайзера-Бесселя и др.). Однако любые методы низкочастотной фильтрации данных снижают пространственную разрешающую способность интерпретации, так как кроме снижения статистических флуктуаций приводят к определенной деформации частотных составляющих полезной части сигнала, спектр которого по условиям деконволюции должен иметь вещественные значения вплоть до частоты Найквиста. В определенной мере ликвидировать этот негативный фактор позволяет метод адаптивной регуляризации данных (АРД).

Выражения (11.2.5-6) получены в предположении полной неизвестности априорного распределения  $P(z)$  для отсчетов в каждой текущей экспозиции  $\tau$ . Между тем, при обработке данных непрерывных измерений, и тем более каротажных данных, которые обычно являются многопараметровыми, для каждого текущего отсчета при обработке данных может проводиться определенная оценка распределения  $P(z)$ . Как минимум, можно выделить два способа оценки распределения  $P(z)$ .

Способ 1. По массивам данных параллельных измерений каких-либо других информационных параметров, значения которых достаточно четко коррелированы с обрабатываемым массивом данных либо в целом по пространству измерений, либо в определенном скользящем интервале сравнения данных. К таким массивам относятся, например, предварительные каротажные измерения в процессе бурения скважин, измерения другим прибором, с другой скоростью каротажа, в другом спектральном интервале излучения, и даже другим методом каротажа. При гамма-опробовании оценка распределения  $P(z)$  может производиться по параллельным измерениям интенсивности потока  $m$  в низкочастотном интервале спектра горных пород.

Способ 2. При единичной диаграмме ГК оценка распределения  $P(z)$  в каждой текущей точке обработки данных может выполняться по ближайшим окрестностям данной точки, захватывающим более широкий пространственный интервал по сравнению с интервалом отсчетов.

**Использование априорных данных.** Допустим, что кроме основного массива данных  $N$ , подлежащего обработке (подготовке к интерпретации), мы располагаем дополнительным массивом данных  $M$ , значения которого в определенной степени коррелированы с массивом  $N$ . При отсутствии дополнительных массивов способ 2 позволяет получить массив  $M$  обработкой массива  $N$  цифровым фильтром МНК (или любым другим весовым фильтром) со скользящим временным окном  $T \geq 3\tau$  ( $M(k) = m(k)\tau$  сглаженного сигнала  $m(k) = n(k) \otimes h$ , где  $h$  – оператор симметричного цифрового фильтра). Отметим также, что 2-ой способ всегда может использоваться для



регуляризации данных независимо от наличия данных для 1-го метода.

Массив  $M$  позволяют дать оценку статистических характеристик распределения  $P(z)$ . Так, если для тех же интервалов времени  $\tau$  в массиве  $M$  имеются отсчеты  $M = m_{k\tau}$  (или приведенные к ним отсчеты какого-либо другого параметра), то можно записать:

$$P_M(z) = \int_x x \frac{(xz)^M}{M!} P(x) e^{-xz} dx, \quad (11.2.7)$$

где  $P(x)$  – априорная плотность распределения значений  $x_k = m_k/n_k$ , которые в общем случае также могут быть случайными. При равномерном распределении  $P(x)$  от 0 до  $\infty$  для отсчета  $M$  равновероятно любое значение  $z$ , т.е. эффект от измерений в потоке  $m$  отсутствует. Однако по исходным условиям задачи в потоке  $m$  обязательно присутствие полезной информации, а, следовательно, и существование, как минимум, определенных границ распределения  $P(x)$  от  $x_{\min} > 0$  до  $x_{\max} \ll \infty$ , и среднего значения  $\bar{x}$  по пространству измерений. При этом из выражения (11.2.7) следует, что наиболее вероятное значение  $z_a$ , "априорное" для отсчетов  $z = n\tau$  в потоке  $n$  по измерениям в потоке  $m$  (отсчетам  $M$ ), должно быть равно:

$$z_a = (M+1)/\bar{x} \cong M/\bar{x}. \quad (11.2.8)$$

При статистической независимости величин  $x$  и  $M$  относительная средняя квадратическая погрешность определения значений  $z_a$  по отсчетам в массиве  $M$ :

$$\delta_{z_a}^2 = \delta_M^2 + \delta_x^2. \quad (11.2.9)$$

Отсюда дисперсия распределения значений  $z_a$ :

$$D_{z_a} = (D_M + M^2 \delta_x^2) / \bar{x}^2 = D(M) / \bar{x}^2, \quad (11.2.10)$$

$$D(M) = D_M + M^2 \delta_x^2 = D_M + D_{x_m}, \quad (11.2.11)$$

$$D_M = M+1 \cong M, \quad D_{x_m} = M^2 \delta_x^2,$$

где значение дисперсии  $D_M$  определяется статистикой отсчетов в массиве  $M$  при  $x = \text{const}$ , значение  $D_{x_m}$  представляет собой дисперсию значений  $M$  за счет флуктуаций величины  $x$ , а сумма  $D(M)$  определяет полную дисперсию отсчетов  $M$ .

Влияние  $P(x)$  на форму распределения  $P_M(z)$  сказывается в его "растягивании" по координате  $z$  относительно модального значения, при этом решение интеграла (11.2.7) в первом приближении может быть представлено в следующем виде:

$$P_M(z) \cong b \frac{(bz)^{a-1}}{(a-1)!} e^{-bz}. \quad (11.2.12)$$

Для данного распределения:

$$\bar{z} = z_a = a/b, \quad (11.2.13)$$

$$D_{z_a} = a/b^2, \quad (11.2.14)$$

С учетом выражений (11.2.8) и (11.2.10):

$$a = MD_M / (D_{z_a} \bar{x}^2) = MD_M / D(M), \quad (11.2.15)$$

$$b = D_M / (D_{z_a} \bar{x}) = \bar{x} D_M / D(M). \quad (11.2.16)$$

Значение 'а' в выражении (11.2.15) принимается целочисленным. Выражение (11.2.12) может быть принято для распределения (11.2.4) в качестве априорного распределения вероятностей  $P(z)$ , при этом:

$$P_N(z) = (b+1) \frac{[z(b+1)]^{a+N-1}}{(a+N-1)!} e^{-z(b+1)}. \quad (11.2.17)$$

Отсюда, математическое ожидание и дисперсия  $z$ :

$$z = (N+a)/(b+1), \quad (11.2.18)$$

$$D_z = (N+a)/(b+1)^2. \quad (11.2.19)$$

С использованием выражений (11.2.15-16):

$$z = \beta N + (1-\beta)M / \bar{x}, \quad (11.2.20)$$

где  $\beta$  и  $(1-\beta)$  – весовые коэффициенты доверия отсчетам  $N$  и  $M$ :

$$\beta = D(M)/(D_N \bar{x}^2 + D(M)). \quad (11.2.21)$$

Дисперсия и относительная средняя квадратическая погрешность отсчетов  $z$ :

$$D_z = D(M) \frac{ND(M) + MD_M}{[D_N \bar{x}^2 + D(M)]^2}, \quad (11.2.22)$$

$$\delta_z^2 = 1/(N + MD_M/D(M)). \quad (11.2.23)$$

**Эффективность метода.** Сравнение выражений (11.2.20-23) и (11.2.5-6) позволяет дать оценку эффекта использования дополнительной информации из статистически независимого от  $N$  потока  $M$  (произвольная дополнительная информация).

1. При  $\bar{x} \Rightarrow \text{const}$  имеет место  $\delta_x^2 \Rightarrow 0$ ,  $D_{x_m} \Rightarrow 0$  и дисперсия отсчетов в массиве  $M$  определяется только статистикой потока:

$$D(M) \Rightarrow D_M = M, \quad z = (N+M) / (\bar{x} + 1),$$

$$\delta_z^2 \Rightarrow 1/(N+M) < \delta_N^2 = 1/N, \quad (11.2.24)$$

$$\eta = \delta_N^2 / \delta_z^2 = [N + M^2/D(M)] / N \Rightarrow 1 + M/N,$$

что соответствует определению  $z$  по двум независимым измерениям и эффект использования дополнительной информации максимален. Так, при  $M \approx N$ ,  $\eta \Rightarrow 2$  и погрешность измерений уменьшается в  $\sqrt{\eta} \sim 1.4$  раза.

2. В общем случае  $D_{x_m} \neq 0$ , при этом  $D(M) > D_M$  и положительный эффект снижается. В пределе:  $\delta_x \Rightarrow \infty$ ,  $D_{x_m} \Rightarrow \infty$ ,  $D(M) \Rightarrow \infty$ ,  $\eta \Rightarrow 1$ ,  $z \Rightarrow N$ ,  $\delta_z \Rightarrow \delta_N$  и положительный эффект полностью вырождается. Во всех остальных случаях  $\eta > 1$  и  $\delta_z < \delta_N$ . Отсюда следует, что при наличии коррелированной информации в массиве  $M$  положительный эффект, в той или иной мере, всегда имеет место.

3. Положительный эффект тем больше, чем больше значение  $x = m/n$ , меньше флуктуации  $x$  (величина  $\delta_x$ ), и меньше значения отсчетов  $N = nt$ . Положительный эффект увеличивается именно в тех случаях, когда особенно остро ощущается недостаток информации: при малых значениях плотности потока излучения и/или экспозиции измерений.

Аналогичный эффект будет иметь место и при формировании отсчетов  $M$  по окрестностям текущих точек обработки данных путем определения их среднего значения (низкочастотное сглаживание массива  $n$ ). Предварительное низкочастотное сглаживание может применяться и для статистически независимого дополнительного массива  $m$ , что будет повышать достоверность прогнозных отсчетов и увеличивать глубину регуляризации, если это сглаживание при регуляризации по формулам (11.2.20 и 21) не сказывается на изменении формы основного сигнала. Последнее определяется соотношением частотных спектров основного сигнала и оператора сглаживания.

Возможны два способа реализации уравнения (11.2.20): непосредственно в процессе измерений методом статистической группировки полезной информации (СГПИ) в реальном масштабе времени, или методом статистической регуляризации данных (СРД), зарегистрированных в виде временного (пространственного) распределения в параллельных массивах отсчетов.

### 11.3. СТАТИСТИЧЕСКАЯ РЕГУЛЯРИЗАЦИЯ ДАННЫХ.

Как следует из выражения (11.2.21), для практического использования информации из дополнительных потоков данных необходимо установить значения  $\bar{x}$  и дисперсию  $D(M)$ , причем, исходя из задания последней по выражению (11.2.11), должно быть известно значение  $\delta_x$  - относительной средней квадратической флуктуации величины  $x$ .

Применительно к СРД определение значений  $\bar{x}$  и  $\delta_x$  по зарегистрированным массивам данных не представляет затруднений как в целом по пространству измерений, так и в виде распределений в скользящем окне усреднения данных. Последнее эквивалентно

приведению  $D_{xm} \Rightarrow 0$  для текущей точки обработки данных по информации ее ближайших окрестностей и позволяет производить максимальное извлечение полезной информации из дополнительных потоков сигналов, если частотный спектр распределения величины  $x$  по пространству измерений много меньше частотного спектра полезного сигнала. Отметим, что информация о распределении  $x$  также может иметь практическое значение (в частности, при гамма-опробовании с дополнительным потоком сигналов в низкоэнергетическом диапазоне спектра излучения - для оценки эффективного атомного номера горных пород).

**Проверка теоретических положений метода** АРД проводилась путем статистического моделирования соответствующих массивов данных и их обработки цифровыми фильтрами.

В таблице 1 приведены 4 группы результатов обработки по формулам (11.2.20-21) двух статистически независимых и постоянных по средним значениям массивов данных  $n$  и  $m$  (модели постоянных полей) при различных установках СРД по скользящему окну  $K_c$  счета текущих значений  $\bar{x}_i = m_i/n_i$  и  $D_i(M)$  по массиву  $m$ . Текущая точка обработки данных – по центру окна. Количество отсчетов в каждом массиве – 1000, распределение значений отсчетов соответствует закону Пуассона. Определение прогнозных отсчетов  $M_i$  по массиву  $m$  для использования в уравнении (11.2.20) проводилось со сглаживанием отсчетов в скользящем окне  $K_s$  низкочастотного цифрового фильтра (вариант без сглаживания при  $K_s = 1$ ). В качестве низкочастотного фильтра в алгоритме СРД используется (здесь и в дальнейшем) весовое окно Лапласа-Гаусса. Теоретическое значение  $D_{z.t.}$  дисперсии результатов  $z$  определялось по выражению (11.2.22) с расчетом дисперсии  $D(M)$  по выражению  $D(M) = \bar{M} [1 + \bar{M} (1/(K_c \bar{N}) + 1/(K_c \bar{M}))]$ . При сглаживании прогнозных отсчетов значение  $D_M$  в выражении (11.2.22) принималось равным  $D_M = \bar{M} \cdot H_s$ , где  $H_s$  – коэффициент усиления сглаживающим фильтром дисперсии шумов (сумма квадратов коэффициентов цифрового фильтра). Дополнительно в таблице приводятся зарегистрированные средние значения коэффициента снижения статистических флюктуаций  $\eta = \delta_n^2 / \delta_z^2$ .

Таблица 1. Статистика результатов моделирования СРД.

(Основной массив  $\bar{n} = 9.9$ ,  $D_n = 9.7$ , дополнительный массив  $\bar{m} = 9.9$ ,  $D_m = 9.9$ , 1000 отсчетов.)

$K_c$	$K_s$	$z$	$D_z$	$D_{z.t.}$	$\eta$	$K_c$	$K_s$	$z$	$D_z$	$D_{z.t.}$	$\eta$
3	1	9,7	5,7	6,19	1,7	11	3	9,6	3,6	3,80	2,8
5	1	9,7	5,4	5,78	1,8	11	5	9,6	3,3	3,55	3,0
11	1	9,6	5,1	5,36	1,9	11	11	9,6	3,1	3,22	3,2
21	1	9,6	5,0	5,18	2,0	11	21	9,6	3,0	3,11	3,3
51	1	9,6	5,0	5,05	2,0	11	51	9,6	3,0	2,99	3,3
3	3	9,7	4,1	4,71	2,4	3	11	9,8	4,5	4,26	2,2
5	5	9,7	3,6	4,01	2,8	5	11	9,7	3,5	3,78	2,8
11	11	9,6	3,1	3,22	3,2	11	11	9,6	3,1	3,22	3,2
21	21	9,6	2,9	2,91	3,4	21	11	9,6	3,1	3,12	3,2
51	51	9,6	2,7	2,66	3,7	51	11	9,6	3,1	2,99	3,2

Как видно из данных таблицы, практические результаты фильтрации достаточно хорошо совпадают с ожидаемыми по данным теоретических расчетов. Некоторое уменьшение среднего значения  $z$  по отношению к исходному среднему значению  $n$  определяется асимметричностью пуассоновского типа модели. При малых средних значениях модельных отсчетов в массиве  $m$  это приводит к определенной статистической асимметрии в работе СРД, т.к. при  $(+\sigma_m)^2 > (-\sigma_m)^2$  среднестатистическое доверие к дополнительной информации с отсчетами  $M_i + \sigma$  меньше, чем с отсчетами  $M_i - \sigma$ . Этим же

фактором, по-видимому, вызвано и большее расхождение между теоретическими и фактическими значениями  $D_z$  при малых значениях окна  $K_c$ . Можно также заметить, что по значению коэффициента  $\eta$  фильтрация выходит на теоретические значения ( $\Rightarrow 1+M/N$ ) только при достаточно точном определении значений  $\bar{x}_i$  и  $D_i(M)$ , что требует увеличения окна  $K_c$  счета этих параметров для полного использования дополнительной информации.

Таблица 2.

$n/D_n$	$m/D_m$	Окно $K_c$	Окно $K_s$	Результаты МАРД			
				$z$	$D_z$	$z/n$	$D_n/D_z$
9,80   10,12	9,76   9,19	5	1	9,62	5,63	0,981	1,80
			5	9,71	2,61	0,990	3,88
			15	9,78	2,05	0,997	4,93
			55	9,80	2,01	1,0	5,03
			205	9,80	2,01	1,0	5,03
		15	1	9,56	5,23	0,975	1,94
			5	9,56	3,47	0,985	4,35
			15	9,56	3,12	0,992	8,36
			55	9,56	2,98	0,999	14,2
			205	9,57	2,95	1,0	14,8

Эффект использования дополнительной информации, в полном соответствии с выражением (11.2.22), усиливается при предварительном сглаживании статистических вариаций отсчетов  $M_i$  и при увеличении значений отсчетов дополнительного массива (материалы по последнему случаю не приводятся, т.к. не имеют какой-либо дополнительной информации). В спокойных по динамике полях еще большая глубина регуляризации может быть достигнута при счете значений  $\bar{x}_i$

и  $D_m$  по сглаженному массиву  $M$ , что позволяет повысить вес прогнозных отсчетов  $M_i$ . Результаты моделирования данного варианта в тех же условиях, что и для таблицы 1, приведены в таблице 2. Такой же эффект, в принципе, может достигаться и непосредственным введением дополнительного коэффициента веса в выражение (11.2.20) в качестве множителя для значения  $D(M)$ , что позволяет осуществлять внешнее управление глубиной регуляризации.

**Оценка сохранения разрешающей способности** полезной информации была проведена на фильтрации детерминированных сигналов  $n$  и  $m$  предельной формы – в виде прямоугольных импульсов. Оценивались два фактора: сохранение формы полезного сигнала и подавление статистических шумов, наложенных на полезный сигнал.

При установке СРД без усреднения данных по массиву  $M$  ( $K_s = 1$ , прогноз  $M_i$  по текущим значениям массива  $M$ ) при любых значениях окна  $K_c$  выходной массив  $Z$  без всяких изменений повторяет массив  $N$ , т.е. не изменяет полезный сигнал и полностью сохраняет его частотные характеристики. Естественно, при условии, что массив  $M$  пропорционален массиву  $N$ .

При  $K_s > 1$  форма выходных кривых несколько изменяется и приведена на рис. 11.3.1. В индексах выходных кривых  $z$  приведена информация по установкам окон СРД: первая цифра - окно счета дисперсии  $D_m$  и текущего значения  $\bar{x}_i$  (в количестве точек отсчетов), вторая цифра (через флеш) - окно сглаживания отсчетов  $M$  весовой функцией Лапласа-Гаусса и определения прогнозных отсчетов  $M_i$ . Для сравнения с результатами типовой низкочастотной фильтрации на рисунке приведена кривая  $n_{25}$  отсчетов  $N$ , сглаженных весовой функцией Лапласа-Гаусса с окном 25 точек.

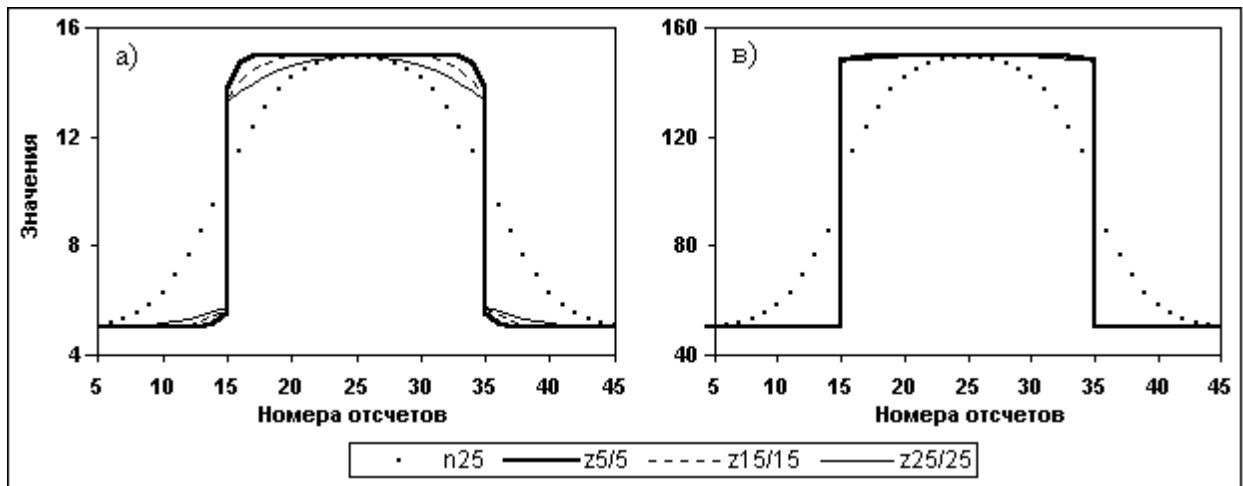


Рис. 11.3.1. СРД прямоугольного импульса. Счет  $D_m$  по несглаженному массиву  $M$ .

На рис. 11.3.1а приведен результат СРД прямоугольного импульса с амплитудным значением 10 на фоне 5 при отношении  $m/n = 1$  (равные значения отсчетов  $N$  и  $M$ ). Дисперсия  $D_N$  в выражении (11.2.21) принималась равной значению отсчетов  $N$  (статистика Пуассона). Как видно на рисунке, при сохранении фронтов сигнальной функции сглаживание прогнозных значений  $M_i$  приводит к появлению искажения формы сигнала по обеим сторонам скачка, интервал которого тем больше, чем больше значение  $K_s$ . Амплитудное значение искажений, как это и следует из выражения (11.2.21), в первую очередь зависит от соотношения текущих значений  $D_N$  и  $D(M)$  и в меньшей степени от глубины сглаживания прогнозных отсчетов.

Максимальную величину искажения для точек скачка в первом приближении можно оценить из следующих соображений. Значения  $D(M)$  между точками скачка равны  $D(M) = A^2/4$ , где  $A$  - амплитуда скачка, при этом значения коэффициента  $\beta$  для нижней и верхней точек скачка определяются выражениями  $\beta \approx A^2/(4D_N + A^2)$ , где  $D_N = N$  точки скачка (для статистики Пуассона). Отсюда, при прогнозном значении  $M \approx N + A/2$  для нижней точки скачка и  $M \approx N - A/2$  для верхней точки относительная величина изменений  $N$  определится выражением  $\delta \approx 1/(2N/A + A)$ , т.е. будет тем меньше, чем больше значения  $A$  и  $N$  и больше отношение  $N/A$ , что можно наглядно видеть на рис. 11.3.1в. Из этого выражения также следует, что максимальные искажения скачков, вносимые системой СРД, будут всегда в несколько раз меньше, чем статистические флуктуации непосредственных отсчетов  $\delta = 1/\sqrt{N}$  на краях скачков.

При увеличении глубины регуляризации введением счета дисперсии  $D(M)$  по сглаженному массиву  $M$  картина искажений несколько изменяется и приведена на рис. 11.3.2. Реакция СРД на сглаживание дисперсии  $D(M)$  проявляется в своеобразной компенсации абсолютных отклонений отсчетов непосредственно по сторонам скачка отклонениями противоположного знака в более дальней зоне от скачка. Максимальные значения искажений остаются примерно на таком же уровне, как и для работы по несглаженной дисперсии  $D(M)$ , с несколько меньшей зависимостью от увеличения значений  $N$  и  $A$ .

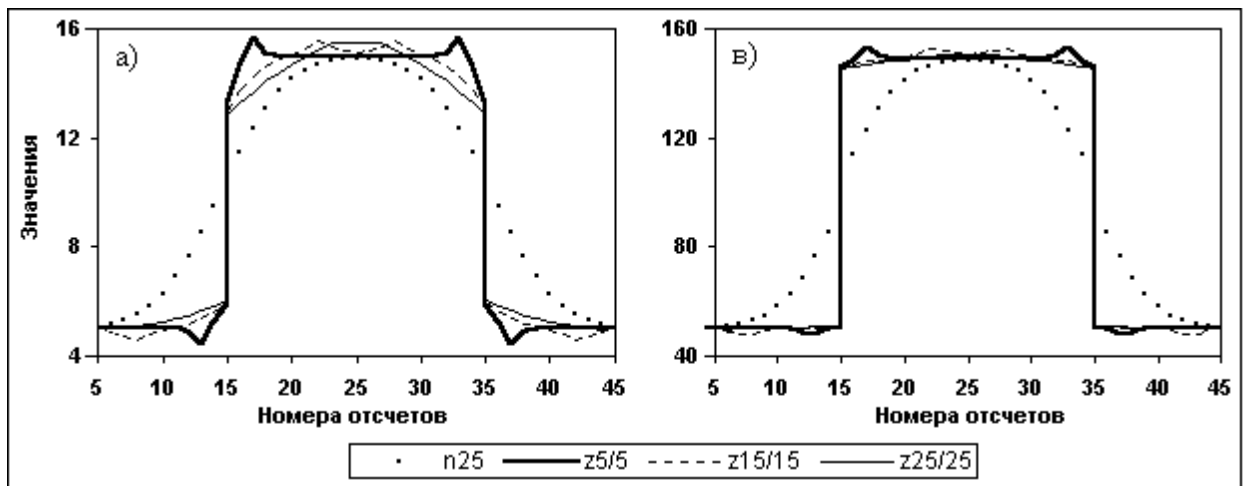


Рис. 11.3.2. СРД прямоугольного импульса. Счет  $D_m$  по сглаженному массиву  $M$ .

В приведенных примерах значение окна счета  $K_c$  принималось равным значению окна сглаживания  $K_s$  дополнительного массива  $M$ . При  $K_c > K_s$  картина процесса практически не изменяется. При обратном соотношении размеров окон вступает в действие второй фактор - отклонение от фактических значений счета текущих значений  $x_i = m/n$  в малом окне  $K_c$  по массиву отсчетов, сглаженных с большим окном  $K_s$ . На расстояниях от скачка функции, больших  $K_c/2$ , СРД переходит в режим предпочтения сглаженных значений массива  $M$ , т.к.  $D(M) \Rightarrow 0$ , что при  $K_c < K_s$  может приводить к появлению существенной погрешности – выбросов на расстояниях  $\pm K_c/2$  от скачков. Естественно, что при практических измерениях таких условий наблюдаться не будет и эффект резко уменьшится, но для полного его исключения вариант  $K_c \geq K_s$  можно считать предпочтительным.

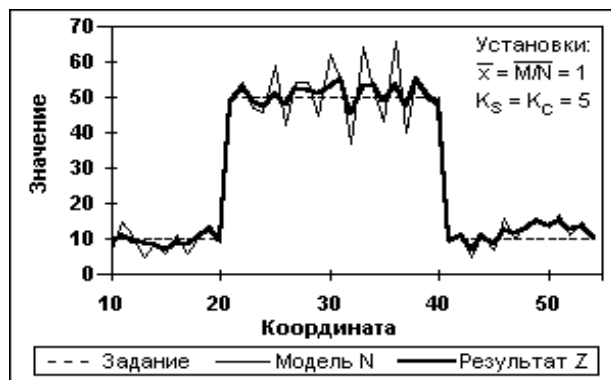


Рис. 11.3.3. СРД сигнала  $N$  по массиву  $M$ .  
(Счет  $D_m$  по несглаженному массиву  $M$ ).

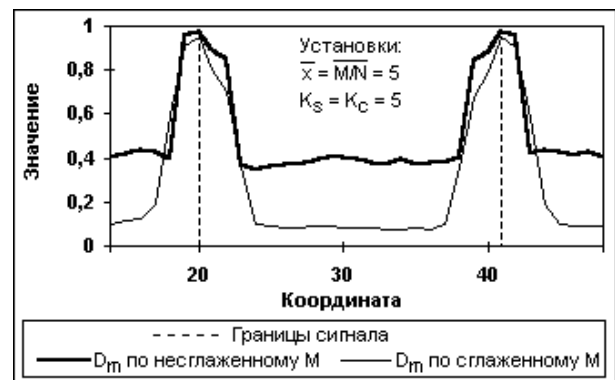


Рис. 11.3.4. Коэффициент  $\beta$ .  
(Среднее статистическое по 50 циклам)

На рис. 11.3.3 приведен пример регистрации рандомизированного модельного сигнала в виде прямоугольного импульса амплитудой 40 на фоне 10, на котором виден принцип работы СРД. Как и следовало ожидать, СРД производит сглаживание статистических флуктуаций фона и сигнала за пределами зоны  $\pm K_c$  от скачка, отдавая предпочтение сглаженным прогнозным значениям  $M_i$ , и не изменяет значения фона и сигнала в пределах этой зоны в связи с резким возрастанием текущих значений  $D(M)$  в выражении (11.3.21). Изменение коэффициента  $\beta$  в зоне скачка, управляющего формированием выходных отсчетов, приведено на рис. 11.3.4 (среднестатистическое по 50-ти циклам рандомизации для модельного импульса на рис. 11.3.3) и наглядно показывает принцип адаптации СРД к динамике изменения значений обрабатываемых сигналов.

**Статистическая оценка регуляризации данных** по прямоугольным импульсам проводилась по 50-ти циклам рандомизации исходных массивов  $N$  и  $M$ . В качестве примера на рисунках 11.3.5 и 6 приведены результаты обработки статистики массивов  $N$  и  $Z$ . Кроме статистики циклов рандомизации проводилась суммарная обработка всех циклов по общей статистике фона и вершины импульсов. Результаты обработки для тех же установок фильтров приведены в таблице 3.

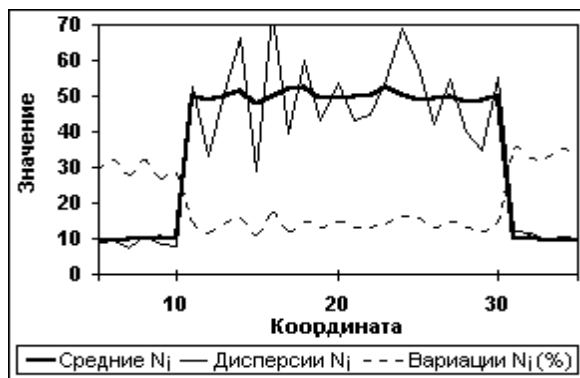


Рис. 11.3.5. Статистика сигнала  $N$  (Измерения по 50-ти циклам).

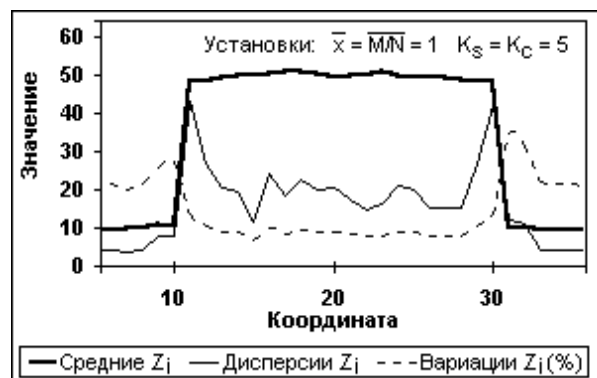


Рис. 11.3.6. Статистика сигнала  $Z$  (50 циклов. Счет  $D_m$  по несглаженному  $M$ )

Таблица 3.

Статистика значений фона и вершины импульсов (50 циклов).

Массивы и условия обработки	Фон		Сигнал	
	Сред. отсчет	Дисперсия	Сред. отсчет	Дисперсия
Основной входной массив $N$	9,96	9,97	50,1	52,0
Дополнительный входной массив $M$	9,89	9,49	50,2	47,4
Массив $Z$ , счет $D_m$ по несглаженному $M$	9,87	5,47	49,7	22,3
Массив $Z$ , счет $D_m$ по сглаженному $M$	9,84	4,76	49,9	18,6
Массив $N$ , сглаженный весовым окном	11,5	17,9	48,5	29,2

**Результаты моделирования** подтверждают преимущество СРД перед простыми методами сглаживания. В числовой форме это наглядно проявляется в снижении дисперсии отсчетов выходного массива  $Z$  при практическом сохранении средних значений массива  $N$  и для фоновых отсчетов, и для амплитудных значений сигнала. При простом сглаживании "развал" фронтов сигнала (подавление высокочастотных составляющих спектра сигнала), как и должно быть при использовании низкочастотных фильтров, вызывает снижение по отношению к исходному массиву средних значений в максимумах и повышение фоновых значений сигнала, которое тем больше, чем больше окно весовой функции. Этот эффект особенно отчетливо проявляется в интервале окна фильтра по обе стороны от резких изменений сигнала.

При отсутствии дополнительных массивов  $M$ , коррелированных с регуляризуемым массивом  $N$ , формирование прогнозных значений  $M_i$  может производиться по ближайшим окрестностям текущих значений  $N_i$  в скользящем окне  $K_s$ . При строго корректном подходе текущая точка  $N_i$  не должна включаться в число счета прогнозных значений  $M_i$ , но, как показало моделирование, это практически не влияет на результаты регуляризации. При прогнозировании  $M_i$  по всем точкам окна  $K_s$  массив  $M$  формируется любым методом сглаживания из массива  $N$ , и все особенности работы СРД по сглаженным массивам  $M$ , рассмотренные выше, остаются без изменений при условии счета значений  $D_m$  в окне  $K_c$  по массиву  $M$ . Для исключения выбросов по обе стороны от скачков полезного сигнала счет  $D_m$  как дисперсии прогнозных значений  $M_i$  необходимо выполнять непосредственно по массиву  $N$ .

Фундаментальной особенностью СРД является возможность последовательной многократной фильтрации данных, при которой может осуществляться преимущественное повышение степени регуляризации данных с минимальными искажениями формы полезного сигнала. Для выполнения последнего размер окна  $K_c$  счета  $x_i$  и  $D_m$  устанавливается минимальным (3-5 точек), а глубина регуляризации данных (степень подавления шумов) устанавливается количеством последовательных операций фильтрации (до 3-5 проходов). Пример регуляризации модельного массива  $N$  в три прохода приведен на рис. 11.3.7.

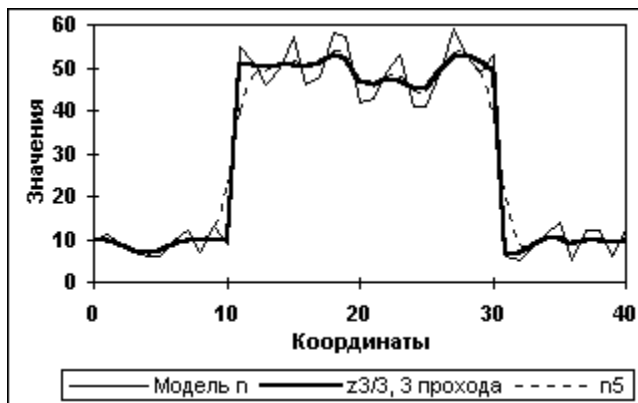


Рис. 11.3.7. СРД одиночного массива  $N$  (3 прохода. Счет  $D_m$  по массиву  $n$ )

Для сравнения пунктиром на рисунке приведено сглаживание массива 5-ти точечным фильтром Лапласа-Гаусса, который имеет коэффициент подавления шумов, эквивалентный 3-х проходному СРД (см. рис. 11.3.9).

На рисунках 11.3.8 и 11.3.9 приведены результаты статистической обработки 3-х проходной СРД для 25 циклов моделирования в сравнении с 1-м проходом и с 5-ти точечным фильтром Лапласа-Гаусса (кривая  $n5$ ).

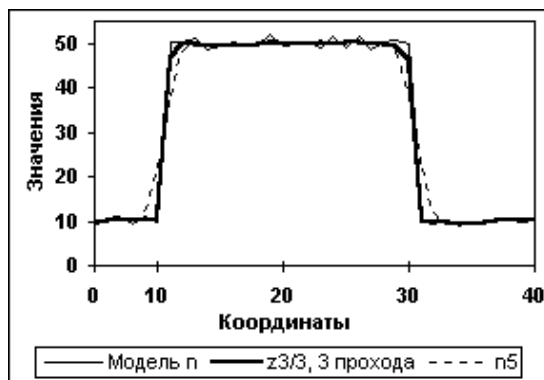


Рис. 11.3.8. Статистика средних значений (25 циклов. Счет  $D_m$  по массиву  $n$ )

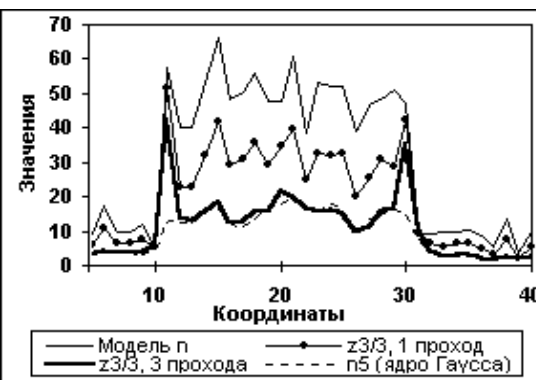


Рис. 11.3.9. Статистика дисперсий (25 циклов. Счет  $D_m$  по массиву  $n$ )

Количество проходов может ограничиваться в автоматическом режиме, например, по среднеквадратическому значению корректирующих отсчетов  $\Delta z_i = N_i - z_i$  в каждом проходе по сравнению с предыдущим проходом, которое сначала резко уменьшается за счет сглаживания флуктуаций, а затем, в зависимости от динамики сигнальной функции, стабилизируется или даже начинает увеличиваться за счет искажения самого сигнала.

**Частотное представление** работы СРД хорошо видно на рис. 11.3.10, где приведены модули спектров рандомизированного сигнала в виде меандра (средние значения в минимуме - 20, в максимуме - 100, 25 периодов по 40 отсчетов, всего 1000 отсчетов) и результатов его обработки СРД (окно  $K_c=3$ , окно  $K_s=3$ ).



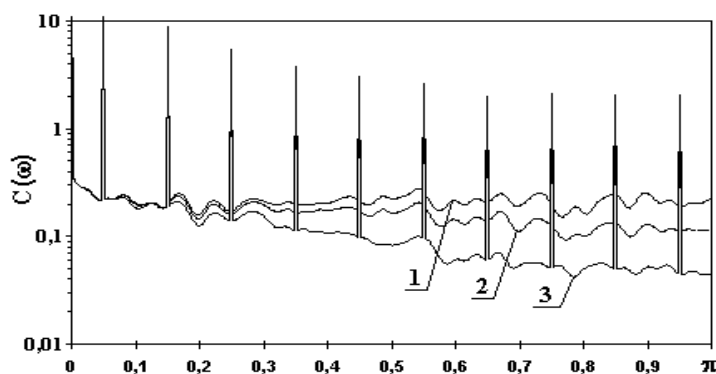


Рис. 11.3.10. Модули спектров модельных сигналов.

(1– входной массив N, 2– выходной массив Z, один цикл СРД,

3– выходной массив Z, три цикла СРД), 4 – массив нерандомизированного меандра).

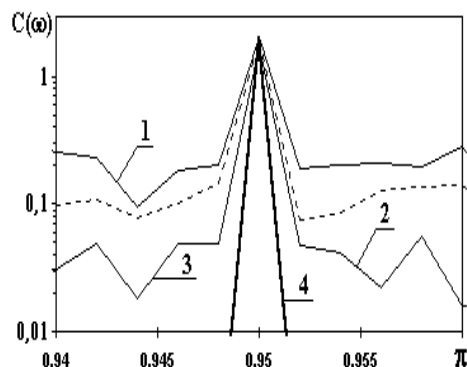


Рис.11.3.11. Участок спектра.

Модуль спектра основного полезного сигнала (в данном случае чистого меандра) представляет собой последовательность отдельных частотных гармоник по всему диапазону спектра. В спектре рандомизированного меандра эти частотные гармоники суммируются со спектром шума, статистически равномерно распределенным по всему частотному диапазону (спектр шума на рисунке для наглядности сглажен). СРД осуществляет подавление шумовых составляющих сигнала, практически не затрагивая частотных гармоник меандра и не изменяя их по амплитуде. Последнее можно видеть на рис. 11.3.11, где представлен отрезок спектра сигналов в высокочастотной части главного диапазона в области одной гармоники меандра (частотные составляющие шума не сглажены). При 3-х цикловом СРД высокочастотные составляющие шумов подавляются практически на порядок.

**Пример практического использования** СРД приведен на рис. 11.3.12 при опробовании участка скважины, пересекающей пласты каменной соли, на содержание сильвинита по гамма-излучению Калия-40. По данным геологического опробования пласты сильвинита в толще вмещающих пород (галита) имеют достаточно резкие границы и однородны по содержанию сильвинита в пределах пластов. Исходная диаграмма ГК (детектор CsJ(Tl) со свинцовым фильтром толщиной 2 мм) и результаты фильтрации исходного массива данных ГК с использованием СРД и низкочастотного фильтра с весовым окном Лапласа-Гаусса приведены на рис. 11.3.12.

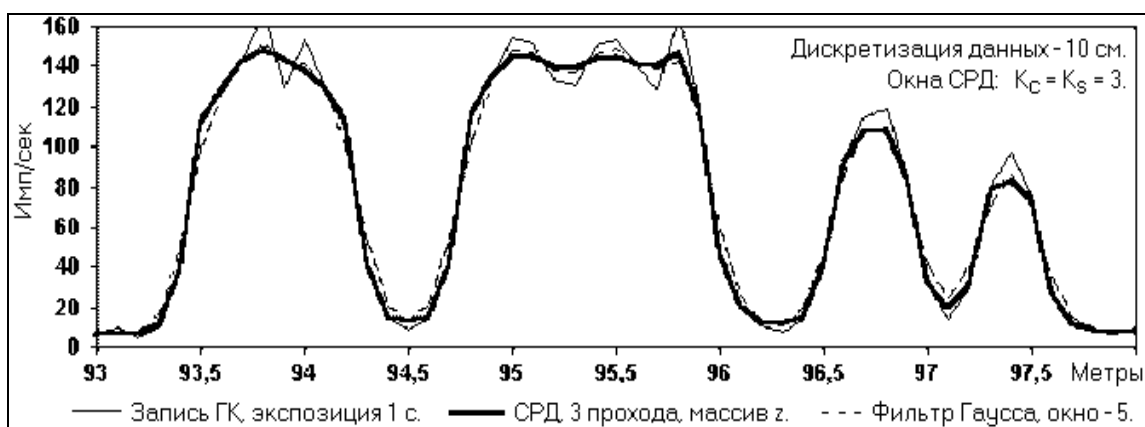


Рис. 11.3.12. Диаграммы ГК.

Результаты интерпретации диаграмм ГК симметричным деконволюционным цифровым фильтром (окно 13 точек) приведены на рис. 11.3.13. Как видно на рисунке, деконволюция по несглаженной диаграмме ГК дает существенные вариации содержания сильвинита в пределах пластов. Применение низкочастотной фильтрации диаграммы ГК снимает флюктуации содержания в пределах пластов, но существенно сглаживает границы

пластов. Использование СРД позволяет устранить этот недостаток.

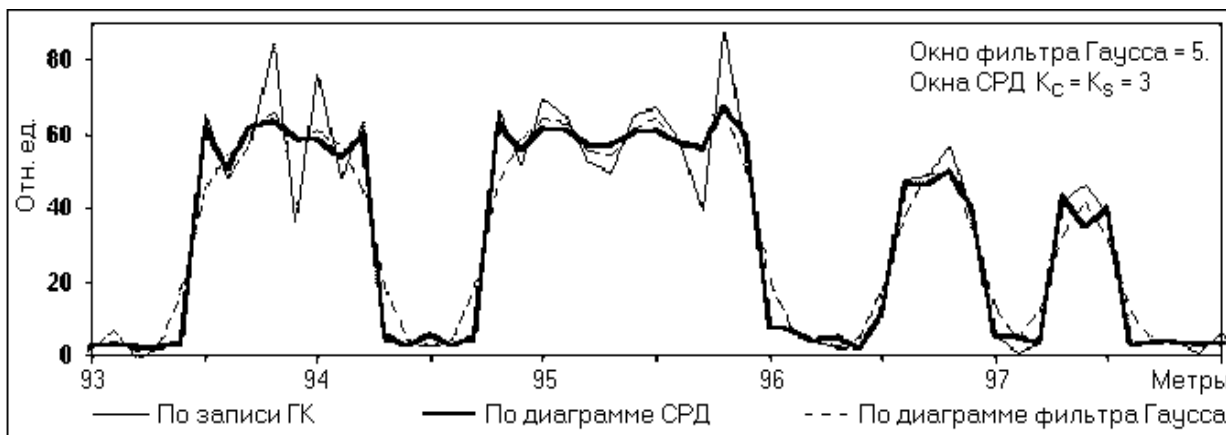


Рис. 11.3.13. Результаты интерпретации диаграмм ГК.

В заключение отметим, что СРД может использоваться для регуляризации не только ядернофизических данных, но и любых других числовых массивов непрерывных измерений, если радиус их корреляции не менее 3-5 отсчетов. В качестве примера на рис. 11.3.14 приведена диаграмма акустического каротажа, зарегистрированная с шагом дискретизации данных 20 см, сглаживание которой проведено СРД без потери пространственного разрешения.

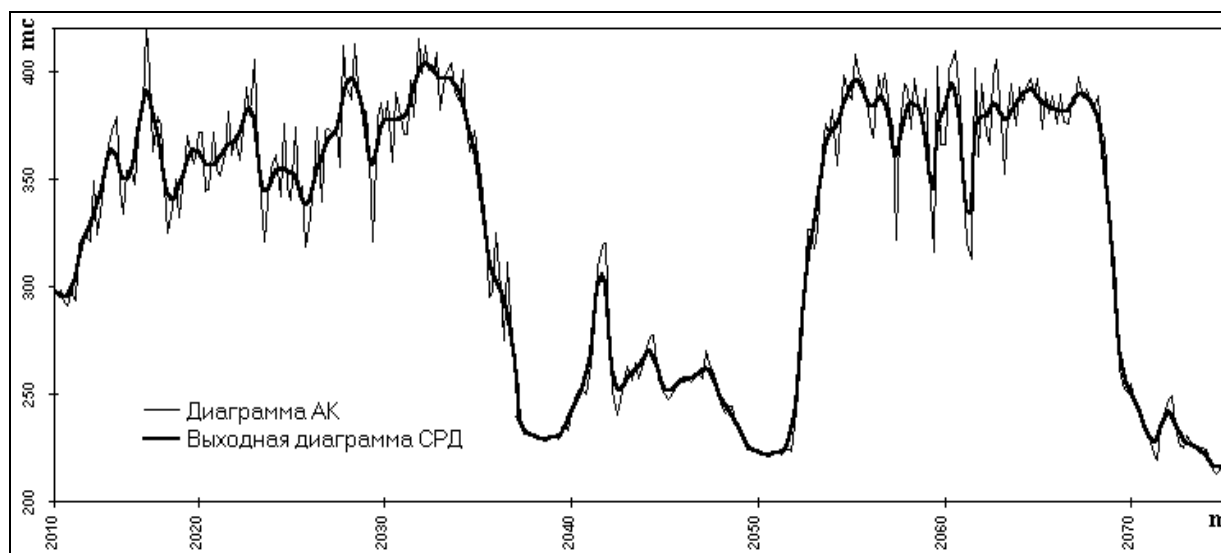


Рис. 11.3.14. Диаграмма акустического каротажа и результат ее обработки СРД (5 циклов,  $K_c = K_s = 3$ , физическое окно 0.6 м).

### 11.3. СТАТИСТИЧЕСКАЯ ГРУППИРОВКА ПОЛЕЗНОЙ ИНФОРМАЦИИ.

Что касается аппаратных способов реализации СГПИ, то он может быть выполнен в реальном масштабе времени, если информация представлена потоком импульсов и основным информативным параметром является скорость следования импульсов.

**Сущность аппаратной реализации** заключается в статистической (близкой к статистической) нормированной выборке импульсов из дополнительного потока  $m$  и их суммировании с основным потоком  $n$  с заданием условий выборки по отношению частоты следования импульсов в потоках. Полагая для непрерывного режима измерений  $M+1 = M$ , перепишем выражение (5.2.20) с подстановкой значения  $\beta$  в следующем виде:

$$z = N + (M/\bar{x} - N) \cdot M\bar{x} / (M\bar{x} + D(M)). \quad (11.3.1)$$

Умножим левую и правую части выражения на нормировочный коэффициент размножения выходного потока  $K = 1+R$ :

$$Z = K \cdot z = N + RN + (M/\bar{x} - N) \cdot KM\bar{x} / (M\bar{x} + D(M)). \quad (11.3.2)$$

Заменим отсчеты RN выборкой сигналов из потока m:

$$RN = P_B M, \quad (11.3.3)$$

где  $P_B$  - вероятность выборки сигналов из потока m. Если вероятность выборки сигналов поддерживать равной значению

$$P_B = R/\bar{X}, \quad (11.3.4)$$

то при этом будет иметь место

$$M/\bar{X} - N = P_B M/R - N \rightarrow 0, \quad (11.3.5)$$

и соответственно для выражения (11.3.2) имеем:

$$(M/\bar{X} - N) \cdot KM\bar{X}/(M\bar{X} + D(M)) \rightarrow 0, \quad (11.3.6)$$

$$Z = N + P_B M \rightarrow N + RN. \quad (11.3.7)$$

При статистической независимости величины x от частоты потоков n и m приведенные выражения действительны при определении значения  $\bar{X}$  как в целом по пространству измерений, так и для скользящих окон текущих значений  $\bar{X}$  по определенным интервалам предшествующих измерений. Действительно и обратное заключение: если по определенному интервалу измерений выражение (11.3.5) обращается в нуль, то установленная вероятность выборки соответствует условию (11.3.4). На этом принципе может проводиться аппаратная реализация СГПИ с автоматической адаптацией к условиям измерений: управление процессом выборки импульсов из потока m и направление их на суммирование с потоком n по сигналам обратной связи с устройства, следящего за обращением в нуль выражения (11.3.5).

**Особенности аппаратной реализации** СГПИ с автоматической адаптацией под условия измерений заключаются в следующем.

Значение вероятности выборки  $P_B$  не может быть больше 1. Отсюда из (11.3.3) следует, что для любых интервалов измерений должно выполняться условие  $M \geq RN$ , а соответственно по всему пространству измерений должно выполняться условие  $\bar{X} \geq R$ , чем и обуславливается выбор коэффициента R. Значение коэффициента R принципиально ограничивает степень положительного эффекта СГПИ ( $k_{\max} \rightarrow 1+R$ ), в отличие от СРД, где такого ограничения не имеется.

Относительная статистическая погрешность измерений выходного потока отсчетов Z соответствует выражению (11.2.23) при условии постоянного значения величины  $P_B$ , т.е. при установке значения  $P_B$  по среднему значению величины  $\bar{X}$  в целом по пространству измерений. При автоматической адаптации под условия измерений значение вероятности  $P_B$  по текущему среднему значению отношения n/m определенного предшествующего интервала измерений также является статистически флюктуирующей величиной с дисперсией распределения (без учета изменений действительного значения x):

$$D_p = R^2(n+m)n/(m^3T), \quad (11.3.8)$$

где T- интервал усреднения информации при определении текущего значения  $\bar{X}$ . Соответственно, дисперсия и средняя квадратическая погрешность текущих отсчетов Z:

$$D_z = D_N + P_B D_M + M^2 D_p = N + P_B M + M^2 D_p, \quad (11.3.9)$$

$$\delta_z^2 = (N + P_B M + M^2 D_p)/(N + P_B M)^2. \quad (11.3.10)$$

При постоянной экспозиции измерений  $\tau$  положительный эффект возрастает с увеличением значения T:

$$k = K^2/(K + R^2(n+m)\tau/mT). \quad (11.3.11)$$

$$k_{\max} \rightarrow 1+R, \quad \delta_z^2 \rightarrow 1/(N + P_B M) \text{ при } T \rightarrow \infty. \quad (11.3.12)$$

В общем случае, с учетом средней квадратической ошибки прогнозирования  $\delta_{xi}$  значений  $x_i$  для текущих точек измерений по значениям  $\bar{X}$  в предшествующих интервалах при  $T > \tau$ :

$$D_z = N + P_B M + M^2(D_p + P_B^2 \delta_{xi}^2). \quad (11.3.13)$$

Формирование значения  $P_B$  на основе информации по средним значениям  $\bar{X}$  интервалов измерений, предшествующих текущим, определяет СГПИ как динамическую систему с соответствующей постоянной времени реакции на изменение условий измерений. Учитывая, что, во-первых, для любой точки пространства измерений должно выполняться условие  $m > nR$ , и, во-вторых, увеличение интервала  $T$  приводит к возрастанию времени реакции на изменение условий измерений, значение  $T$  целесообразно ограничивать величиной порядка (5-10) значений текущих экспозиций. Чем меньше пространственная частота распределения  $x$  по отношению к распределению  $n$ , тем большее значение  $T$  допустимо.

**Реализация систем СГПИ** значительно облегчается при чисто практическом ограничении целевой задачи: получение максимального положительного эффекта в экстремально неблагоприятных условиях производства измерений (при низких значениях регистрируемой плотности потока излучения, при высокой скорости измерений) с вырождением положительного эффекта по мере снижения статистической погрешности измерений в основном потоке. Так, например, если при скважинном гамма-опробовании статистическая погрешность измерений основного потока сигналов в зонах с повышенной интенсивностью излучения снижается до 2-3%, то ее дальнейшее уменьшение не имеет практического смысла, т.к. основная погрешность каротажной радиометрической аппаратуры обычно не превышает 5%.

Использование данного целевого ограничения позволяет применить формирование параметра  $P_B$  не в скользящем окне временного или пространственного усреднения информации, а по определенному зарегистрированному объему предшествующей информации, т.е. с автоматической вариацией интервала усреднения информации и постоянной регулирования  $P_B$  в зависимости от частоты потоков сигналов, при этом объем информации формирования  $P_B$  может задаваться с учетом характера вариаций величины  $\bar{X}$  и допустимого значения динамической погрешности измерений.

Для реализации такой возможности преобразуем выражение (11.3.5) по интервалу усреднения  $t$  к виду:

$$P_B m t / R - n t + Q = q, \quad (11.3.14)$$

$$P_B = nR/m = q/\mu, \quad (11.3.15)$$

$$q \rightarrow Q \text{ при } t \rightarrow \infty,$$

где  $Q$  - средний уровень смещения числового эквивалента сигнала обратной связи системы АРВ - автоматического регулирования вероятности выборки  $P_B$ , при котором обеспечивается выполнение равенства (11.3.15),  $\mu$  - коэффициент пропорциональности преобразования цифрового сигнала АРВ в сигнал  $P_B$ . Дифференциальное уравнение для системы АРВ:

$$dq/dt = n - m q / \mu R. \quad (11.3.16)$$

Решение дифференциального уравнения при начальных условиях  $t = 0$  и  $q = 0$  (переходная функция АРВ):

$$q = \mu R (n/m) [1 - \exp(-mt/\mu R)]. \quad (11.3.17)$$

$$P_B = R(n/m) [1 - \exp(-mt/\mu R)] = R(n/m) [1 - \exp(-nt/q)]. \quad (11.3.18)$$

Как видно из этих выражений, значение сигнала обратной связи АРВ пропорционально отношению  $(n/m)$  частот потоков, а постоянная времени АРВ  $\mu R/m$  прямо пропорциональна значению коэффициента преобразования  $\mu$  при обратной пропорциональности от значения частоты дополнительного потока  $m$ , равно как и, с учетом (11.3.15), прямо пропорциональна текущему значению сигнала обратной связи  $q$  при обратной пропорциональности от значения частоты основного потока  $n$ . Первое полностью эквивалентно второму при  $(n/m) \rightarrow \text{const}$  и  $q = \mu R n/m \rightarrow Q$ . В первом приближении, с использованием выражения (11.3.8) и эквивалентности значения

статистических флюктуаций при  $T \approx 2\tau$  для скользящего прямоугольного временного окна и окна интенсивметра с экспоненциальной переходной функцией, для относительных флюктуации значения  $P_B$  получаем:

$$\delta_p^2 = (n+m)/(2\mu Rn) = (n+m)/(2qm). \quad (11.3.19)$$

Выражение действительно для прямого измерения  $2\tau$ -интенсиметром отношения  $(n/m)$  и является максимальной оценкой. Для более точной оценки следует учитывать, что в данном случае интенсиметр является устройством с отрицательной обратной связью по цепи АРВ, что несколько уменьшает значение флюктуации. Точная оценка может быть проведена с использованием формулы Кэмпбелла для дисперсии случайной величины  $x(t)$ , образованной сложением импульсов пуассоновского потока [5], отдельно для потока  $n$  при  $m = \text{const}$  и потока  $m$  при  $n = \text{const}$ , с последующим сложением квадратов относительного среднего квадратического значения флюктуации. Так, для схемы, приведенной ниже, получено значение  $\delta_p^2 \approx (R+1)m/(2\mu nR^2)$ .

При выбранном для пространства измерений значении коэффициента  $R \leq (m/n)_{\min}$  с использованием выражения (11.3.19) параметры системы АРВ (коэффициент  $\mu$  и среднее значение  $Q$  для средней по пространству величины отношения  $n/m$ ) могут устанавливаться под заданное значение допустимых флюктуаций вероятности выборки импульсов  $P_B$ :

$$\mu \leq (1+(m/n)_{\max})/(2R\delta_p^2). \quad (11.3.20)$$

В процессе измерений АРВ осуществляет непрерывную адаптацию под текущие условия измерений ( $n\tau \rightarrow q$ ,  $m\tau \rightarrow mR$ ,  $P_B \rightarrow q/\mu$ ) с регулированием текущего значения  $P_B$  по объему информации  $q = (n/m)\mu R = n\tau$  предшествующего интервала измерений путем соответствующего изменения постоянной времени интегрирования этой информации в зависимости от изменения частот потоков сигналов. При  $n/m \rightarrow \text{const}$  последнее имеет абсолютный характер:  $\delta_p \rightarrow \text{const}$ ,  $\tau \rightarrow (1/n + 1/m)/(2\delta_p^2)$ .

Следует отметить, что во многих методах геофизики существуют достаточно благоприятные условия использования как СГПИ, так и СРД. Так, например, применительно к скважинному гамма-опробованию с извлечением дополнительной информации из низкоэнергетической части спектра излучения условия достаточно точной реакции на изменения параметра  $\bar{X}$  по стволу скважины являются весьма хорошими, т.к. основной фактор вариации значений  $x$  - эффективный атомный номер среды, изменяется в небольшом диапазоне с низкой пространственной частотой вариаций, причем в зонах расположения активных пород, где требуется наиболее высокая точность интерпретации результатов измерений и возможны значительные изменения атомного номера пород, за счет увеличения плотностей потоков излучения постоянная времени АРВ будет существенно уменьшаться, а пространственная разрешающая способность измерений соответственно увеличиваться. Аналогичные условия характерны, как правило, и для других методов ядерной геофизики.

**Пример исполнения системы СГПИ** для двух импульсных потоков сигналов приведен на рис. 11.3.1. Функциональная схема СГПИ содержит реверсивный счетчик импульсов 1, на вход суммирования которого подаются импульсы основного потока  $n$ , а на вход вычитания - импульсы дополнительного потока  $m$ , предварительно проходящие через схему выборки импульсов 3 и счетчик-делитель частоты следования импульсов 4 с коэффициентом пересчета  $R$ .



Рис. 11.3.1. Базовая функциональная схема СПИ.  
 1- реверсивный счетчик импульсов, 2- блок формирования сигнала выборки импульсов, 3- схема выборки импульсов, 4- счетчик-делитель частоты на R, 5- блок суммирования потоков импульсов.

Информация о состоянии счетчика 1 (сигнал q) с выходов счетчика подается на блок формирования сигнала выборки импульсов 3. В простейшем случае этот блок может представлять собой пороговое устройство (по коду числа Q), открывающее схему 3, однако выборка в этом случае имеет характер, близкий к статистическому, только при

достаточно малых различиях частоты потоков n и m/R (порядка  $n < m/R < 1.5n$ ). По мере роста отношения  $m/n$  независимость выборки от импульсов потока n в такой схеме вырождается, и требуются дополнительные устройства адаптации под текущие условия измерений. В общем случае блок 3 выполняется в виде пересчетной схемы с регулируемым коэффициентом пересчета импульсов потока m в интервале  $0 < P_B < 1$  по значению кода q с выхода счетчика 1 (например, с использованием статистических пересчетных схем, выполненных в виде генератора случайных кодов со схемой формирования сигнала выборки очередного импульса из потока m путем сравнения на больше-меньше текущих кодов с выхода счетчика 1 и с выхода генератора кодов).

Импульсы основного потока n и импульсы выборки из потока m, частота которых равна  $P_B m = R \cdot n$ , поступают на вход блока 5 суммирования потоков сигналов. Интенсивность потока импульсов на выходе блока 5 равна  $z = n + P_B m = (1+R)n$ . Блок 5 может содержать пересчетную схему с коэффициентом  $K=(1+R)$ , при этом выходной поток будет приводиться к масштабу основного потока n и появляется возможность синхронного переключения коэффициентов пересчета схем 4 и 5 под различные условия измерений, при этом установка оптимального значения коэффициента R может быть переведена в режим автоматической с управлением по текущему значению (в определенном интервале) информационного кода схемы 1. Альтернативное решение - подача на вход суммирования схемы 5 потока импульсов с выхода схемы 4, при этом частота потока z будет всегда в 2 раза больше потока n.

Попутно отметим, что при выводе информации  $q = \mu R(n/m)$  в цифровом коде со счетчика 1 данная схема может выполнять функции универсального цифрового интенсиметра: средней частоты импульсов (n-var, m-const от генератора тактовой частоты), среднего временного интервала между импульсами (m-var, n-const) и отношения частот n/m двух статистически распределенных потоков импульсов.

## Тема 12. ОПТИМАЛЬНЫЕ ЛИНЕЙНЫЕ ЦИФРОВЫЕ ФИЛЬТРЫ.

Результаты практических измерений, подлежащие обработке, содержат определенный полезный сигнал на фоне различного рода помех (шумов), при этом спектр помех в общем случае представлен по всему интервалу главного частотного диапазона и наложен на спектр полезного сигнала. В этих условиях ставится задача реализации оптимальных фильтров, которые позволяют достаточно надежно производить обнаружение сигнала, наилучшим образом выделять сигнал на фоне помех или подавлять помехи без существенного искажения сигнала.

Главным критерием при проектировании оптимальных фильтров, как правило, является минимизация среднеквадратичной ошибки восстановления полезного сигнала. Линейные оптимальные фильтры, которые рассматриваются в настоящей теме, обычно базируются на оптимальном фильтре Колмогорова-Винера.

### 12.1. СЛУЧАЙНЫЕ ПРОЦЕССЫ И ШУМЫ

Случайные процессы и шумы описываются функциями автокорреляции и

спектрами мощности. Модели случайных процессов и сигналов с заданными статистическими характеристиками обычно получают фильтрацией белого шума.

**Белый шум** является стационарным случайным процессом  $q(t)$ , у которого автокорреляционная функция описывается дельта - функцией Дирака, спектральная плотность мощности не зависит от частоты и имеет постоянное значение  $W_q(f) = \sigma^2$ , равное дисперсии значений  $q(t)$ . Все спектральные составляющие белого шума имеют одинаковую мощность. По существу, это идеализированный случайный процесс с бесконечной энергией. Но в случае постоянства спектральной плотности мощности случайного процесса в конечном диапазоне частот такая идеализация позволяет достаточно просто разрабатывать оптимальные методы фильтрации. Многие помехи в радиотехнике, в технике связи и в других отраслях, в том числе в информатике, рассматривают как белый шум, если эффективная ширина спектра сигналов  $B_s$  много меньше эффективной ширины спектра шумов  $B_q$ , а спектральная плотность мощности шумов слабо изменяется в интервале спектра сигнала. Понятие "белый шум" определяет только спектральную характеристику случайного процесса и под это понятие подпадают любые случайные процессы, имеющие равномерный энергетический спектр и различные законы распределения.

Если частотный диапазон спектра, на котором рассматриваются сигналы и помехи, равен  $0-B$ , то спектральная плотность шума задается в виде:

$$W_q(f) = \sigma^2, \quad 0 \leq f \leq B; \quad W_q(f) = 0, \quad f > B, \quad (12.1.1)$$

при этом корреляционная функция шума определяется выражением:

$$R_q(\tau) = \sigma^2 B \cdot \sin(2\pi B\tau) / 2\pi B\tau. \quad (12.1.2)$$

Эффективный интервал корреляции:

$$T_k = 2 \int_0^{\infty} |R_q(\tau)| d\tau / R_q(0). \quad (12.1.3)$$



Рис. 12.1.1. Функции корреляции белого шума в частотном интервале  $0-B$ .

Реальный интервал корреляции целесообразно определять по ширине главного максимума функции  $R_q(\tau)$  (значения  $\tau$  при первых пересечениях нулевой линии), в котором сосредоточена основная часть энергии шумов, при этом  $T_k = 1/B$  и  $BT_k = 1$ .

Как следует из всех этих выражений и наглядно видно на рис. 12.1.1, при ограничении частотного диапазона в шумах появляется определенная корреляция между значениями. Чем меньше частотный диапазон шумов, тем больше их радиус корреляции. Ограничение шумов определенным частотным диапазоном эквивалентно фильтрации белого шума частотным фильтром с соответствующей шириной полосы пропускания, при этом корреляционная функция импульсного отклика фильтра свертывается с дельта – функцией белого шума.

**Модель белого шума**  $q(t)$  можно формировать как случайную по времени (аргументу) последовательность дельта - импульсов  $\delta(t_i)$  со случайными амплитудными значениями  $a_i$ :

$$q(t) = \sum_i a_i \delta(t-t_i), \quad (12.1.4)$$

которая удовлетворяет условиям статистической однородности: постоянное среднее число импульсов в единицу времени и статистическая независимость появления каждого импульса от предыдущих. Такой поток импульсов называют пуассоновским, он является некоррелированным и имеет равномерный спектр плотности мощности:

$$W_q(\omega) = c^2 = N\sigma_a^2,$$

где  $N$  - число импульсов на интервале  $T$  реализации случайного процесса,  $\sigma_a^2$  - дисперсия

амплитуд импульсов.

**Фильтрация белого шума.** Если на входе фильтра с импульсным откликом  $h(t)$  действует белый шум  $q(t)$ , то сигнал на выходе фильтра:

$$g(t) = h(t) \otimes q(t) = h(t) \otimes \sum_i a_i \delta(t-t_i) = \sum_i a_i h(t-t_i), \quad (12.1.5)$$

т.е. выходной сигнал будет представлять собой последовательность сигналов импульсной реакции фильтра  $h(t)$  с амплитудой  $a_i$ , при этом автокорреляционная функция и спектр мощности выходного потока также становятся подобными ФАК и спектру мощности импульсной реакции фильтра, и в первом приближении определяются выражениями:

$$R_g(\tau) \approx N \sigma_a^2 R_h(\tau) = c^2 R_h(\tau), \quad (12.1.6)$$

$$W_g(\omega) \approx N \sigma_a^2 |H(\omega)|^2 = c^2 |H(\omega)|^2. \quad (12.1.7)$$

Этот результат известен как теорема Кэмпбелла.

## 12.2. КРИТЕРИИ ПОСТРОЕНИЯ ОПТИМАЛЬНЫХ ФИЛЬТРОВ.

В практике обработки данных используются три основных критерия построения оптимальных фильтров: минимум среднего квадратического отклонения профильтрованного сигнала от его действительного или заданного значения, максимум отношения сигнал/шум и максимум энергетического отношения сигнал/шум на выходе фильтра. Критерии исходят из вероятностно - статистической модели обрабатываемых данных. При анализе и синтезе фильтров используется аддитивная модель входного сигнала:  $x(k) = s(k)+q(k)$ , где  $s(k)$  - полезная составляющая сигнала,  $q(k)$  - составляющая шумов и помех. Синтез оптимальных фильтров производится с максимальным использованием известной информации как о сигналах, которые необходимо выделить, так и о шумах и помехах. Как правило, используется информация о природе полезного сигнала и шума, об их спектральном составе, о корреляционных и взаимных корреляционных характеристиках. Наличие определенных особенностей (различий) в характеристиках сигнала и шума позволяет реализовать фильтр вообще и оптимальный фильтр в частности.

В геофизической практике априорные данные о полезных сигналах являются достаточно определенными, особенно для активных методов геофизики (сейсмические методы, электроразведка на переменном токе, индукционные методы ядерной геофизики и пр.). Определение характеристик действующих помех представляет собой более сложную проблему, но даже при полной неопределенности можно допустить, что помеха является нормальным стационарным процессом с нулевым средним значением.

**Среднее квадратическое отклонение.** При наличии помех точное выделение полезного сигнала методами линейной фильтрации, как правило, невозможно. Результат фильтрации

$$y(k) = h(n) \otimes x(k-n) \quad (12.2.1)$$

отличается от  $s(k)$  на величины  $\varepsilon(k) = y(k)-s(k)$ , которые являются абсолютными значениями погрешности воспроизведения полезного сигнала по координатам  $k$ . Качество фильтра оценивается средним значением квадрата величины  $\varepsilon(k)$ :

$$\overline{\varepsilon^2} = \overline{[y(k) - s(k)]^2}. \quad (12.2.2)$$

Во многих задачах анализа данных не требуется восстановления исходной формы сигнала  $s(k)$ , т.к. в процессе его дальнейшей обработки осуществляется преобразование сигнала  $s(k)$  в сигнал  $z(k)$ , форма которого может быть более удобной для извлечения (измерения) каких-либо информационных параметров сигнала (амплитуды, частоты, длительности и т.п.). В этом случае фильтр может проектироваться непосредственно на получение выходного сигнала  $z(k)$ . Качество таких формирующих фильтров оценивается средним значением квадрата величины  $\varepsilon(k)$  получения сигнала заданной формы:

$$\overline{\varepsilon^2} = \overline{[y(k) - z(k)]^2}. \quad (12.2.2')$$

Выражения (12.2.2) дают возможность определить значения  $h(k)$  фильтра по критерию минимума среднего квадратического отклонения выходного сигнала от его



действительной или заданной формы.

**Амплитудное отношение сигнал/шум.** При постановке задачи обнаружения (установления факта наличия) в экспериментальных данных сигнала известной формы для проектирования фильтра используется критерий максимума пикового отношения сигнал/шум на выходе фильтра:

$$\rho_a = y_{\text{экс}}/\sigma,$$

где  $y_{\text{экс}}$  - экстремальное значение амплитуды сигнала,  $\sigma$  - средний квадратический уровень значений помех. Если в полезном сигнале отсутствует четко выраженный экстремум, а сам сигнал достаточно протяженный по аргументу, то в качестве критерия используется отношение средних квадратов амплитуд сигнала и шума:

$$\rho^2 = \overline{y^2}/\sigma^2, \quad (12.2.3)$$

где  $y^2$  - средний квадрат амплитуды сигнала в пределах его формы.

**Энергетическое отношение сигнал/шум.** При узко конкретной задаче обнаружения сигнала степень искажения самого сигнала может не ограничиваться. Если кроме обнаружения сигнала, как основной цели обработки данных, ставится и задача оценки его формы, то в этом случае для проектирования фильтра обычно используется критерий максимума энергетического отношения сигнал/шум:

$$\rho = E_y/E_n, \quad (12.2.4)$$

где  $E_y$  и  $E_n$  - энергия соответственно сигнала и шума на выходе фильтра.

### 12.3. ФИЛЬТР КОЛМОГОРОВА-ВИНЕРА.

**Условие оптимальности фильтра.** Фильтр Колмогорова-Винера является оптимальным фильтром формирования из входного сигнала  $x(t)$  выходного сигнала  $z(t)$  при известной форме полезного сигнала  $s(t)$ , который содержится во входном сигнале в сумме с шумами. В качестве критерия его оптимизации используется среднее квадратическое отклонение сигнала  $y(t)$  на выходе фильтра от заданной формы сигнала  $z(t)$ . Подставим уравнение свертки (12.2.1) в раскрытой форме весового суммирования в выражение (12.2.2') и получим отклонение  $\varepsilon^2$  выходного сигнала  $y(k) = h(n) \otimes x(k-n)$  от заданной формы выходного сигнала  $z(k)$  по всем точкам массива данных:

$$\varepsilon^2 = \overline{\left[ \sum_n h(n) x(k-n) - z(k) \right]^2}. \quad (12.3.1)$$

В частном случае воспроизведения формы полезного сигнала в качестве функции  $z(k)$  принимается функция  $s(k)$ . Минимум выражения (12.3.1) определяет значения коэффициентов  $h(n)$  оптимального фильтра. Для нахождения их значений продифференцируем выражение (12.3.1) по коэффициентам фильтра и приравняем полученные уравнения нулю. В итоге получаем:

$$d(\varepsilon^2)/dh(n) = -\overline{z_k x_{k-m}} + \sum_n h_n \overline{x_{k-m} x_{k-n}} = 0,$$

где  $\overline{x_{k-m} x_{k-n}} = R(m-n)$  - корреляционная функция входного сигнала,  $\overline{z_k x_{k-m}} = B(m)$  - взаимная корреляционная функция сигналов  $z(k)$  и  $x(k)$ . Отсюда:

$$\sum_n h_n R(m-n) = B(m), \quad n = m = 0, 1, 2, \dots, M. \quad (12.3.2)$$

В краткой форме символической записи:

$$h(n) \otimes R(m-n) = B(m). \quad (12.3.3)$$

Другими словами, *свертка функции отклика оптимального фильтра с функцией автокорреляции входного сигнала должна быть равна функции взаимной корреляции выходного и входного сигналов.*

**Система линейных уравнений фильтра.** Выражение (12.3.2) может быть записано в виде системы линейных уравнений - однострочных равенств левой и правой части для фиксированных значений координаты  $m$  коэффициентов фильтра. При расчете симметричных фильтров, не сдвигающих фазы выходного сигнала, фильтр может вычисляться только одной половиной своих значений:



диапазону равномерно, в отличие от плотности мощности сигнала  $W_s$ , которая, в зависимости от формы сигнала, может занимать любые частотные интервалы спектрального диапазона. Частотная передаточная функция фильтра воспроизведения сигнала формируется отношением  $W_s(\omega)/[W_s(\omega)+W_q(\omega)]$ . На частотах, где сосредоточена основная энергия сигнала, имеет место  $W_s(\omega) \gg W_q(\omega)$  и  $H(\omega) \Rightarrow 1$  (как минимум, больше 0.5). Там, где значение  $W_s(\omega)$  меньше  $W_q$ , коэффициент передачи фильтра становится меньше 0.5. В пределе  $H(\omega)=0$  на всех частотах, где полностью отсутствуют частотные составляющие сигнала.

Аналогичный процесс имеет место и при формировании произвольного сигнала  $z(t)$  на выходе фильтра. В этом случае из частот входного сигнала устанавливаются на выделение и усиление частоты взаимной мощности входного и выходного сигнала, необходимые для формирования сигнала  $z(t)$ , причем коэффициент усиления на этих частотах может быть много больше 1, а подавляться могут не только шумы, но и частоты основного сигнала, если их нет в сигнале  $z(t)$ .

Таким образом, оптимальные фильтры учитывают особенности спектрального состава сигналов и способны формировать передаточные функции любой сложности на выделение полезных частот сигналов из любых диапазонов спектра с максимальных подавлением шумов на всех частотах спектрального диапазона, не содержащих полезных сигналов, при этом границы усиления-подавления устанавливаются автоматически по заданному уровню шумов.

**Задание мощности шумов.** Следует внимательно относиться к заданию функции шумов  $W_q(\omega)$ . При полной неопределенности шума необходимо, как минимум, выполнять оценку его дисперсии  $\sigma^2$  и распространять на весь частотный диапазон с соответствующей нормировкой ( $2 \int_0^{\Omega} W_q(\omega) d\omega = \sigma^2$ ), т.е. считать его белым шумом. При известной функции полезного сигнала в зарегистрированной реализации значение дисперсии шумов в первом приближении может быть оценено по разности дисперсий реализации и функции полезного сигнала. Можно выделить и выделение шумов из входного сигнала в отдельный шумовой массив, например, вейвлетным преобразованием. Однако использовать выделенный шум непосредственно для вычисления функции  $W_q(\omega)$  допустимо только по достаточно представительной реализации при условии стационарности и эргодичности шума. В противном случае функция  $W_q(\omega)$  будет отображать только распределение шумов в зарегистрированной реализации сигнала, а соответственно фильтр будет оптимален только к этой реализации, что не гарантирует его оптимальности к любой другой реализации. Но для обработки единичной зарегистрированной реализации сигнала такой метод не только вполне допустим, но может существенно повысить точность формирования выходного сигнала.

**Эффективность фильтра.** Из выражений (12.3.5-12.3.7) следует, что с позиции минимального искажения полезного сигнала при максимальном подавлении шумов фильтр Колмогорова-Винера эффективен тем в большей степени, чем больше отношение сигнал/шум на входе фильтра. В пределе, при  $W_q(\omega) \ll W_s(\omega)$  имеем  $H(\omega) \Rightarrow 1$  и фильтр воспроизводит входной (или заданный) сигнал без искажений (помех нет, исправлять нечего). Отметим также, что помеха, коррелированная с полезным сигналом, как это следует из (12.3.5), используется фильтром для повышения точности воспроизведения сигнала. С другой стороны, при  $W_q(\omega) \gg W_s(\omega)$  имеем  $H(\omega) \Rightarrow 0$  и сигнал будет сильно искажен, но никакой другой фильтр не обеспечит лучшего результата.

---

**Пример. Расчет оптимального фильтра воспроизведения сигнала. Выполняется в среде Mathcad.**

Форма входного сигнала считается известной и близка к гауссовой. На входной сигнал наложен статистический шум с равномерным распределением мощности по всему частотному диапазону (белый шум), некоррелированный с сигналом, и функцию  $W_zq$  принимаем равной нулю. Для наглядного

просмотра связи параметров фильтра с параметрами сигнала задаем модели сигналов в двух вариантах:

$K := 1000$   $k := 0 .. K$   $A := 50$

$s1_k := A \cdot \exp[-0.0005 \cdot (k-500)^2]$   $s2_k := A \cdot \exp[-0.00003 \cdot (k-500)^2]$   $\leftarrow$  информационные сигналы

$Q := 30$   $q_k := \text{rnd}(Q) - Q/2$   $x1_k := s1_k + q_k$   $x2_k := s2_k + q_k$   $\leftarrow$  входные сигналы

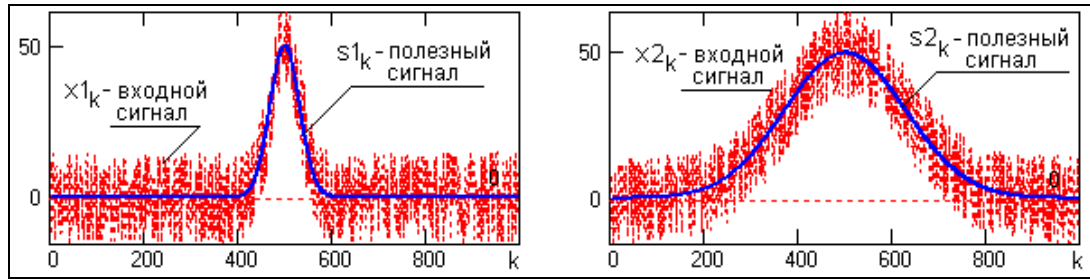


Рис. 12.3.1. Модельные сигналы.

В качестве выходных сигналов задаем те же самые функции  $s1$  и  $s2$ . Быстрым преобразованием Фурье вычисляем спектры сигналов и формируем спектры плотности мощности:

$S1 := \text{CFFT}(s1)$   $S2 := \text{CFFT}(s2)$   $Q := \text{CFFT}(q)$   $\leftarrow$  спектры сигналов

$Ws1 := \overline{(S1 \cdot S1^*)}$   $Ws2 := \overline{(S2 \cdot S2^*)}$   $Wq := \overline{(Q \cdot Q^*)}$   $\leftarrow$  спектры мощности

$Ds1 := \text{var}(s1)$   $Ds2 := \text{var}(s2)$   $Dx1 := \text{var}(x1)$   $Dx2 := \text{var}(x2)$   $Dq := \text{var}(q)$   $\leftarrow$  дисперсии

$Ds1 = 124.308$   $Ds2 = 310.264$   $Dx1 = 202.865$   $Dx2 = 386.78$   $Dq = 79.038$   $\leftarrow$  информация

$\text{mean}(Wq) = 0.079$   $Wq1 := (Dx1 - Ds1)/(K+1)$   $Wq1 = 0.078$   $\leftarrow$  информация

$Wq2 := (Dx2 - Ds2)/(K+1)$   $Wq2 = 0.076$   $\leftarrow$  информация

$Wq_k := Wq1$   $\leftarrow$  замена на постоянное распределение

Для воспроизведения сигналов вычисления функций  $Wzs$  не требуется, т.к.  $Wzs = Ws$ . Вычисление  $Wq$  также имеет только информативный характер.

Передаточные функции оптимальных фильтров (приведены на рис. 12.3.2):

$$H1_k := \frac{Ws1_k}{Ws1_k + Wq_k} \quad H2_k := \frac{Ws2_k}{Ws2_k + Wq_k}$$

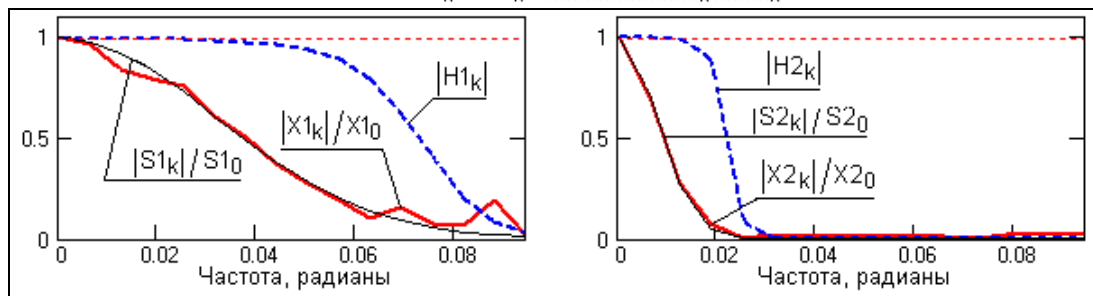


Рис. 12.3.2. Передаточные функции оптимальных фильтров

в сопоставлении с нормированными модулями спектров сигналов

Как следует из рисунка 12.3.2, для плавных монотонных функций, основная энергия которых сосредоточена в низкочастотной части спектра, передаточные функции оптимальных фильтров, по существу, представляют собой низкочастотные сглаживающие фильтры с автоматической подстройкой граничной частоты пропускания под основные частоты входного сигнала. Операторы фильтров можно получить обратным преобразованием Фурье:

$h1 := \text{ICFFT}(H1)/(K+1)$   $h2 := \text{ICFFT}(H2)/(K+1)$   $\leftarrow$  обратное преобразование Фурье

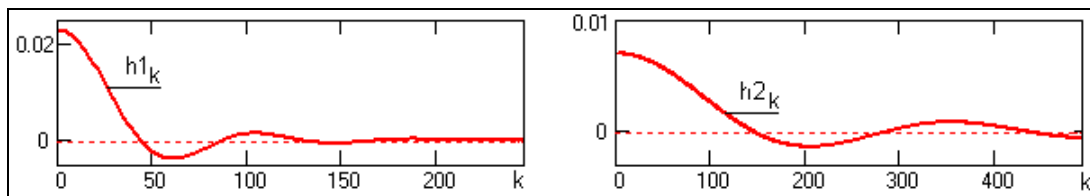


Рис. 12.3.3. Импульсные отклики фильтров.

Оператор фильтра, в принципе, бесконечен. В данном случае, при использовании БПФ максимальное число отсчетов равно  $K/2 = 500$ . Усечение размеров оператора может выполняться по типовым методам с применением весовых функций (в расчете операторы нормируются к 1, весовые функции не применяются).

$N1 := 160$   $n1 := 0 .. N1$   $N2 := 500$   $n2 := 0 .. N2$   $\leftarrow$  размеры и нумерация операторов

$$hm1 := h1_0 + 2 \cdot \sum_{n1} h1_{n1} \quad hm1=0.988 \quad h1 := h1/hm1 \quad \leftarrow \text{нормировка}$$

$$hm2 := h2_0 + 2 \cdot \sum_{n2} h2_{n2} \quad hm2=1.001 \quad h2 := h2/hm2 \quad \leftarrow \text{нормировка}$$

$$y_k := h_0 \cdot x_k + \sum_{n=1}^N h_n \cdot \text{if}(k-n < 0, 0, x_{k-n}) + \sum_{n=1}^N h_n \cdot \text{if}(k+n > K, 0, x_{k+n}) \quad \leftarrow \text{свертка}$$

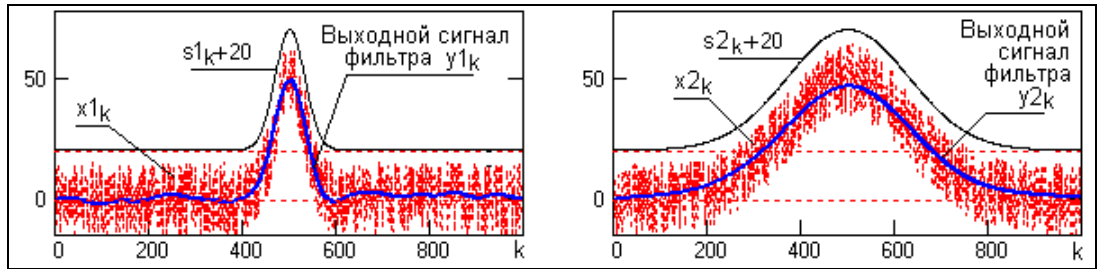


Рис. 12.3.4. Проверка действия оптимальных фильтров.

$$\text{Коэффициент усиления дисперсии помех} \Rightarrow Kd := (h_0)^2 + 2 \cdot \sum_{n=1}^N h_n \quad Kd1=0.021 \quad Kd2= 0.0066$$

Среднеквадратическое отклонение воспроизведения сигнала:

$$e := \sqrt{\frac{1}{K+1} \cdot \sum_k (s_k - y_k)^2} \quad e1= 1.238 \quad e2 = 0.701$$

Проверка действия оператора фильтра приведена на рис. 12.3.4.

Особую эффективность оптимальный фильтр имеет при очистке от шумов сигналов, имеющих достаточно сложный спектральный состав. Оптимальный фильтр учитывает конфигурацию спектра сигнала и обеспечивает максимальное подавление шумов, в том числе внутри интервала основного частотного диапазона сигнала. Это наглядно видно на рис. 12.3.5 для сигнала, близкого к прямоугольному, спектр которого имеет кроме основной низкочастотной части затухающие боковые осцилляции. Расчет фильтра выполнялся по методике, приведенной в примере 1.

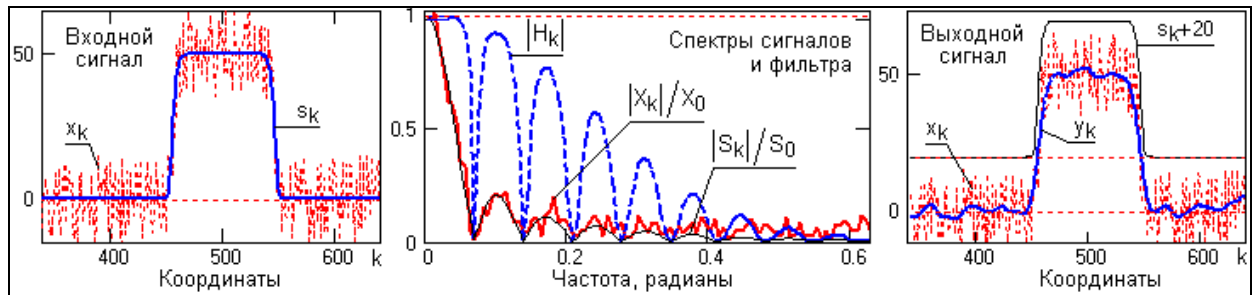


Рис. 12.3.5. Оптимальная фильтрация сигнала со сложным спектральным составом.

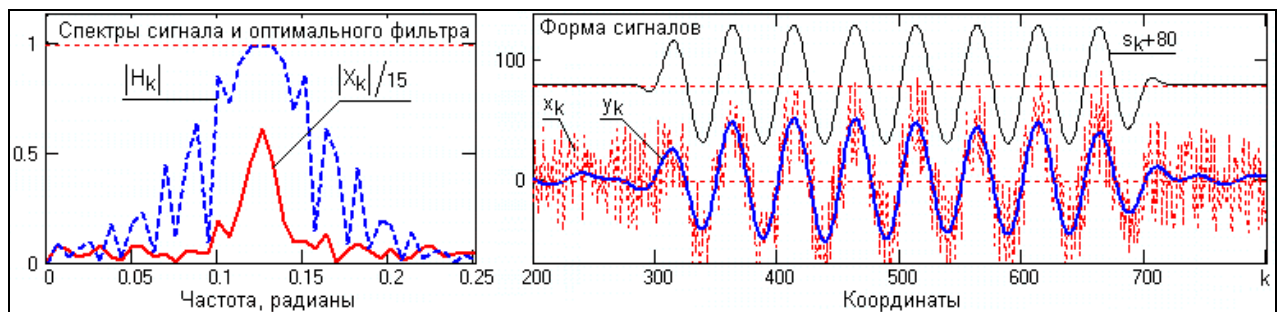


Рис. 12.3.6. Оптимальная фильтрация радиоимпульса.

На рис. 12.3.6 приведен пример фильтрации оптимальным фильтром радиоимпульса. Основной пик спектра радиоимпульса находится в области несущей частоты, а боковые полосы определяются формой модулирующего сигнала, в данном случае – прямоугольного импульса. На графике модулей сигнала и передаточной функции

фильтра можно видеть, что оптимальный фильтр превратился в полосовой фильтр, при этом учитывается форма боковых полос спектра сигнала.

**Фильтры прогнозирования и запаздывания.** Если в правой части уравнения (12.3.3) желаемым сигналом задать входной сигнал со сдвигом на величину  $k\Delta t$ , то при этом  $V(m) = R(m+k)$  и уравнение принимает вид:

$$h(n) \otimes R(m-n) = R(m+k). \quad (12.3.8)$$

При  $k > 0$  фильтр называется фильтром прогнозирования и вычисляет будущие значения сигнала по его предшествующим значениям. При  $k < 0$  фильтр является фильтром запаздывания. Реализация фильтра заключается в решении соответствующих систем линейных уравнений для каждого заданного значения  $k$ . Фильтр может использоваться для интерполяции геофизических полей, в том числе в наперед заданные точки, а также для восстановления утраченных данных.

#### 12.4. ОПТИМАЛЬНЫЕ ФИЛЬТРЫ СЖАТИЯ СИГНАЛОВ.

**Условие оптимальности.** Фильтр сжатия сигнала  $x(t)$ , по существу, представляет собой фильтр формирования сигнала  $z(t)$  с эффективной шириной длительности, меньшей по сравнению с эффективной шириной длительности полезного сигнала  $s(t)$  во входном сигнале  $x(t)$ . Расчет оптимального фильтра сжатия может выполняться непосредственно по выражениям (12.3.3).

В предельном случае сжатия сигнала до импульса Кронекера положим, что  $z(k) = \delta(k)$  при статистической независимости сигнала и шума. Отсюда:

$$B_{sz}(m) = \delta(m) \otimes s(k+m) = s(-m).$$

$$h(n) \otimes (R_s(m-n) + R_q(m-n)) = s(-m). \quad (12.4.1)$$

$$H(\omega) = S^*(\omega) / (|S(\omega)|^2 + W_q(\omega)). \quad (12.4.2)$$

При некоррелированной помехе с дисперсией  $\sigma^2$  система уравнений для определения значений коэффициентов  $h(n)$ :

$$h_0(R(0) + \sigma^2) + h_1R(1) + h_2R(2) + h_3R(3) + \dots + h_MR(M) = s(0), \quad (12.4.3)$$

$$h_0R(1) + h_1R(0) + h_2R(1) + h_3R(2) + \dots + h_MR(M-1) = 0,$$

$$h_0R(2) + h_1R(1) + h_2R(0) + h_3R(1) + \dots + h_MR(M-2) = 0,$$

.....

$$h_0R(M) + h_1R(M-1) + h_2R(M-2) + \dots + h_MR(0) = 0.$$

При расчете коэффициентов фильтра значение  $s(0)$  обычно принимается произвольным, чаще всего равным площади сигнала  $s(t)$ . Тем самым делается попытка полного сжатия площади сигнала до весовой функции Кронекера, что возможно только для сигналов со спектром в главном диапазоне до частоты Найквиста.

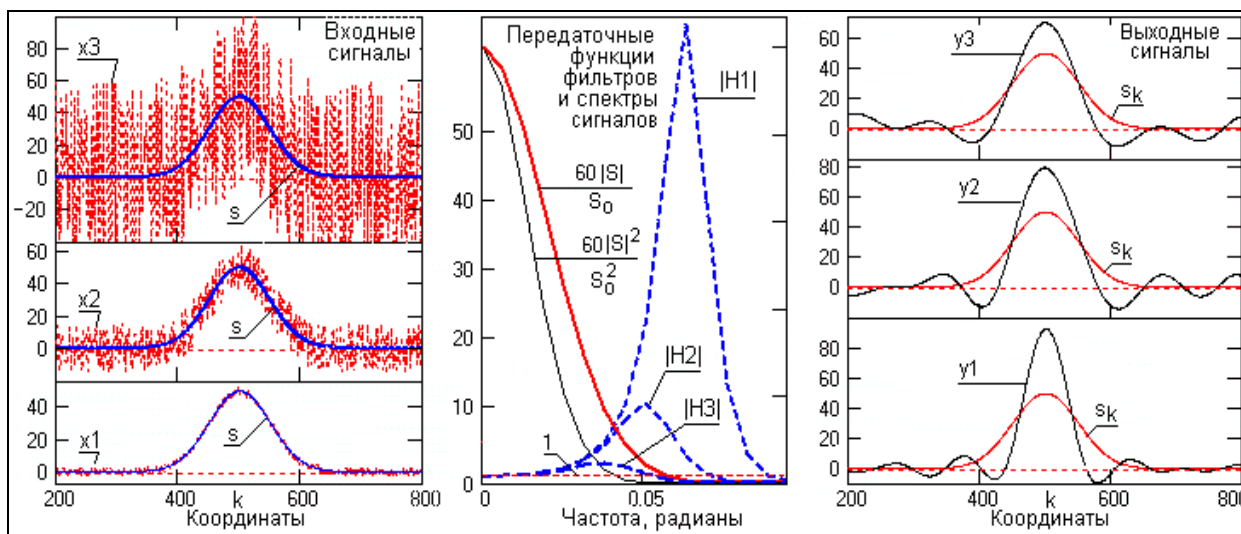


Рис. 12.4.1. Сжатие гладких сигналов с разным уровнем шума.

Для гладких и монотонных функций со спектром в низкочастотной части главного

диапазона сжатие до импульса Кронекера невозможно, и в зависимости от уровня шумов фильтр поднимает насколько возможно высокие частоты сигнала, учитывая значение уровня шумов. При этом нарушаются условия нормировки площади оператора фильтра к 1, о чем можно судить по значению передаточной функции  $H(\omega)$  при  $\omega=0$ , которое становится меньше 1, и при обратном преобразовании  $H(\omega) \Rightarrow h(m)$  оператор  $h(m)$  требует нормировки к 1. Все эти факторы можно наглядно видеть на рис. 12.4.1.

На рисунке приведены три сигнала с одной и той же базовой функцией, на которую наложены шумы разного уровня. При малом уровне шумов (сигнал  $x_1$ ) фильтр в максимальной степени использует высокие частоты сигнала ( $|H_1| \gg 1$  на этих частотах), сохраняя устойчивость работы фильтра при достаточно удовлетворительном (хотя и больше 1) коэффициенте усиления дисперсии помех при максимально возможном сжатии сигнала. При повышении уровня шумов (сигналы  $x_2$  и  $x_3$ ) подъем высоких частот сигнала уменьшается, и сжатие сигнала соответственно также уменьшается, предпочтение отдается максимальному подавлению шумов.

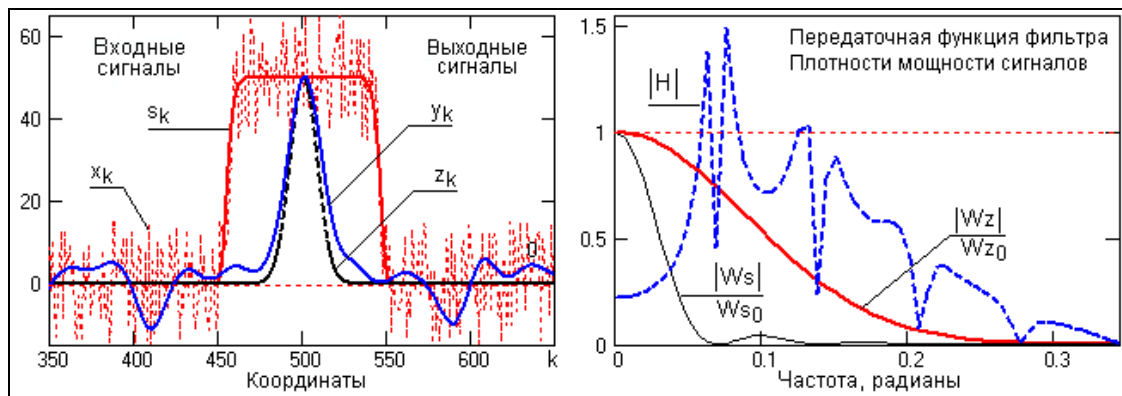


Рис. 12.4.2. Сжатие сигнала с высокочастотным спектром

На рис. 12.4.2. приведен пример сжатия сигнала, близкого к прямоугольному импульсу. Базовая функция сигнала  $s(k)$  имеет достаточно высокочастотный спектр мощности  $W_s(\omega)$ , и при задании формы выходного сигнала сжатия в виде гауссовой функции  $z(k)$  передаточная функция фильтра  $H(\omega)$  обеспечивает уверенное сжатие сигнала (при уменьшении уровня шумов практически до заданной формы).

В пределе, при  $W_q=0$  фильтр сжатия превращается в фильтр деконволюции:

$$H(\omega) = S^*(\omega) / |S(\omega)|^2 = 1/S(\omega), \quad (12.4.4)$$

На выходе такого фильтра имеем:

$$Y(\omega) = H(\omega)X(\omega) \rightarrow 1, \quad \text{при } X(\omega) \rightarrow S(\omega).$$

Реализация фильтра возможна только при условии  $S(\omega) > 0$  на всех частотах в главном частотном диапазоне. В противном случае, при  $S(\omega_i) \rightarrow 0$ ,  $H(\omega_i) \rightarrow \infty$  и фильтр становится неустойчивым. Для исключения возможности такого явления в фильтр (12.4.4) вводится стабилизатор  $\alpha$ :

$$H(\omega) = S^*(\omega) / [|S(\omega)|^2 + \alpha], \quad (12.4.5)$$

где  $|S(\omega)|^2 + \alpha > 0$  во всем частотном диапазоне.

Фильтры деконволюции могут использоваться не только для повышения разрешающей способности данных, но и для интерпретации геофизических данных, если формирование полезного входного сигнала удовлетворяет принципу суперпозиции данных по зависимости от искомым параметров.

## 12.5. ФИЛЬТР ОБНАРУЖЕНИЯ СИГНАЛОВ .

Фильтр используется при решении задач обнаружении сигналов известной формы на существенном уровне шумов, значение которых может превышать значения сигналов. В процессе фильтрации необходимо зафиксировать наличие сигнала в массиве данных, если он там присутствует (может не присутствовать), при этом сохранения формы сигнала не требуется. Сама форма сигнала полагается известной либо по теоретическим данным,



либо по результатам измерений на моделях. Для уверенного обнаружения сигнала фильтр должен обеспечить максимально возможную амплитуду выходного сигнала над уровнем помех и выполняется на основе критерия максимума пикового отношения сигнал/шум.

**Частотная характеристика.** Для расчета фильтра требуется задать известную форму полезного сигнала  $s(k) \Leftrightarrow S(\omega)$  и функцию автокорреляции или спектр мощности помех  $R_q(m) \Leftrightarrow W_q(\omega)$ . Полный входной сигнал принимается по аддитивной модели:  $x(t) = s(t)+q(t)$ . На выходе проектируемого фильтра  $h(n) \Leftrightarrow H(\omega)$  для составляющих выходного сигнала имеем:

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) S(\omega) \exp(j\omega t) d\omega, \quad (12.5.1)$$

$$\sigma^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 W_q(\omega) d\omega, \quad (12.5.2)$$

где  $\sigma$  - средняя квадратическая амплитуда выходной помехи.

Оптимальным в задаче обнаружения одиночного сигнала конечной длительности является фильтр, обеспечивающий на выходе максимальное отношение пиковой мощности сигнала к мощности шума в момент окончания импульса. Значения (12.5.1, 12.5.2) используются для задания критерия максимума отношения сигнал/шум (12.2.3) для произвольной точки  $t_i$ :

$$\rho = [y(t_i)]^2/\delta^2. \quad (12.5.3)$$

Исследование функции (12.5.3) на максимум показывает, что он достигается при частотной характеристике фильтра:

$$H(\omega) = \exp(-j\omega t_i) |S^*(\omega)| / W_q(\omega), \quad (12.5.4)$$

Для физически реализуемых фильтров в качестве точки  $t_i$  целесообразно использовать интервал длительности импульса  $\tau$ , при этом:

$$H(\omega) = \exp(-j\omega\tau) |S^*(\omega)| / W_q(\omega) = \exp(-j\varphi_s - j\omega\tau) |S(\omega)|/W_q(\omega). \quad (12.5.4')$$

Аргумент  $\varphi_s$  в этом выражении компенсирует фазовые сдвиги составляющих спектра сигнала, а  $\omega\tau$  обеспечивает их задержку на время длительности сигнала. Таким образом, на концевой части сигнала фильтр выполняет синфазное суммирование всех полезных частотных составляющих входного сигнала с весами, пропорциональными отношению  $|S(\omega)|/W_q(\omega)$ , что обеспечивает накопление амплитуды полезного сигнала на интервале всей длительности входного импульса и формирует максимум сигнала на момент его окончания. Вместе с тем фильтр ослабляет спектральные составляющие шума тем сильнее, чем меньше модуль  $|S(\omega)|$ , и полная мощность шума на выходе фильтра оказывается меньшей, чем на входе.

Для получения линейных уравнений расчета коэффициентов фильтра без потери общности можно принять  $t_i=0$ , при этом:

$$H(\omega) = S^*(\omega)/W_q(\omega) = |S(\omega)|\exp(j\varphi_s(\omega)) / W_q(\omega). \quad (12.5.5)$$

При переходе во временную (координатную) область:

$$H(\omega)W_q(\omega) = S^*(\omega) \Leftrightarrow h(n) \otimes R_q(n-m) = s(-m). \quad (12.5.6)$$

**Система линейных уравнений** для расчета фильтра:

$$\begin{aligned} h_0R_q(0)+ h_1R_q(1)+ h_2R_q(2)+ h_3R_q(3)+ \dots+ h_MR_q(M) &= S(-M), \\ h_0R_q(1)+ h_1R_q(0)+ h_2R_q(1)+ h_3R_q(2)+ \dots+ h_MR_q(M-1) &= S(-M+1), \\ h_0R_q(2)+ h_1R_q(1)+ h_2R_q(0)+ h_3R_q(1)+ \dots+ h_MR_q(M-2) &= S(-M+2), \\ \dots & \\ h_0R_q(M)+ h_1R_q(M-1)+ h_2R_q(M-2)+ \dots+ h_MR_q(0) &= S(0). \end{aligned}$$

При задании  $t_i$  по центру симметричных входных сигналов можно получить симметричные двусторонние фильтры, не изменяющие фазы сигнала.

На рис. 12.5.1 приведен пример фильтрации фильтром обнаружения сигнала радиоимпульса (информационный сигнал) в сумме с шумами (входной сигнал) при отношении сигнал/шум по средним амплитудным значениям на входе фильтра порядка 1.



Аналогичное отношение сигнал/шум на выходе фильтра повышается до 7 по интервалу полезного сигнала в целом, и превышает 8 в центральной части интервала сигнала.

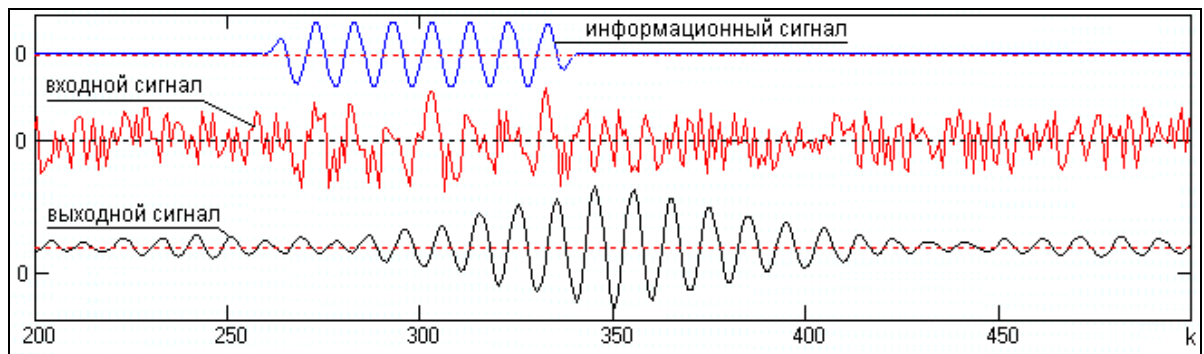


Рис. 12.5.1. Фильтр обнаружения сигнала.

**Эффективность фильтра.** Из выражения (12.5.5) можно видеть, что фильтр имеет максимальный коэффициент передачи на частотах доминирования сигнала и минимальный коэффициент передачи на частотах доминирования помех. Синфазность суммирования всех частотных составляющих выходного сигнала обеспечивает максимальную амплитуду выходного сигнала в заданный момент времени  $t_i$ . Значение максимальной амплитуды можно оценить, приняв  $t_i=0$ , при этом выходной сигнал:

$$y(0) \Leftrightarrow S(\omega) H(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S(\omega) S^*(\omega)}{W_q(\omega)} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{W_q(\omega)} d\omega.$$

Коэффициент передачи фильтра прямо определяется спектром подлежащего обнаружению сигнала, его формой и длительностью. Для оценки эффективности фильтра зададим входной сигнал в виде прямоугольного импульса амплитудой  $u_0$  длительностью  $\tau$  на интервале  $0-\tau$ . Спектральная плотность прямоугольного импульса при интегральном преобразовании Фурье:

$$P(\omega) = (1 - \exp(-j\omega\tau))/j\omega, \quad P^*(\omega) = (\exp(j\omega\tau) - 1)/j\omega.$$

При подстановке в (12.5.4'), принимая  $W_q(\omega) = \text{const}$ , коэффициент передачи фильтра:

$$H(\omega) = \alpha[(\exp(j\omega\tau) - 1)\exp(-j\omega\tau)]/j\omega = \alpha(1 - \exp(-j\omega\tau))/j\omega,$$

где  $\alpha$  - коэффициент пропорциональности с размерностью, обратной спектральной плотности, для получения безразмерных значений коэффициента  $H(\omega)$ . При  $\alpha=1$  (нормировка оператора фильтра производится, как правило, по коэффициенту усиления постоянной составляющей входного сигнала) сигнал на выходе фильтра:

$$u_{\text{вых}}(t) = (u_0/2\pi) \int_{-\infty}^{\infty} P(\omega)H(\omega) d\omega = (u_0/2\pi) \int_{-\infty}^{\infty} (1 - \exp(-j\omega\tau))^2 \exp(j\omega t) d\omega,$$

$$u_{\text{вых}}(t) = U_0 \{ t|_{t>0} - 2(t-\tau)|_{t>\tau} + (t-2\tau)|_{t>2\tau} \}.$$

Как можно видеть на рис 12.5.2, выходной сигнал для входного прямоугольного импульса представляет собой треугольный импульс длительностью  $2\tau$  по основанию с максимальным значением амплитуды на концевой части входного импульса. Это определяется тем, что при  $W_q(\omega)=1$  оператор фильтра полностью повторяет форму входного сигнала (прямоугольного импульса), а выходной сигнал в отсутствие шумов представляет собой свертку двух одинаковых импульсов, максимальное значение которой

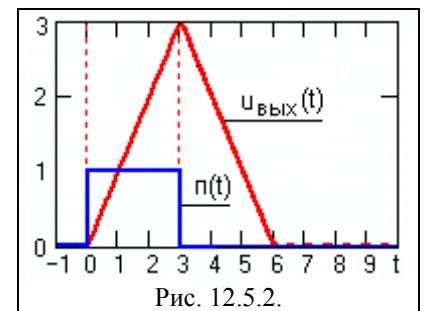


Рис. 12.5.2.

достигается при полном входе сигнала в интервал оператора фильтра ( $t=\tau$ ) и равно полной энергии входного импульса:

$$U_0 = \int_0^\tau p(t) \cdot h(t) dt = \int_0^\tau p(t)^2 dt = u_0^2 \cdot \tau.$$

Значение  $U_0$  определяется нормировкой оператора фильтра  $\alpha$ . Что касается усиления дисперсии (мощности) шумов, то, как известно, дисперсия шума на выходе фильтра равна дисперсии входных шумов  $\sigma^2$ , умноженной на интеграл квадрата импульсного отклика фильтра (для цифровых систем – сумма квадратов коэффициентов оператора фильтра):

$$\sigma_{\text{вых}}^2 = \sigma^2 \int_0^\tau h^2(t) dt = (\sigma^2/2\pi) \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega.$$

Для вычисления интеграла модуль передаточной функции фильтра для прямоугольного импульса может быть представлен в виде интегрального синуса:

$$\int_{-\infty}^{\infty} |H(\omega)|^2 d\omega = 2\tau u_0^2 \int_{-\infty}^{\infty} \text{sinc}^2(\omega\tau/2) d(\omega\tau/2) = 2\pi u_0^2 \tau.$$

Дисперсия шумов на выходе:

$$\sigma_{\text{вых}}^2 = \sigma^2 u_0^2 \tau.$$

С использованием этого выражения для отношения мощности сигнала к мощности шума для сигналов на входе и выходе фильтра имеем:

$$\rho_{\text{вх}} = u_0^2/\sigma^2, \quad \rho_{\text{вых}} = u_0^4 \tau^2 / \sigma^2 u_0^2 \tau = u_0^2 \tau / \sigma^2.$$

Для отношения амплитудных значений сигнала к среднеквадратическим значениям шума:

$$\rho_{\text{вх}} = u_0/\sigma, \quad \rho_{\text{вых}} = (u_0/\sigma) \sqrt{\tau}.$$

Отсюда следует, что эффективность фильтра тем выше, чем больше длительность взаимодействия сигнала с оператором фильтра. Фильтр жестко настраивается под форму сигнала, и любое изменение формы сигнала понижает его эффективность.

Отметим также, что коэффициент передачи фильтра тем больше и эффективность его работы тем выше, чем больше различия в форме частотных спектров сигнала и шумов. При постоянной форме спектров сигнала и шума любой другой фильтр уступает данному фильтру, как по пиковому, так и по энергетическому отношению сигнал/шум на выходе фильтра.

**Согласованный фильтр.** При помехах типа белого шума  $W_q(\omega) = \sigma^2$  и из (12.5.5) следует  $H(\omega) = S^*(\omega)/\sigma^2$ . Постоянный множитель  $1/\sigma^2$  может быть опущен. Частотная характеристика фильтра с точностью до постоянного множителя определяется только спектром сигнала, вследствие чего он и получил название согласованного (со спектром сигнала).

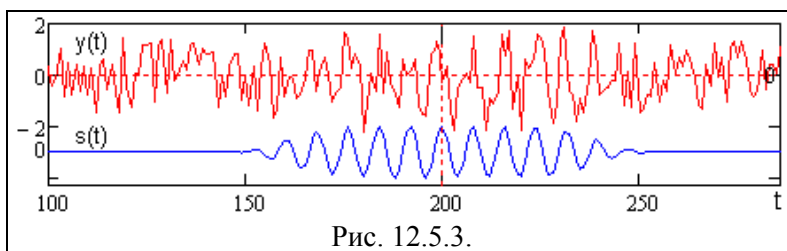


Рис. 12.5.3.

В качестве примера на рис. 12.5.3 приведен радиоимпульсный сигнал  $s(t)$ , форма которого хорошо известна априорно, и зашумленный входной сигнал  $y(t) = s(t) + q(t)$ , при этом мощность шумового сигнала

$q(t)$  в 2 раза больше мощности радиопульса.

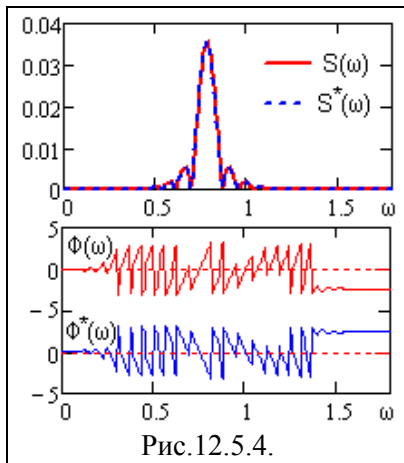


Рис.12.5.4.

и выходного сигнала согласованного фильтра:

$$u(t) = y(t) \otimes h(t') = s(t) \otimes h(t') + q(t) \otimes h(t'). \quad (12.5.8)$$

Первый член этой формулы  $s(t) \otimes h(t')$ , с учетом реверса оператора  $h(t') = s(-t')$  при свертке относительно  $s(t')$ , это функция автокорреляции сигнала  $s(t')$ , сдвинутая относительно начала сигнала  $s(t)$  в 12.5.8 на длину оператора  $h(t')$ , т.е. с максимумом на концевой части сигнала  $s(t)$ . Что касается второго члена формулы, то свертка  $q(t) \otimes h(t')$  представляет собой интегрирование по интервалу оператора случайных шумовых компонент, умножаемых на весовую функцию  $h(t')$  с нулевым средним значением, что приводит к их уменьшению, хотя спектральная характеристика этих шумов перекрывается со спектром выделяемого сигнала.

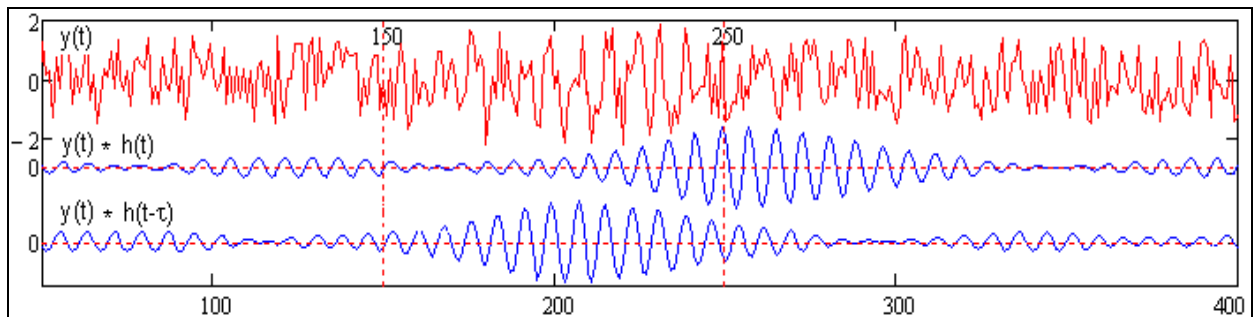


Рис. 12.5.6.

Пример согласованной фильтрации приведен на рис. 12.5.6. Для исключения сдвига по фазе выходного сигнала относительно входного можно выполнять сдвиг оператора фильтра влево по координатам на половину интервала длины оператора (второй график свертки на рисунке).

Фильтр мало эффективен при выделении коротких импульсных или длинных гармонических сигналов.

**Обратный фильтр.** Допустим, что помеха имеет такой же частотный состав, что и полезный сигнал, т.е.:

$$W_q = \sigma^2 |S(\omega)|^2.$$

Выделение полезного сигнала в таких условиях сомнительно. Тем не менее, определим оптимальный фильтр:

$$H(\omega) = S^*(\omega) / [\sigma^2 |S(\omega)|^2] = 1 / [\sigma^2 S(\omega)]. \quad (12.5.9)$$

Выражение (12.5.9) с точностью до постоянного множителя соответствует фильтру сжатия сигнала. Но если согласованный фильтр и фильтр сжатия рассматривать в качестве предельных случаев при полной неопределенности характеристики помех, то в качестве

На рис. 12.5.4 приведены модули спектров  $S(\omega) \leftrightarrow s(t)$ ,  $S^*(\omega)$  и аргументы этих спектров  $\Phi(\omega)$  и  $\Phi^*(\omega)$  в правой половине главного частотного диапазона. Согласованный фильтр имеет АЧХ, подобную фильтру сигнала и фазово-частотную характеристику, комплексно сопряженную с ФЧХ сигнала. При этом оператор фильтра определяется выражением:

$$S^*(\omega) \leftrightarrow s(-t') = h(t'), \quad (12.5.7)$$

что представляет собой реверсирование известной формы сигнала в интервале его значимых значений (для данного примера на рис. 12.5.3  $t'=0 \dots 100$  из интервала  $t=150 \dots 250$ )

На рис. 12.5.5 приведены модули спектров исходного и

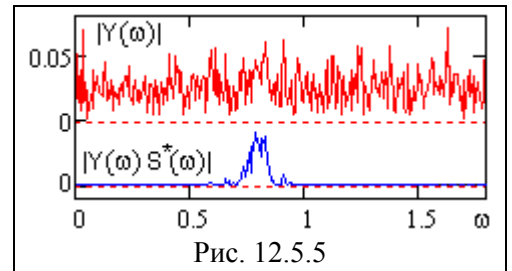


Рис. 12.5.5

модели помех можно принять их суперпозицию:

$$W_q = a^2 |S(\omega)|^2 + b^2.$$

Подставляя это выражение в (12.5.5), с точностью до множителя получаем:

$$H(\omega) = S^*(\omega) / [|S(\omega)|^2 + \gamma^2], \quad (12.5.10)$$

где  $\gamma = b/a$  - отношение дисперсий шума и сигнала. Фильтр стремится к согласованному при больших  $\gamma$ , и к обратному (фильтру сжатия) при малых.

### 12.6. ЭНЕРГЕТИЧЕСКИЙ ФИЛЬТР.

Энергетический фильтр максимизирует отношение сигнал/помеха по всей длине фильтра (а не в отдельной точке), и если сигнал по своей протяженности укладывается в окно фильтра, то тем самым обеспечивается оценка формы сигнала. Фильтр занимает промежуточное положение между фильтром воспроизведения сигнала Колмогорова-Винера и согласованным фильтром и требует задания корреляционных функций сигнала и помех. Сигнал может быть представлен и в детерминированной форме с соответствующим расчетом его автокорреляционной функции.

**Критерий оптимальности.** Энергия сигнала на выходе фильтра:

$$E_{sh} = \sum_k s_k^2 = \sum_k (\sum_n h_n s_{k-n})^2 = \sum_k h_k \sum_n h_n R_s(k-n), \quad (12.6.1)$$

где  $R_s$ - функция автокорреляции сигнала. В векторной форме:

$$E_{sh} = \bar{\mathbf{h}}^T \bar{\mathbf{R}}_s \bar{\mathbf{h}}. \quad (12.6.2)$$

Аналогично, выражение для энергии помех на выходе:

$$E_{qh} = \sum_k h_k \sum_n h_n R_q(k-n) = \bar{\mathbf{h}}^T \bar{\mathbf{R}}_q \bar{\mathbf{h}}, \quad (12.6.3)$$

где  $R_q$  - функция автокорреляции помех. При некоррелированной помехе  $E_{qh} = \sigma^2$ .

Подставим (12.6.2, 12.6.3) в выражение (12.2.4):

$$\rho = \bar{\mathbf{h}}^T \bar{\mathbf{R}}_s \bar{\mathbf{h}} / \bar{\mathbf{h}}^T \bar{\mathbf{R}}_q \bar{\mathbf{h}}. \quad (12.6.4)$$

**Расчет векторов операторов фильтров.** Для определения значений вектора  $\bar{\mathbf{h}}$  продифференцируем  $\rho$  по  $\bar{\mathbf{h}}$ , и приравняем производную к нулю:

$$[(\bar{\mathbf{h}}^T \bar{\mathbf{R}}_q \bar{\mathbf{h}})(\bar{\mathbf{R}}_s \bar{\mathbf{h}}) - (\bar{\mathbf{h}}^T \bar{\mathbf{R}}_s \bar{\mathbf{h}})(\bar{\mathbf{R}}_q \bar{\mathbf{h}})] / (\bar{\mathbf{h}}^T \bar{\mathbf{R}}_q \bar{\mathbf{h}})^2 = 0.$$

$$(\bar{\mathbf{R}}_s - \rho \bar{\mathbf{R}}_q) \bar{\mathbf{h}} = \mathbf{0}. \quad (12.6.5)$$

В системе уравнений (12.6.5) неизвестны собственные значения  $\rho$  матрицы  $(\bar{\mathbf{R}}_s - \rho \bar{\mathbf{R}}_q)$  и значения коэффициентов  $h_n$ . Система имеет  $N+1$  ненулевых решений относительно значений  $\rho$  и соответствующих этим значениям векторов  $\bar{\mathbf{h}}$ . Для определения коэффициентов фильтра приравнивается к нулю и решается относительно  $\rho$  определитель матрицы  $(\bar{\mathbf{R}}_s - \rho \bar{\mathbf{R}}_q)$ , после чего максимальное значение  $\rho_{max}$  подставляется в (12.6.5) и система уравнений решается относительно коэффициентов  $h_i$  вектора  $\bar{\mathbf{h}}_1$ . При фильтрации сигнала вектор  $\bar{\mathbf{h}}_1$  обеспечивает выделение первой по мощности главной компоненты сигнала, т.е. составляющей сигнала, которая имеет наибольшую энергию и отношение сигнал/шум. В сложных полях такая компонента, как правило, соответствует региональному фону.

В принципе, расчет может быть продолжен и для других значений  $\rho < \rho_{max}$ , и определены значения коэффициентов векторов  $\bar{\mathbf{h}}_2$ ,  $\bar{\mathbf{h}}_3$  и т.д., с использованием которых могут выделяться вторая и прочие компоненты сигнала. Наиболее эффективно такой метод используется для разделения сигналов (полей) при некоррелированных помехах. В этом случае корреляционная матрица помех является единичной (единицы по диагонали, остальное - нули) и уравнение (12.6.5) имеет вид:

$$(\bar{\mathbf{R}}_s - \rho \bar{\mathbf{I}}) \bar{\mathbf{h}} = \mathbf{0}. \quad (12.6.6)$$

В развернутой форме:



восстановления первоначальной формы сигнала операция деконволюции должна выполнить усиление данной гармоники в выходном сигнале в 2 раза и осуществить обратный сдвиг фазы на  $\pi/4$ . Для одной гармоники выполнение такой операции труда не представляет. Но практические задачи деконволюции значительно сложнее, так как требуется, как правило, восстановление полного спектра исходного сигнала, имеющего непрерывный характер.

**Определение деконволюции.** Если для прямой свертки дискретного сигнала  $x(k)$  с импульсным откликом  $h(n)$  линейной системы (фильтра) имеем уравнение:

$$y(k) = h(n) \otimes x(k) \Leftrightarrow H(z)X(z) = Y(z), \quad Y(z) = \sum_{k=-\infty}^{\infty} y_k z^k, \quad z = \exp(-j\varphi),$$

то задача деконволюции в общей форме - определение сигнала на входе линейной системы по значениям выходного сигнала, т.е. снятие реакции (импульсного отклика) системы на сигнал и восстановление исходной формы сигнала, что весьма актуально, например, для регистрирующих систем:

$$X(z) = Y(z)/H(z) = Y(z)H^{-1}(z) \Leftrightarrow y(k) \otimes h^{-1}(n) = x(k), \quad (13.1.1)$$

где индексом "-1" символически обозначена передаточная функция оператора обратного фильтра  $H^{-1}(z) = 1/H(z)$ , т.е. оператора деконволюции, инверсного прямому оператору свертки (импульсному отклику системы). Соответственно, при обратном  $z$ -преобразовании получаем оператор деконволюции:

$$H^{-1}(z) = 1/H(z) \Leftrightarrow h^{-1}(n). \quad (13.1.2)$$

Очевидно, что если имеет место  $H(z)H^{-1}(z) = 1$ , то при обратном  $z$ -преобразовании этого выражения мы должны иметь:

$$H(z)H^{-1}(z) = 1 \Leftrightarrow h(n) \otimes h^{-1}(n) = \delta_0(n), \quad (13.1.3)$$

где  $\delta_0(n)$  - импульс Кронекера. При этом последовательная свертка сигнала  $x(k)$  с оператором системы  $h(k)$  и оператором деконволюции  $h^{-1}(k)$  будет эквивалентна свертке сигнала  $x(k)$  с импульсом Кронекера, что не должно изменить форму сигнала  $x(t)$ .

При  $z = \exp(-j\varphi)$  все изложенное действительно и для спектрального представления операторов. Пример инверсии оператора через спектральное представление приведен на рис. 13.1.1 (исходный оператор  $h_n \rightarrow$  спектральная плотность  $H(\omega) \rightarrow$  инверсная спектральная плотность  $H^{-1}(\omega) \rightarrow$  инверсный оператор  $h^{-1}_n$  на начальном интервале отсчетов).

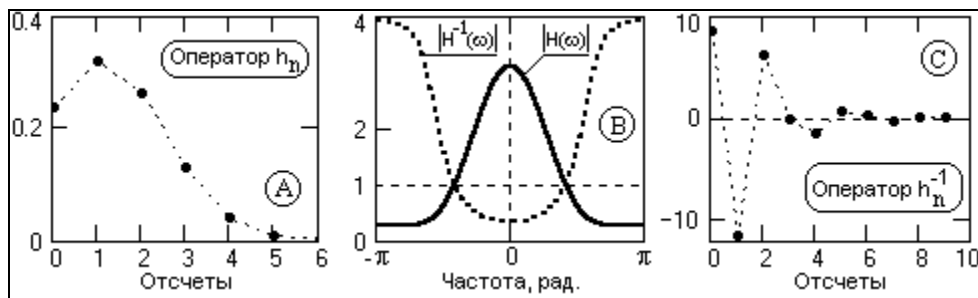


Рис. 13.1.1.

**Особенности деконволюции.** Выражение (13.1.2) позволяет сделать некоторые выводы об особенностях выполнения деконволюции.

При ограниченной импульсной реакции  $h(n)$  инверсный оператор  $h^{-1}(n)$  в общем случае не ограничен. Так, например, если импульсная реакция представлена нормированным диполем  $h(n) = \{1, a\} \Leftrightarrow (1+az) = h(z)$ , то имеем:

$$H^{-1}(z) = 1/(1+az) = 1 - az + a^2z^2 - a^3z^3 + \dots$$

$$h^{-1}(n) = \{1, -a, a^2, -a^3, \dots\}.$$

Это действительно практически для любых операторов свертки, энергия которых на каких-либо ограниченных участках главного частотного диапазона близка к нулевой. При инверсии спектральной функции таких операторов на этих участках

возникают резкие спектральные пики, которые при обратном преобразовании Фурье дают медленно затухающие функции операторов. Пример такого явления приведен на рис. 13.1.2.

Отсюда следует, что для точного выполнения деконволюции необходимо располагать бесконечно длинным инверсным оператором фильтра. Практически, деконволюция выполняется, если инверсный оператор достаточно быстро затухает и может быть ограничен. Но использование усеченных операторов приводит к появлению определенной погрешности деконволюции, величину которой следует контролировать.

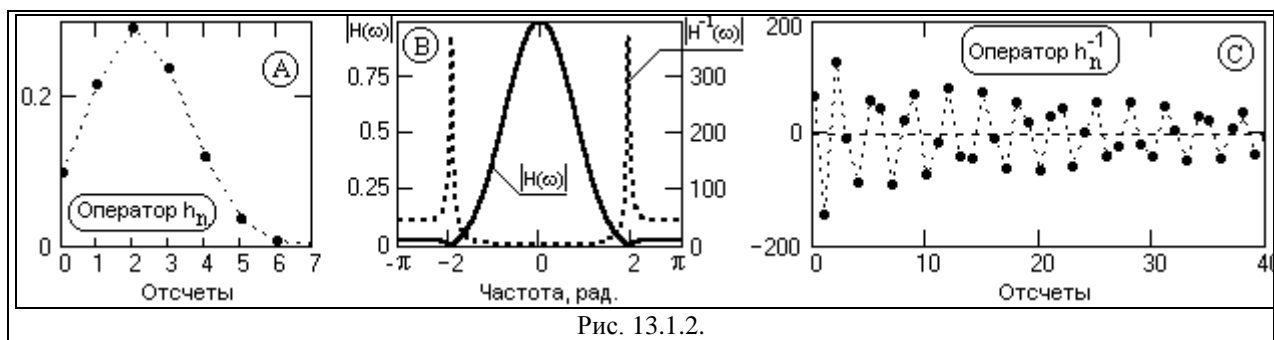


Рис. 13.1.2.

Заметим также, что передаточные функции систем, как правило, имеют низкочастотный характер. Инверсия операторов таких систем всегда связана с усилением высоких частот, что приводит к высоким коэффициентам усиления инверсными фильтрами дисперсии помех, что может приводить к потере полезного сигнала среди усиленных шумовых флюктуаций.

Кроме того, быстрое затухание оператора деконволюции является необходимым, но еще не достаточным условием реализации деконволюции.

**Устойчивость фильтров деконволюции.** Функция  $H(z)$  в выражении (13.1.2) имеет особые точки - нули функции, которые становятся полюсами функции  $H^{-1}(z) = 1/H(z)$  и определяют устойчивость инверсного фильтра. Для того чтобы фильтр деконволюции был устойчивым, ряд  $1/H(z)$  должен сходиться, т.е. полюса функции должны находиться вне единичного круга на  $z$ -плоскости (внутри круга при использовании символики  $z^{-1}$ ).

Многочлен  $H(z)$  порядка  $N$  может быть разложен на  $N$  простых сомножителей - двучленов (диполей):

$$H(z) = (a-z)(b-z)(c-z)\dots, \quad (13.1.4)$$

где  $a, b, c, \dots$  - корни полинома. Обращение передаточной функции:

$$H^{-1}(z) = \frac{1}{a-z} \cdot \frac{1}{b-z} \cdot \frac{1}{c-z} \dots \quad (13.1.5)$$

Если каждый из диполей функции (13.1.4) является минимально-фазовым диракоидом, т.е. корни диполей находится вне единичного круга на  $z$ -плоскости и модули нулевых членов диполей всегда больше следующих за ними первых членов (в данном случае:  $|a|>1, |b|>1, |c|>1$ ), то функция  $H(z)$  также является минимально-фазовым диракоидом. Максимум энергии импульсного отклика минимально-фазового диракоида сосредоточен в его начальной части и последовательность отсчетов представляет собой затухающий ряд. При этом функция  $1/H(z)$  также будет представлять собой передаточную функцию минимально-фазового оператора, обеспечивающую выполнение условия (13.1.3), а ее обратное преобразование - сходящийся ряд устойчивого фильтра. Так, например, фильтр, реализующий передаточную функцию (13.1.5), в самой общей форме может быть выполнен в виде включенных последовательно фильтров, каждый из которых имеет передаточную функцию следующего типа (для первого фильтра):

$$H_1^{-1}(z) = 1/(a-z) = (1+z/a+z^2/a^2+\dots)/a.$$

Отсюда, равно как и непосредственно из выражения (13.1.5), следует также, что чем больше значения модулей корней фильтра (дальше от единичного круга полюс фильтра),

тем быстрее затухает инверсный оператор.

**Пример 1.** Оценить возможность инверсии оператора свертки  $h_n = \{0.131, 0.229, 0.268, 0.211, 0.111, 0.039, 0.009, 0.001\}$ ,  $N = 7$ .

Проверка устойчивости оператора инверсного фильтра в среде Mathcad приведена на рис. 13.1.3. Модули всех корней больше 1, полюсы инверсного полинома находятся за пределами единичной окружности на  $z$ -плоскости, и инверсный оператор должен быть устойчив.

На рисунке показан также оператор деконволюции, который получен обратным преобразованием Фурье передаточной функции  $H_i(\omega)$ . Оператор бесконечен, но затухает достаточно быстро, что обеспечивает высокую точность деконволюции при использовании ограниченного числа членов оператора фильтра (устанавливается пользователем по заданной точности).

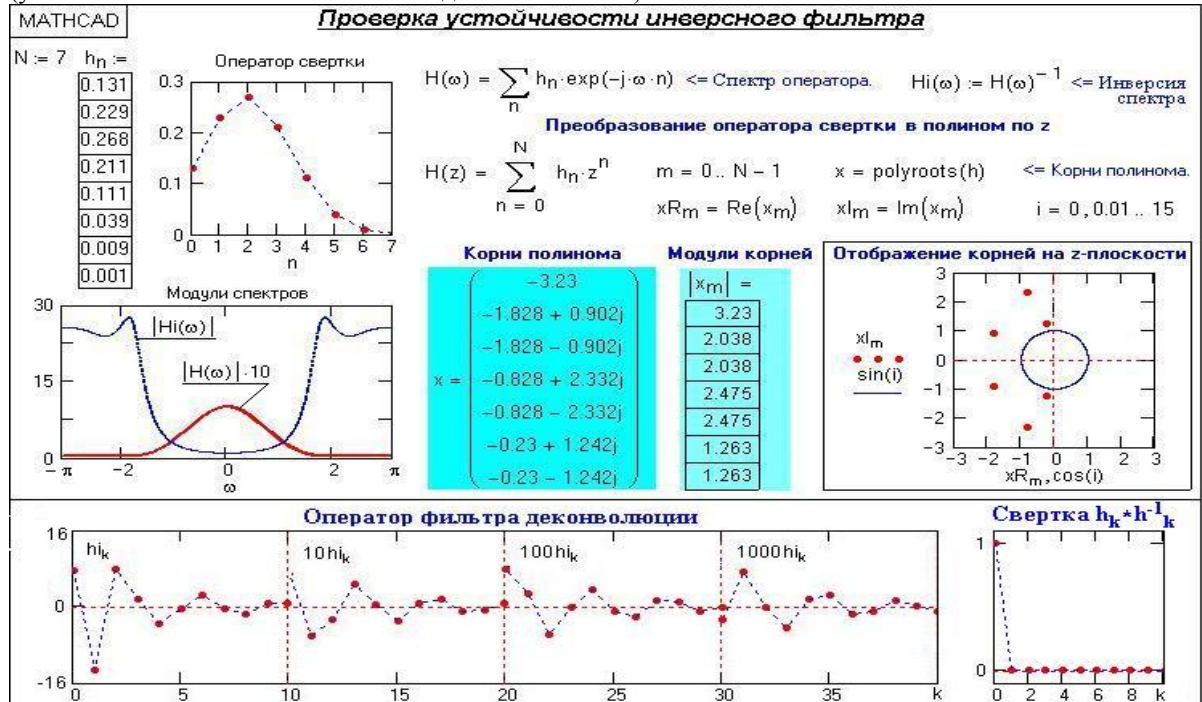


Рис. 13.1.3. Пример устойчивого инверсного фильтра.

**Пример 2.** Сдвинем вышеприведенный оператор на одну позицию вправо и для сохранения той же энергии оператора примем  $h_0 = 0.001$ . Требуется оценить устойчивость инверсного оператора свертки  $h_n = \{0.001, 0.131, 0.229, 0.268, 0.211, 0.111, 0.039, 0.009, 0.001\}$ ,  $N =$



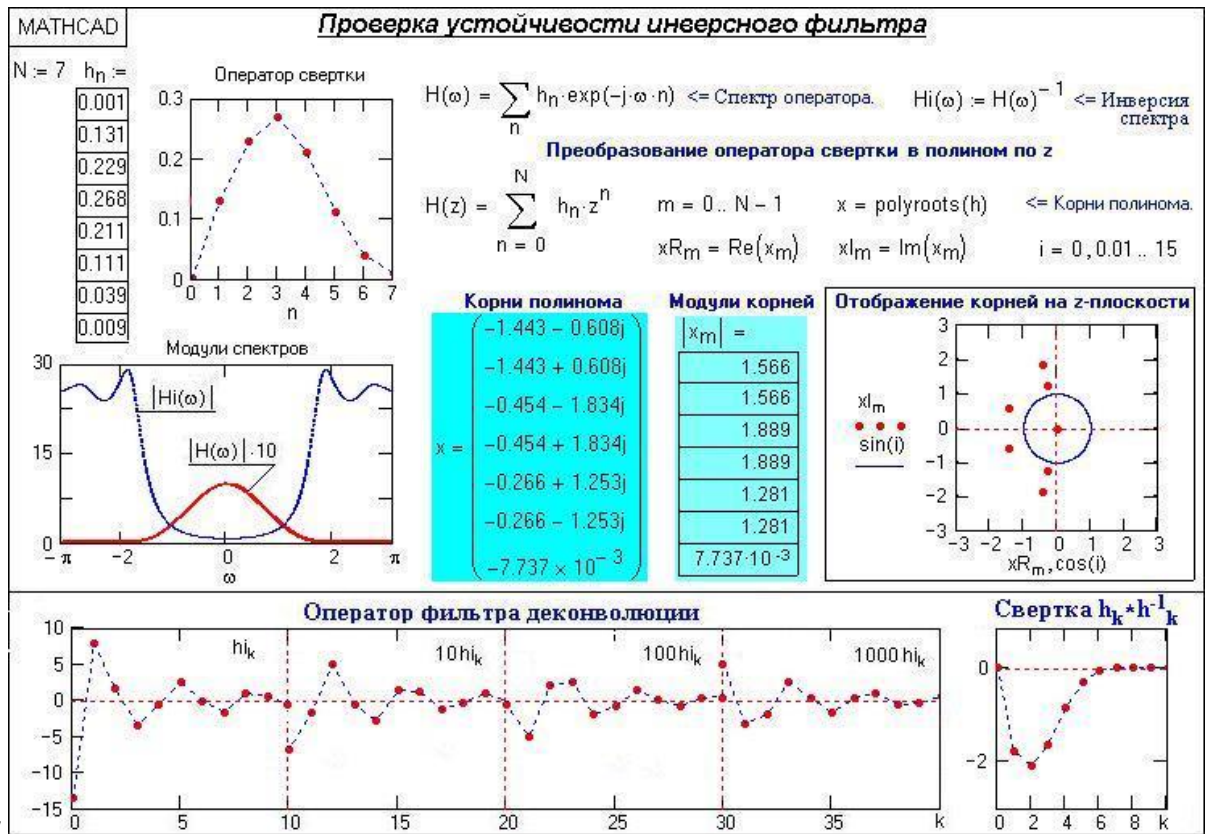


Рис. 13.1.4. Пример неустойчивого инверсного фильтра

Результаты проверки на устойчивость нового оператора деконволюции приведены на рис. 13.1.4. При сходной форме операторов свертки модули спектров операторов практически не отличаются. Но за счет сдвига по фазе данного оператора относительно первого (на рис. 13.1.3) корни его z-полинома претерпели существенное изменение, а модуль одного из корней меньше 1. И хотя вычисленный оператор деконволюции также представляет собой ряд с конечной энергией, но условие (13.1.3) не выполняется, о чем наглядно свидетельствует результат свертки  $h_k * h^{-1}_k$ .

**Обращение недиракоидных функций.** Если  $H(z)$  - реверсоид, т.е. корни составляющих его диполей находятся внутри и на единичном круге в z-плоскости, то устойчивое обращение  $H(z)$  является антиимпульсом (с отрицательными степенями  $z$ ) и для его использования необходимо располагать "будущими" значениями входного сигнала.

Если диполи функции (13.1.4) представляют собой и диракоиды, и реверсоиды, то обращение будет центроидом и определяется полным рядом Лорана:

$$H^{-1}(z) = \dots + h_{-2}z^{-2} + h_{-1}z^{-1} + h_0 + h_1z^1 + h_2z^2 + \dots,$$

т.е. оператор инверсного фильтра является двусторонним и не обязательно симметричным, как мы обычно рассматривали ранее двусторонние операторы.

**Пример 3.** Передаточная функция фильтра:  $H(z) = 1 - 2z$ . Инверсная функция  $H^{-1}(z) = 1/(1 - 2z)$ . Частотные спектры функций приведены на рис. 13.1.5.

Полос функции  $z_p = 1/2$  и находится внутри единичного круга на z-плоскости.

Перепишем выражение для инверсного фильтра в следующем виде:

$$H^{-1}(z) = -(1/2z) [1 + 1/2z + 1/(2z)^2 + \dots].$$

Это выражение является разложением в ряд по степеням  $1/z$  и сходится к кругу радиусом  $1/2$  при  $z \rightarrow \infty$ . Коэффициенты при степенях  $1/z$  являются, соответственно, коэффициентами инверсного фильтра с координатами  $(-n)$ , т.е. фильтр оперирует с "будущими" отсчетами входного сигнала (см. рис. 13.1.5).

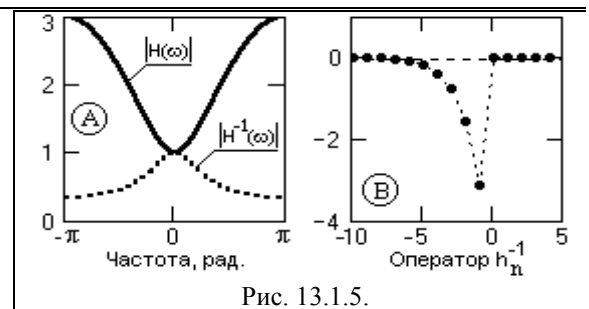


Рис. 13.1.5.

## 13.2. ИНВЕРСИЯ ИМПУЛЬСНОГО ОТКЛИКА ФИЛЬТРА.

**Вычисление коэффициентов инверсного фильтра** по значениям каузального (одностороннего) оператора  $h(n)$  может быть проведено на основе выражения (13.1.3):

$$\sum_{k=0}^{\infty} h^{-1}(k)h(n-k) = \delta_0(n), \quad (13.2.1)$$

для чего достаточно развернуть его в систему  $n$ -уравнений при  $n = 0, 1, 2, \dots, k \leq n$ :

$$\begin{aligned} n = 0: & \quad h^{-1}(0)h(0) = 1, & \quad h^{-1}(0) = 1/h(0). \\ n = 1: & \quad h^{-1}(0)h(1)+h^{-1}(1)h(0) = 0, & \quad h^{-1}(1) = h^{-1}(0)h(1) / h(0). \\ n = 2: & \quad h^{-1}(0)h(2)+h^{-1}(1)h(1)+h^{-1}(2)h(0) = 0, & \quad h^{-1}(2) = (h^{-1}(0)h(2)+h^{-1}(1)h(1))/h(0), \text{ и т.д.} \end{aligned}$$

Продолжая последовательно, можно вычислить любое количество значений коэффициентов инверсного фильтра. Рекуррентная формула вычислений:

$$h^{-1}(n) = (1/h_0) \sum_{k=0}^{n-1} h^{-1}(k)h(n-k). \quad (13.2.2)$$

Если фильтр деконволюции устойчив и ряд  $h^{-1}(n)$  сходится, то появляется возможность ограничения количества членов ряда с определенной ошибкой восстановления исходного сигнала. Метрика приближения  $E$  (квадратичная норма разности) определяется выражением:

$$E^2 = \sum_n [\delta_0(n) - h(n) \otimes h^{-1}(n)]^2. \quad (13.2.3)$$

Ошибка восстановления исходного сигнала появляется со сдвигом на длину оператора фильтра деконволюции и проявляется на интервале длины прямого оператора.

**Пример инверсии оператора фильтра**  $h_n = \{0.219, 0.29, 0.257, 0.153, 0.061, 0.016, 0.003\}$ .

Полином по  $z^n$ :  $H(z) = \sum_n h_n z^n$ .  
 Модули корней полинома:  $z_n = \{2.054, 2.054, 2.485, 2.485, 1.699, 1.699\}$ .  
 Модули корней больше 1, инверсный фильтр устойчив и, судя по значениям корней (достаточно существенно отличаются от 1), быстро затухает.  
 Двенадцать значений оператора по (13.2.2):  $h^{-1}(n) = \{4.56, -6.033, 2.632, 0.417, -0.698, -0.062, 0.267, -0.024, -0.11, 0.051, 0.018, -0.019, 0.004\}$ .

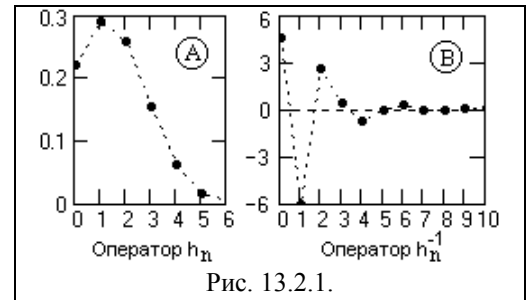


Рис. 13.2.1.

Значения прямого и инверсного оператора фильтра приведены на рис. 13.2.1.

Значения свертки прямого оператора с инверсным при длине  $N=10$  инверсного фильтра и метрика приближения:  $\sigma_n = \{1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.004, 0.006, 0.004, 0.002, <0.001, \dots\}$ .  $E=0.0086$ .

На рис. 13.2.2 приведены графики абсолютных значений концевой ошибки деконволюции при разной длине  $N$  оператора деконволюции.

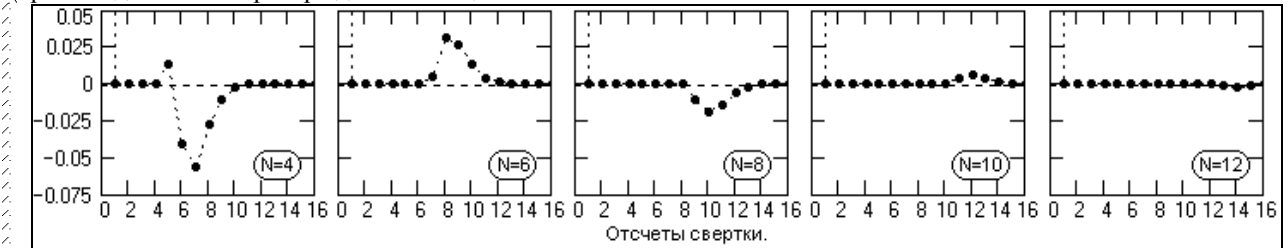


Рис. 13.2.2.

### 13.3. ОПТИМАЛЬНЫЕ ФИЛЬТРЫ ДЕКОНВОЛЮЦИИ

Можно рассчитать оптимальные фильтры деконволюции, метрика приближения которых меньше, чем у усеченных фильтров деконволюции. Для получения общего уравнения оптимальной деконволюции будем считать, что число коэффициентов оператора  $h_n$  равно  $M+1$ , а число коэффициентов инверсного оператора  $h_n^{-1}$  равно  $N+1$ .

**Принцип оптимизации.** Выходная функция приближения при использовании уравнения свертки (13.1.2) с ограничением числа членов оператора фильтра:

$$F = E^2 = \sum_{k=0}^{M+N} [\delta_o(k) - \sigma_k]^2. \quad \sigma_k = \sum_{k=0}^N h_n^{-1} h_{k-n}. \quad (13.3.1)$$

Чтобы определить минимум функции, приравняем нулю частные производные от E по неизвестным коэффициентам фильтра:

$$dF/dh_j^{-1} = \sum_{k=0}^{M+N} -2h_{k-j} [\delta_o(k) - \sum_{k=0}^N h_n^{-1} h_{k-n}] = 0. \quad (13.3.2)$$

$$\sum_{k=0}^{M+N} h_{k-j} \sum_{k=0}^N h_n^{-1} h_{k-n} = \sum_{k=0}^{M+N} h_{k-j} \delta_o(k) = h_{-j}. \quad (13.3.3)$$

$$\sum_{k=0}^N h_n^{-1} \sum_{k=0}^{M+N} h_{k-n} h_{k-j} = \sum_{k=0}^N h_n^{-1} a_{j-n} = h_{-j}, \quad j = 0, 1, 2, \dots, N, \quad (13.3.4)$$

где  $a_{j-n}$  - функция автоковариации импульсной реакции  $h(n)$ . Учитывая также, что  $h_n = 0$  при  $n < 0$  и  $a_j = a_{-j}$  (функция автоковариации является четной функцией), окончательное решение определяется следующей системой линейных уравнений:

$$\begin{array}{cccccc} a_0 h_0^{-1} + & a_1 h_1^{-1} + & a_2 h_2^{-1} + & a_3 h_3^{-1} + \dots & + a_N h_N^{-1} & = h_0 \\ a_1 h_0^{-1} + & a_0 h_1^{-1} + & a_1 h_2^{-1} + & a_2 h_3^{-1} + \dots & + a_{N-1} h_N^{-1} & = 0 \\ a_2 h_0^{-1} + & a_1 h_1^{-1} + & a_0 h_2^{-1} + & a_1 h_3^{-1} + \dots & + a_{N-2} h_N^{-1} & = 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_N h_0^{-1} + & a_{N-1} h_1^{-1} + & a_{N-2} h_2^{-1} + & a_{N-3} h_3^{-1} + \dots & + a_0 h_N^{-1} & = 0 \end{array} \quad (13.3.5)$$

**Пример расчета оптимального фильтра деконволюции.**

Повторим инверсию оператора, приведенного на рис. 13.2.1,  $N=6$ .

Значения оптимального инверсного оператора в сопоставлении с усеченным:

$h^{-1}(n) = \{4.56, -6.033, 2.632, 0.417, -0.698, -0.062, 0.267\}$  – прямой расчет по (13.2.2).

$h^{-1}(n) = \{4.557, -6.026, 2.633, 0.397, -0.693, -0.009, 0.145\}$  – расчет по (13.3.5).

Значения свертки инверсных операторов с прямыми и метрики приближения:

Оператор по (13.2.2) – рис. 13.3.1(A):  $\sigma_n = \{1, 0, 0, 0, 0, 0, 0.005, 0.031, 0.027, 0.013, 0.004, 0.001, 0, \dots\}$ .  $E=0.044$ .

Оператор по (13.3.5) – рис. 13.3.1(B):  $\sigma_n = \{0.999, <0.001, 0.002, -0.003, -0.003, 0.013, -0.008, -0.012, 0.011, 0.013, 0.007, 0.002, <0.001, 0, 0, \dots\}$ .  $E=0.027$ .

Метрика приближения оптимального оператора в 1.6 раза меньше усеченного.

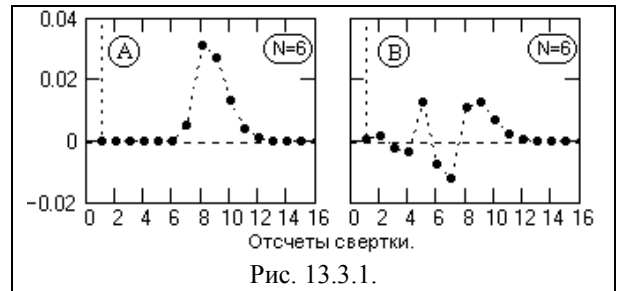


Рис. 13.3.1.

Как видно на рис. 13.3.1, оптимизация инверсного оператора заключается в центрировании ошибок приближения и с распределением по интервалу суммарной длины прямого и инверсного оператора.

**Уравнение оптимальной инверсии.** Оптимальный инверсный фильтр может быть получен непосредственно с использованием  $z$ -образов импульсной реакции и автоковариационной функции прямого фильтра. Если для прямого фильтра мы имеем передаточную функцию  $H(z)$ , то  $z$ -образ автоковариационной функции фильтра (как  $z$ -отображение спектральной плотности мощности) представляет собой произведение:

$$A(z) = H(z)H^*(z), \quad (13.3.6)$$

где  $H^*(z)$ - функция, комплексно сопряженная с  $H(z)$ . Переносим  $H(z)$  в левую часть формулы и для функций диракоидного типа заменяем выражение  $1/H(z) = H^{-1}(z)$ :

$$A(z)H^{-1}(z) = H^*(z). \quad (13.3.7)$$

Запишем последнее равенство в развернутом виде:

$$\begin{aligned} (a_{-N}z^{-N} + \dots + a_{-1}z^{-1} + a_0 + a_1z^1 + \dots + a_Nz^N)(h_0^{-1} + h_1^{-1}z^1 + h_2^{-1}z^2 + \dots + h_N^{-1}z^N) = \\ = h_0^* + h_1^*z^{-1} + h_2^*z^{-2} + \dots + h_N^*z^N. \end{aligned} \quad (13.3.8)$$

В выражении (13.3.8) сумма коэффициентов при одинаковых степенях  $z$  в левой

части равенства должна быть равна коэффициенту при соответствующей степени  $z$  в правой части равенства, что позволяет составить следующую систему из  $N$  уравнений для коэффициентов при степенях  $z^0, z^1, z^2, \dots, z^N$ :

$$\begin{cases} a_0 h_0^{-1} + a_1 h_1^{-1} + a_2 h_2^{-1} + a_3 h_3^{-1} + \dots + a_N h_N^{-1} = h_0^* \\ a_1 h_0^{-1} + a_0 h_1^{-1} + a_1 h_2^{-1} + a_2 h_3^{-1} + \dots + a_{N-1} h_N^{-1} = 0 \\ a_2 h_0^{-1} + a_1 h_1^{-1} + a_0 h_2^{-1} + a_5 h_3^{-1} + \dots + a_{N-2} h_N^{-1} = 0 \\ \dots \\ a_N h_0^{-1} + a_{N-1} h_1^{-1} + a_{N-2} h_2^{-1} + a_{N-3} h_3^{-1} + \dots + a_0 h_N^{-1} = 0 \end{cases} \quad (13.3.9)$$

В случае вещественных фильтров, когда  $a_i = a_i$  и  $h_0^* = h_0$ , уравнение (13.3.9) идентично уравнению (13.3.5).

**Уравнение Левинсона.** Практический способ расчета оптимальных инверсных фильтров по уравнению (13.3.9) предложен в 1947 году Н.Левинсоном.

Перепишем уравнение (13.3.9) в матричной форме:

$$\begin{pmatrix} \mathbf{a}_0 & \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_N \\ \mathbf{a}_1 & \mathbf{a}_0 & \mathbf{a}_1 & \dots & \mathbf{a}_{N-1} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{a}_N & \mathbf{a}_{N-1} & \mathbf{a}_{N-2} & \dots & \mathbf{a}_0 \end{pmatrix} \begin{pmatrix} \mathbf{h}_0^{-1} \\ \mathbf{h}_1^{-1} \\ \dots \\ \mathbf{h}_N^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{h}_0^* \\ \mathbf{0} \\ \dots \\ \mathbf{0} \end{pmatrix} \quad (13.3.10)$$

Так как коэффициенты инверсного фильтра достаточно определить с точностью до произвольного масштабного множителя, приведем  $h_0^{-1}$  к 1, а функцию автоковариации переведем в функцию коэффициентов корреляции делением обеих частей уравнения на  $a_0$ . Обозначая  $A_i = a_i/a_0$ ,  $W_i = h_i^{-1}/h_0^{-1}$  и  $V = h_0^*/(h_0^{-1}a_0) = h_0 h_0^*/a_0$ , получаем:

$$\begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_N \\ \mathbf{A}_1 & \mathbf{A}_0 & \mathbf{A}_1 & \dots & \mathbf{A}_{N-1} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{A}_N & \mathbf{A}_{N-1} & \mathbf{A}_{N-2} & \dots & \mathbf{A}_0 \end{pmatrix} \begin{pmatrix} \mathbf{W}_{0j} \\ \mathbf{W}_{1j} \\ \dots \\ \mathbf{W}_{Nj} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_j \\ \mathbf{0} \\ \dots \\ \mathbf{0} \end{pmatrix} \quad (13.3.11)$$

где для значений  $W$  и  $V$  введен индекс  $j$  номеров предстоящих итераций по циклу вычисления коэффициентов фильтра.

При нулевой итерации ( $N=0, j=0$ ) имеем только одно уравнение:

$$|\mathbf{A}_0| \cdot |\mathbf{W}_{00}| = |\mathbf{V}_0| \quad (13.3.12)$$

Благодаря проведенной нормировке решения уравнения (13.3.12) не требуется:

$$A_0 = 1, V_0 = 1, W_{00} = 1.$$

Составим уравнение для двучленного фильтра ( $N=1, j=1$ ):

$$\begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 \\ \mathbf{A}_1 & \mathbf{A}_0 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{W}_{01} \\ \mathbf{W}_{11} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{0} \end{pmatrix} \quad (13.3.13)$$

Перепишем уравнение (13.3.12) в прямой форме:

$$A_0 W_{00} = V_0. \quad (13.3.14)$$

Запишем вспомогательную систему, для чего к уравнению (13.3.14) добавим вторую строку с новой постоянной  $E_j$ :

$$\begin{aligned} A_0 W_{00} + A_1 \cdot 0 &= V_0, \\ A_1 W_{00} + A_0 \cdot 0 &= E_1. \end{aligned}$$

В матричной форме:

$$\begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 \\ \mathbf{A}_1 & \mathbf{A}_0 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{W}_{00} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_0 \\ \mathbf{E}_1 \end{pmatrix} \quad (13.3.15)$$

Реверсируем уравнение (13.3.15):

$$\begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 \\ \mathbf{A}_1 & \mathbf{A}_0 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{0} \\ \mathbf{W}_{00} \end{pmatrix} = \begin{pmatrix} \mathbf{E}_1 \\ \mathbf{V}_0 \end{pmatrix} \quad (13.3.16)$$

Вычтем (13.3.16) из (13.3.15) с неопределенным множителем  $R_j$ :

$$\begin{pmatrix} \mathbf{A}_0 & \mathbf{A}_1 \\ \mathbf{A}_1 & \mathbf{A}_0 \end{pmatrix} \cdot \left\{ \begin{pmatrix} \mathbf{W}_{00} \\ \mathbf{0} \end{pmatrix} - \mathbf{R}_1 \begin{pmatrix} \mathbf{0} \\ \mathbf{W}_{00} \end{pmatrix} \right\} = \begin{pmatrix} \mathbf{V}_0 \\ \mathbf{E}_1 \end{pmatrix} - \mathbf{R}_1 \begin{pmatrix} \mathbf{E}_1 \\ \mathbf{V}_0 \end{pmatrix} \quad (13.3.17)$$

Из верхней строки уравнения (13.3.16) можно получить значение  $E_1$ :

$$E_1 = A_1 W_{00}. \quad (13.3.18)$$

Уравнение (13.3.13) можно сделать равнозначным уравнению (13.3.17), если правую часть нижней строки уравнения (13.3.17) приравнять к правой части нижней строки уравнения (13.3.13):

$$E_1 - R_1 V_0 = 0, \quad R_1 = E_1 / V_0. \quad (13.3.19)$$

При этом из правых частей верхних строк уравнений (13.3.13, 13.3.17) будем иметь:

$$V_1 = V_0 - R_1 E_1. \quad (13.3.20)$$

Приравнивая друг другу левые части уравнений (13.3.13, 13.3.17), получаем:

$$W_{01} = W_{00} - R_1 \cdot 0 = W_{00} = 1.$$

$$W_{11} = 0 - R_1 W_{00} = -R_1 W_{00}. \quad (13.3.21)$$

Этим заканчивается первая итерация. Аналогично, для второй итерации:

$$\begin{bmatrix} A_0 & A_1 & A_2 \\ A_1 & A_0 & A_1 \\ A_2 & A_1 & A_0 \end{bmatrix} \cdot \begin{bmatrix} W_{02} \\ W_{12} \\ W_{22} \end{bmatrix} = \begin{bmatrix} V_2 \\ 0 \\ 0 \end{bmatrix} \quad (13.3.22)$$

$$\begin{bmatrix} A_0 & A_1 & A_2 \\ A_1 & A_0 & A_1 \\ A_2 & A_1 & A_0 \end{bmatrix} \cdot \begin{bmatrix} W_{01} \\ W_{11} \\ 0 \end{bmatrix} = \begin{bmatrix} V_1 \\ 0 \\ E_2 \end{bmatrix} \quad (13.3.23)$$

$$\begin{bmatrix} A_0 & A_1 & A_2 \\ A_1 & A_0 & A_1 \\ A_2 & A_1 & A_0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ W_{11} \\ W_{01} \end{bmatrix} = \begin{bmatrix} E_2 \\ 0 \\ V_1 \end{bmatrix} \quad (13.3.24)$$

$$\begin{bmatrix} A_0 & A_1 & A_2 \\ A_1 & A_0 & A_1 \\ A_2 & A_1 & A_0 \end{bmatrix} \cdot \left\{ \begin{bmatrix} W_{01} \\ W_{11} \\ 0 \end{bmatrix} - R_2 \begin{bmatrix} 0 \\ W_{11} \\ W_{01} \end{bmatrix} \right\} = \begin{bmatrix} V_1 \\ 0 \\ E_2 \end{bmatrix} - R_2 \begin{bmatrix} E_2 \\ 0 \\ V_1 \end{bmatrix} \quad (13.3.25)$$

Из верхней строки уравнения (13.3.24):

$$E_2 = A_1 W_{11} + A_2 W_{01}.$$

Из правых частей нижней и верхней строк уравнений (13.3.22, 13.3.25):

$$R_2 = E_2 / V_1,$$

$$V_2 = V_1 - R_2 E_2.$$

Новые коэффициенты из левых частей уравнений (13.3.22, 13.3.25):

$$W_{02} = W_{01} - R_2 \cdot 0 = 1,$$

$$W_{12} = W_{11} - R_2 W_{11},$$

$$W_{22} = 0 - R_2 W_{01}.$$

Анализ расчетов позволяет вывести следующие рекуррентные формулы:

$$E_j = \sum_{i=0}^j A_i W_{j-i,j}, \quad j = 1, 2, \dots, M. \quad (13.3.26)$$

$$R_j = E_j / V_{j-1},$$

$$V_j = V_{j-1} - R_j E_j,$$

$$W_{i,j} = W_{i,j-1} - R_j W_{j-1,j-1}, \quad i = 0, 1, \dots, j.$$

Подпрограммы решения уравнений для ЭВМ приведены в литературе /12,22/.

#### 13.4. РЕКУРСИВНАЯ ДЕКОНВОЛЮЦИЯ

**Уравнение фильтра рекурсивной деконволюции.** Запишем уравнение (13.1.2) для инверсного фильтра в развернутой форме:

$$H^{-1}(z) = 1/(h_0 + h_1 z + h_2 z^2 + \dots). \quad (13.4.1)$$

Так как для минимально-фазового оператора всегда выполняется условие  $h_0 \neq 0$ , приведем (13.4.1) к виду:

$$H^{-1}(z) = (1/h_0)/(1 + h_1 z/h_0 + h_2 z^2/h_0 + \dots) = G/(1 + q_1 z + q_2 z^2 + \dots), \quad (13.4.2)$$

где:  $G = 1/h_0$ ,  $q_1 = h_1/h_0$ ,  $q_2 = h_2/h_0$  и т.д. Но уравнение (13.4.2) есть не что иное, как уравнение передаточной функции рекурсивного фильтра, где цепь обратной связи образована коэффициентами нормированного оператора  $h(n)$ . Алгоритм вычислений:

$$y_k = G \cdot x_k - \sum_{n=1}^N q_n \cdot y_{k-n}.$$

Выражение (13.4.2) уникально по своим возможностям. В принципе, оно может реализовать оператор инверсной фильтрации с бесконечным импульсным откликом. На практике оно может использоваться вместо медленно затухающих инверсных операторов, модуль одного из полюсов которого очень близок к 1 (менее 1.1) при высоких требованиях к метрике приближения.

**Пример рекурсивной деконволюции.**

Оператор  $h_n = \{0.41, 0.791, 0.401, -0.193, -0.367, -0.166, 0.032, 0.068, 0.027, -0.001\}$ ,  $N=9$ .

1. Модуль одного из корней фильтра равен 1.032, что приводит к очень слабому затуханию инверсного оператора. Метрика приближения даже при  $N=100$  для усеченного оператора составляет 0.3. Форма операторов приведена на рис. 13.4.1.

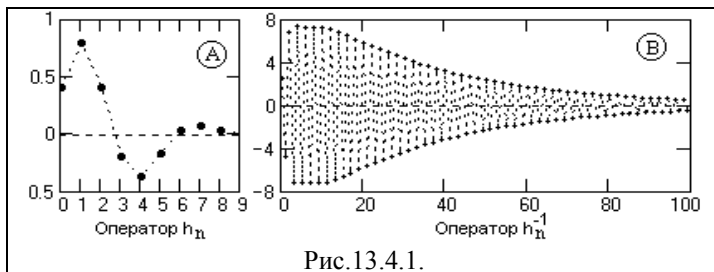


Рис.13.4.1.

2. При использовании оптимального инверсного оператора с  $N=100$  значение погрешности приближения уменьшается более чем в 20 раз, что позволяет уменьшить длину оператора до  $N=35$  при погрешности приближения порядка 0.1 (рис 13.4.2(A)), при этом абсолютные значения погрешностей приближения не превышают 0.03 (рис. 13.4.2(B)).

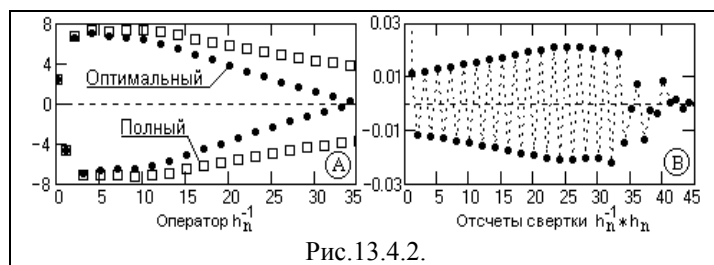


Рис.13.4.2.

3. Расчет коэффициентов фильтра рекурсивной деконволюции:

$$G = 1/h_0 = 2.441$$

$$g_n = h_n \cdot G \quad g_n = \{1.932, 0.978, -0.472, -0.896, -0.405, 0.077, 0.165, 0.065, -0.003\}, \quad n=1, \dots, 9.$$

На рис. 13.4.3 приведен результат рекурсивной деконволюции оператора  $h_n$ . Деконволюция абсолютно точно, с нулевой метрикой, восстанавливает импульс Кронекера, хотя собственный импульсный отклик рекурсивного оператора повторяет оператор  $h_n^{-1}$  при его вычислении по формуле (13.2.2) и длительность его значимой части близка к 200. Коэффициент усиления дисперсии шумов при данной операции вычисляется по значениям импульсного отклика оператора рекурсивной деконволюции и весьма существенен, как и для всех инверсных операторов.



Рис. 13.4.3.

**13.5. ФИЛЬТРЫ СЖАТИЯ СИГНАЛОВ**

Рассмотренные выше методы относятся к расчету так называемых идеальных инверсных фильтров, т.е. фильтров полной инверсии системных операторов. Однако использование идеальных инверсных фильтров на практике не всегда возможно, т.к. регистрируемые данные обычно осложнены влиянием помех (шумов), а инверсные фильтры обычно имеют коэффициент усиления дисперсии шумов значительно больше 1. В этом случае задача точной деконволюции (восстановления истинной первоначальной формы сигнала), как правило, не ставится, а инверсные фильтры считаются оптимальными с точки зрения максимального приближения к форме полезного сигнала с определенным допустимым коэффициентом усиления дисперсии помех. Такие фильтры называются фильтрами неполной (частичной, ограниченной) деконволюции или фильтрами сжатия сигнала. При проектировании фильтров неполной деконволюции учитываются статистические характеристики помех во входном сигнале и их соотношение со статистическими характеристиками самого входного сигнала.

**Передаточная функция фильтра** неполной деконволюции с учетом помех во входном сигнале определяется выражением:

$$H^{-1}(z) = H^*(z) / [ |H(z)|^2 + g^2 ], \tag{13.5.1}$$

где  $g^2 = k \cdot \sigma_h^2$  - дисперсия шумов в единицах дисперсии оператора  $h_n$ ,  $\sigma_h^2$  - дисперсия

значений оператора  $h_n$ , (при условии суммы значений оператора, равной 1),  $k$  - отношение дисперсии шумов к дисперсии оператора  $h_n$ . Коэффициент  $g^2$  играет роль регуляризирующего фактора при выполнении операции деконволюции информации.

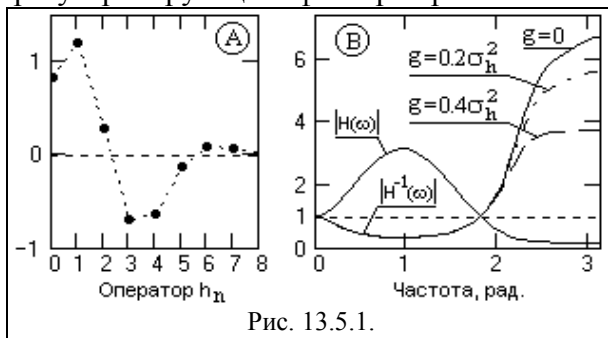


Рис. 13.5.1.

На рис. 13.5.1 пример формы оператора  $h_n$  и спектральных функций (13.5.1) при разных значениях параметра  $g$ . При  $g = 0$  выражение (13.5.1) обращается в идеальный инверсный фильтр  $1/H(z)$ . Во втором крайнем случае, при  $g^2 \gg |H(z)|^2$ , фильтр (13.5.1) переходит в фильтр, согласованный с сигналом по частотному спектру:  $H^{-1}(z) = H^*(z)/g^2$ , который только максимизирует отношение сигнал/помеха.

На рис. 13.5.2 приведена форма инверсных операторов, соответствующая их частотным характеристикам на рис. 13.5.1(B), и результаты свертки инверсных операторов с прямым (для лучшего просмотра графики прямой оператор при свертке сдвинут вправо на 2 значения  $\Delta t$ ). При  $g=0$  коэффициент усиления дисперсии шумов равен 11, при  $g=0.4\sigma_h^2$  равен 4.6. Однако снижение усиления дисперсии шумов сопровождается увеличением погрешности приближения, что можно видеть на рис. 13.5.2(B), при этом уменьшается амплитуда восстановления импульса Кронекера и появляются осцилляции после импульса. Но при наличии шумов и правильном выборе параметра  $g$  общее отношение амплитудных значений сигнал/ шум для оператора по (13.5.1) больше, чем для прямой инверсии по (13.1.2), что объясняется более существенным уменьшением коэффициента усиления дисперсии шумов при увеличении параметра  $g$ , чем увеличением погрешности приближения.

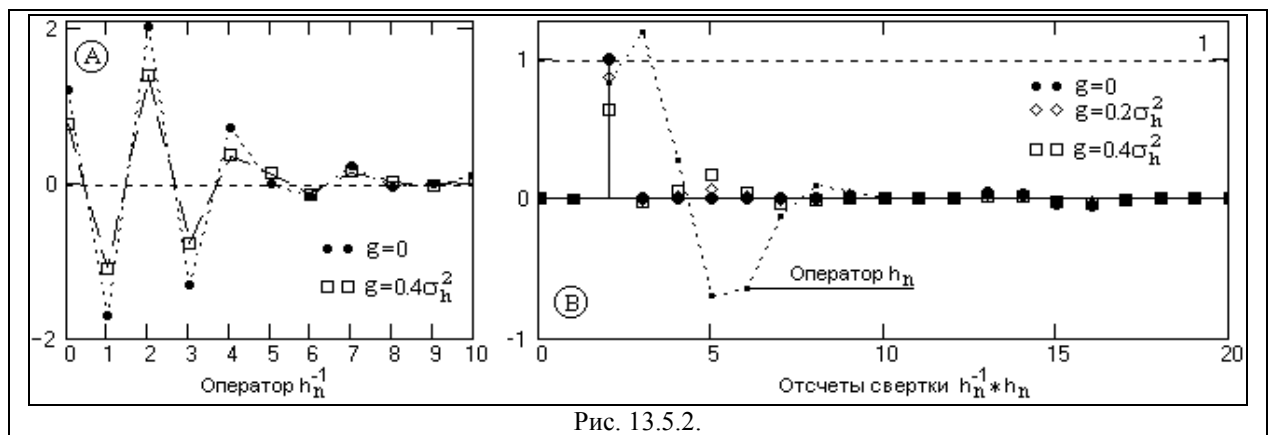


Рис. 13.5.2.

**Операторы оптимальных фильтров** сжатия сигналов также могут вычисляться с учетом помех. Если сигнал  $s(k)$  и помеха статистически независимы, то функция автоковариации сигнала на входе фильтра:

$$a_i = a_{si} + b_i, \quad (13.5.2)$$

где  $a_{si}$  и  $b_i$  - функции автоковариации сигнала и помех. При помехе типа белого шума функция автоковариации помех представляет собой весовую дельта-функцию в точке 0:

$$b_i = c^2 \delta_i, \quad (13.5.3)$$

где  $c^2$  - дисперсия помех. С учетом этого фактора расчет оптимальных инверсных фильтров может проводиться по вышеприведенным формулам (13.3.5, 13.3.9) с изменением значения коэффициента  $a_0$ :

$$a_0 = a_0 + c^2. \quad (13.5.4)$$

На рис. 13.5.3(A) приведены примеры операторов оптимальных инверсных фильтров, вычисленные по прямому оператору, приведенному на рис. 13.5.1(A). Значения

коэффициента  $c^2$  заданы в долях дисперсии прямого оператора. Ввод коэффициента  $c^2$  в функцию автоковариации резко уменьшает значения коэффициентов инверсного оператора и, соответственно, уменьшает коэффициент усиления дисперсии помех.

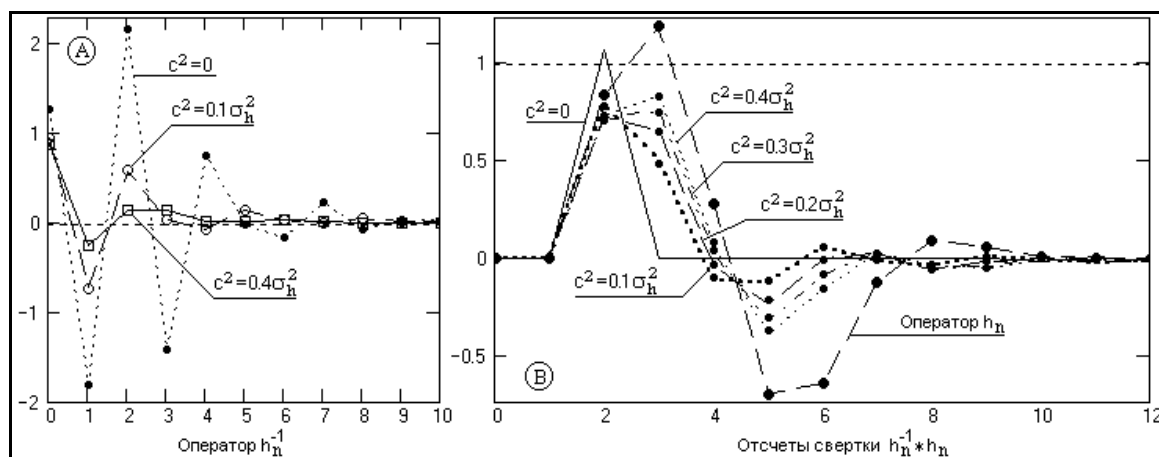


Рис. 13.5.3.

Для приведенного примера, при исходном значении коэффициента усиления дисперсии шумов порядка 12 для  $c^2=0$ , его значение уменьшается до 1.8 при  $c^2=0.1\sigma^2$  и становится меньше 1 при  $c^2 > 0.3\sigma^2$ . Естественно, что общая погрешность приближения деконволюции при этом также существенно изменяется (см. рис. 13.5.3(B)), но амплитуда значения сигнала на месте импульса Кронекера (там, где он должен быть) изменяется много меньше, чем коэффициент усиления дисперсии шумов, а, следовательно, отношение сигнал/шум при введении коэффициента  $c^2$  существенно увеличивается.

#### Тема 14. АППРОКСИМАЦИЯ СИГНАЛОВ И ФУНКЦИЙ

Математика часто оперирует с математическими функциями решения дифференциальных уравнений и интегралов, которые не имеют аналитических выражений и представляются табличными числовыми значениями  $y_i$  для дискретных значений независимых переменных  $x_i$  вместо аналитических выражений в виде  $y=s(x)$ . Это означает, что любому значению  $x$  из области определения поставлено в соответствии определенное числовое значение  $y$ . Аналогичными таблицами  $\{y_i, x_i\}$  могут представляться и экспериментальные данные, для которых неизвестна явная связь между  $y$  и  $x$  или эта связь только подлежит выяснению. Точки, в которых определены числовые значения функций или данных, называются *узловыми*.

Однако на практике могут понадобиться значения данных величин совсем в других точках, отличных от узловых, или с другим шагом дискретизации аргументов. Возникающая задача вычисления значений функции в промежутках между узлами называется задачей *интерполяции*, за пределами семейства узловых точек вперед или назад по переменным – задачей *экстраполяции* или *прогнозирования*.

Даже при известных зависимостях  $y=s(x)$  формулы этих зависимостей, детально и точно описывающие определенные физические объекты и процессы, могут быть очень сложными и мало пригодными для практического использования как при математическом анализе физических данных, так и в прикладных задачах, особенно при расчетах ожидаемых результатов измерений и при математическом моделировании физических процессов. Кроме того, практическая регистрация физических данных выполняется, как правило, с определенной погрешностью или с определенным уровнем шумов, которые по своим значениям могут быть выше теоретической погрешности прогнозирования сигналов при расчетах по сложным, хотя и очень точным формулам. Не имеет смысла и проектирование систем обработки и анализа сигналов по высокоточным формулам, если повышение точности расчетов не дает эффекта в повышении точности обработки данных.



Во всех этих условиях возникает *задача аппроксимации*. Аппроксимация, это представление сложных функций  $s(x)$  или дискретных выборок из этих функций  $s(x_i)$  простыми и удобными для практического использования функциями аппроксимации  $f(x)$  таким образом, чтобы отклонение  $f(x)$  от  $s(x)$  в области ее задания было наименьшим по определенному критерию приближения.

Если приближение строится на заданном множестве точек  $\{x_i\}$ , то аппроксимация называется *точечной*. К ней относятся интерполирование, среднеквадратичное приближение и др. При построении приближения на непрерывном множестве точек, например на отрезке  $[a,b]$ , аппроксимация называется непрерывной или интегральной.

Сглаживание статистических данных или аппроксимация данных с учетом их статистических параметров относится к задачам *регрессии*. Как правило, при регрессионном анализе усреднение данных производится *методом наименьших квадратов* (МНК).

Классические математические методы аппроксимации, интерполяции, экстраполяции и регрессии функций имеют многовековую историю. В рамках настоящего курса мы не будем углубляться в строгую математическую теорию этих операций. Все современные математические системы (Mathcad, Matlab, Maple и пр.) имеют в своем составе универсальный аппарат выполнения таких операций, дающий пользователю возможность реализации любых практических задач по обработке данных. В качестве основной математической системы для примеров будем использовать систему Mathcad.

#### 14.1. ПРИБЛИЖЕНИЕ СИГНАЛОВ РЯДАМИ ТЕЙЛОРА [39]

Разложение функций в ряд Тейлора явилось одним из первых методов приближения функций в окрестностях точек  $x_0$ :

$$f(x) \cong f(x_0) + \frac{f'(x_0)}{1!} (x-x_0) + \frac{f''(x_0)}{2!} (x-x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!} (x-x_0)^n.$$

$$f(x) \cong f(x_0) + \sum_{i=1}^n \frac{f^{(i)}(x_0)}{i!} (x-x_0)^i.$$

При разложении функции в окрестностях точки  $x_0=0$  ряд Тейлора принято называть рядом Маклорена.

Первый член ряда  $f(x_0)$  представляет собой отсчет функции в точке  $x_0$  и грубое приближение к значениям функции в окрестностях этой точки. Остальные члены ряда детализируют значения функции в окрестностях точки  $x_0$  по информации в соседних точках и тем точнее приближают сумму ряда к значениям функции, чем больше членов суммы участвуют в приближении, с расширением интервала окрестностей точного приближения. Наглядно это можно видеть на примере двух функций, приведенном на рис. 14.1.1 (с усечением отображения членов рядов  $f_2(x)$  и  $f_4(x)$ ).

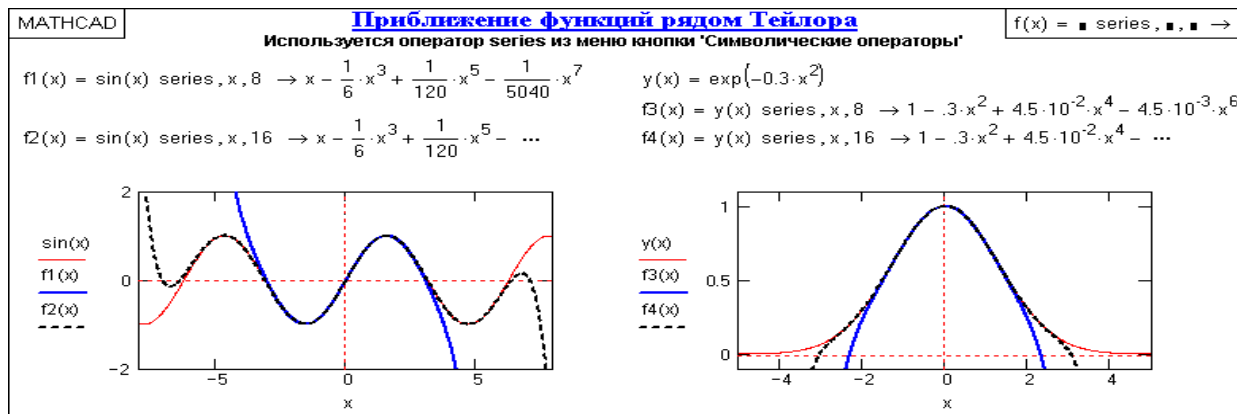


Рис. 14.1.1. Примеры разложения функций в ряд Маклорена.

Приближение функций рядом Тейлора применяется, в основном, для непрерывных гладких функций в локальных интервалах задания. Для разрывных и периодически

повторяющихся функций использовать его практически невозможно, равно как и для непрерывных не дифференцируемых функций. Операция дифференцирования тоже может быть далеко не простой, а получаемые ряды могут сходиться очень медленно.

## 14.2. ИНТЕРПОЛЯЦИЯ И ЭКСТРАПОЛЯЦИЯ СИГНАЛОВ [39, 55].

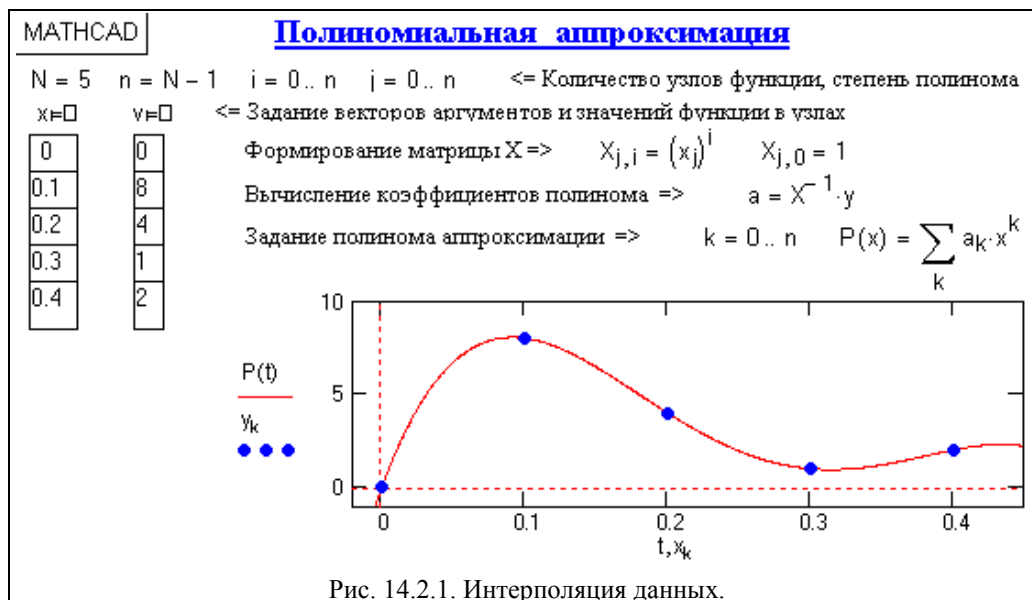
**Полиномиальная интерполяция.** Одним из основных типов точечной аппроксимации является интерполирование. Сущность его заключается в том, что функции  $y_i = s(x_i)$  сопоставляется интерполяционный многочлен

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n = \sum_{i=0}^n a_i \cdot x^i, \quad (14.2.1)$$

принимаящий в точках  $x_i$  те же значения  $y_i$ , что и функция  $f(x)$ . Для выполнения полиномиальной интерполяции достаточно по выражению (14.2.1) составить систему линейных уравнений для  $n$  узловых точек и определить  $n$  значений коэффициентов  $a_i$ .

При  $N$  точках функции  $y_i$  максимальная степень интерполяционного многочлена  $n=N-1$ , и в этом случае говорят о глобальной интерполяции с прохождением  $f(x)$  через все значения точек  $y_i$ . Однако в этом случае при большом количестве узлов получается очень высокая степень многочлена. Кроме того, экспериментальные табличные данные могут содержать ошибки измерений, а глобальная интерполяция повторит все допущенные при измерениях ошибки. Для исключения этого фактора стараются подобрать аппроксимирующий многочлен как можно меньшей степени (как правило,  $n=1, 2, 3$ ), график которого проходит близко от узловых точек. На практике мерой отклонения многочлена  $f(x)$  от заданной функции на множестве точек  $(x_i, y_i)$  является величина  $\sigma$  среднеквадратичного приближения  $\sigma^2 = \sum_i (f(x_i) - y_i)^2$ , минимальное значение которой обеспечивается подбором коэффициентов  $a_i$ .

Пример выполнения глобальной интерполяции приведен на рис.14.2.1. Равномерной дискретизации данных для интерполяции не требуется. Максимальная степень полинома на практике обычно устанавливается не более 8-10, а большие массивы данных интерполируются последовательными *локальными* частями.



Линейная и квадратичная интерполяция являются самыми простыми способами обработки таблиц и выполняются по уравнениям:  
 $f(x)_{\text{лин}} = a_0 + a_1x$ .    $f(x)_{\text{кв}} = a_0 + a_1x + a_2x^2$ .

При линейной интерполяции вычисления дополнительных точек выполняются по линейной зависимости. Графически это означает простое соединение узловых точек отрезками прямых. В системе Mathcad для этого используется функция  $\text{linterp}(X, Y, x)$ , где  $X$  и  $Y$  – вектора узловых точек. Функция  $\text{linterp}(X, Y, x)$  возвращает значение функции при её линейной интерполяции по заданным аргументам  $x$ . При небольшом числе узловых точек

(менее 10) линейная интерполяция оказывается довольно грубой. Для целей экстраполяции функция `linterp` не предназначена.

Для практического использования более удобны формулы аппроксимации, не требующие предварительного определения коэффициентов аппроксимирующих полиномов. К числу таких формул относится интерполяционный многочлен по Лагранжу /40/. При аппроксимации функции  $y(x)$  многочленом  $n$ -й степени  $Y(x)$ :

$$Y(x) = \frac{(x-x_1)(x-x_2)\dots(x-x_n)}{(x_0-x_1)(x_0-x_2)\dots(x_0-x_n)}y_0 + \frac{(x-x_0)(x-x_2)\dots(x-x_n)}{(x_1-x_0)(x_1-x_2)\dots(x_1-x_n)}y_1 + \dots + \frac{(x-x_0)(x-x_1)\dots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\dots(x_n-x_{n-1})}y_n. \quad (14.2.2)$$

Пример интерполяции по Лагранжу приведен на рис. 14.2.2.



Рис. 14.2.2. Интерполяция по Лагранжу.

**Кривые Безье.** Для задач аппроксимации широко применяются кривые Безье. Это связано с их удобством как для аналитического описания, так и для наглядного геометрического построения. При использовании кривых Безье в компьютерной

графике пользователь может задавать форму кривой интерактивно, двигая опорные точки курсором на экране.

Метод построения кривых был предложен де Кастелье в 1959 году и основан на разбиении отрезков, соединяющих исходные точки в отношении  $t$ , а затем в рекурсивном повторении этого процесса для полученных отрезков:

$$P_i^j(t) = (1-t)P_i^{j-1}(t) + tP_{i+1}^{j-1}(t), \quad (14.2.3)$$

где нижний индекс - номер точки, верхний индекс - уровень разбиения. Уравнение кривой  $n$ -ого порядка задается уравнением:  $P(t) = P_0^n(t)$ .

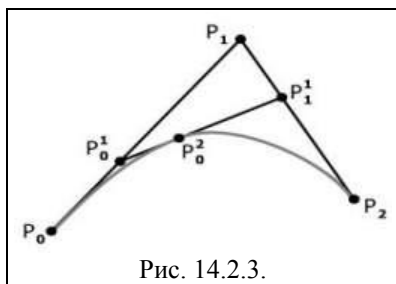


Рис. 14.2.3.

Для примера построим кривую для трех опорных точек  $P_0(t=0)$ ,  $P_1$  и  $P_2(t=1)$  на интервале  $t=[0, 1]$  (рис. 14.2.3).

Для каждого  $t \in [0, 1]$  определим точку  $P_0^2(t)$ :

$$P_0^1(t) = (1-t)P_0 + tP_1, \quad P_1^1(t) = (1-t)P_1 + tP_2,$$

$$P_0^2(t) = (1-t)P_0^1(t) + tP_1^1(t) = (1-t)^2P_0(t) + 2t(1-t)P_1(t) + t^2P_2(t),$$

и тем самым получим кривую второго порядка.

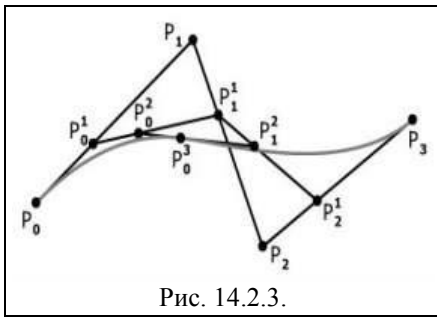


Рис. 14.2.3.

Аналогичным образом построение кривой Безье с четырьмя опорными точками будет определяться следующими выражениями:

$$P_0^1(t) = (1-t)P_0 + tP_1, \quad P_1^1(t) = (1-t)P_1 + tP_2, \quad P_2^1(t) = (1-t)P_2 + tP_3,$$

$$P_0^2(t) = (1-t)P_0^1(t) + tP_1^1(t) = (1-t)^2P_0(t) + 2t(1-t)P_1(t) + t^2P_2(t),$$

$$P_1^2(t) = (1-t)P_1^1(t) + tP_2^1(t) = (1-t)^2P_1(t) + 2t(1-t)P_2(t) + t^2P_3(t),$$

$$P_2^3(t) = (1-t)P_0^2(t) + tP_1^2(t) = (1-t)^3P_0(t) + 3t(1-t)^2P_1(t) + 3t^2(1-t)P_2(t) + t^3P_3(t).$$

$$P_0^3(t) = (1-t)P_0^2(t) + tP_1^2(t) = (1-t)^3P_0(t) + 3t(1-t)^2P_1(t) + 3t^2(1-t)P_2(t) + t^3P_3(t).$$

Общая аналитическая запись для кривых Безье по N+1 опорной точке:

$$P^N(t) = \sum_{i=0}^N P_i B_i^N(t), \quad B_i^N(t) = C_i^N t^i (1-t)^{N-i},$$

$$C_i^N = N! / (i! (N-i)!) - \text{биномиальные коэффициенты.}$$

Кривые Безье всегда проходят через начальную  $P_0$  и конечную  $P_N$  точки. Если рассматривать опорные точки в противоположном порядке, то форма кривой не изменяется. Если опорные точки лежат на одной прямой, то кривая Безье вырождается в отрезок, соединяющий эти точки. Степень многочлена, представляющего кривую в аналитическом виде, на 1 меньше числа точек.

### 14.3. СПЛАЙНОВАЯ ИНТЕРПОЛЯЦИЯ

*Сплайн* - кусочный многочлен степени  $K$  с непрерывной производной степени  $K - 1$  в точках соединения сегментов.

При сплайновой интерполяции обычно используются локальные полиномы не выше третьей степени. Так, например, *кубические сплайны* проходят через три смежные узловые точки (текущие опорные точки вычислений), при этом в граничных точках совпадают как значения полинома и функции, так и значения их первых и вторых производных. Коэффициенты полиномов, проходящих через три узловые точки, рассчитываются так, чтобы непрерывными были их первая и вторая производные. Линия, которую описывает сплайн-функция, напоминает по форме гибкую линейку, закреплённую в узловых точках. Это создает высокую плавность сплайнового полинома по сравнению с другими методами аппроксимации и наглядно видно на рис. 14.3.1. Полиномы более высоких порядков чрезмерно громоздки для практики.



Рис. 14.3.1. Сплайновая интерполяция и интерполяция по Лагранжу.

Сплайновая аппроксимация может применяться для достаточно быстро изменяющихся функций, не имеющих разрывов функции и ее производных. Основным недостатком сплайнов – отсутствие единого аналитического выражения для описания функции. Заметим также, что результаты экстраполяции функций, как это можно видеть на рис. 14.3.1, существенно зависят от метода аппроксимации, и, соответственно, к их достоверности нужно подходить достаточно осторожно.

Сплайновая интерполяция обычно применяется в составе математических пакетов по определенной технологии. Так, в системе Mathcad при выполнении сплайновой интерполяции по узловым точкам функции (векторам X и Y) сначала вычисляется вектор (обозначим его индексом S) вторых производных входной функции  $y(x)$  по одной из программ:

$S := \text{cspline}(X, Y)$  – возвращает вектор S вторых производных при приближении в опорных точках к кубическому полиному;

$S := \text{pspline}(X, Y)$  – возвращает вектор S при приближении в опорных точках к параболической кривой;

$S := \text{lspline}(X, Y)$  – возвращает вектор S при приближении в опорных точках к прямой.

По значениям вектора S функцией  $\text{interp}(S, X, Y, x)$  вычисляются значения аппроксимирующей функции по аргументам x.

На рис. 14.3.2 приведен пример кубической сплайновой интерполяции двумерных цифровых данных с одновременным повышением узловой сетки цифровых данных в 4 раза.

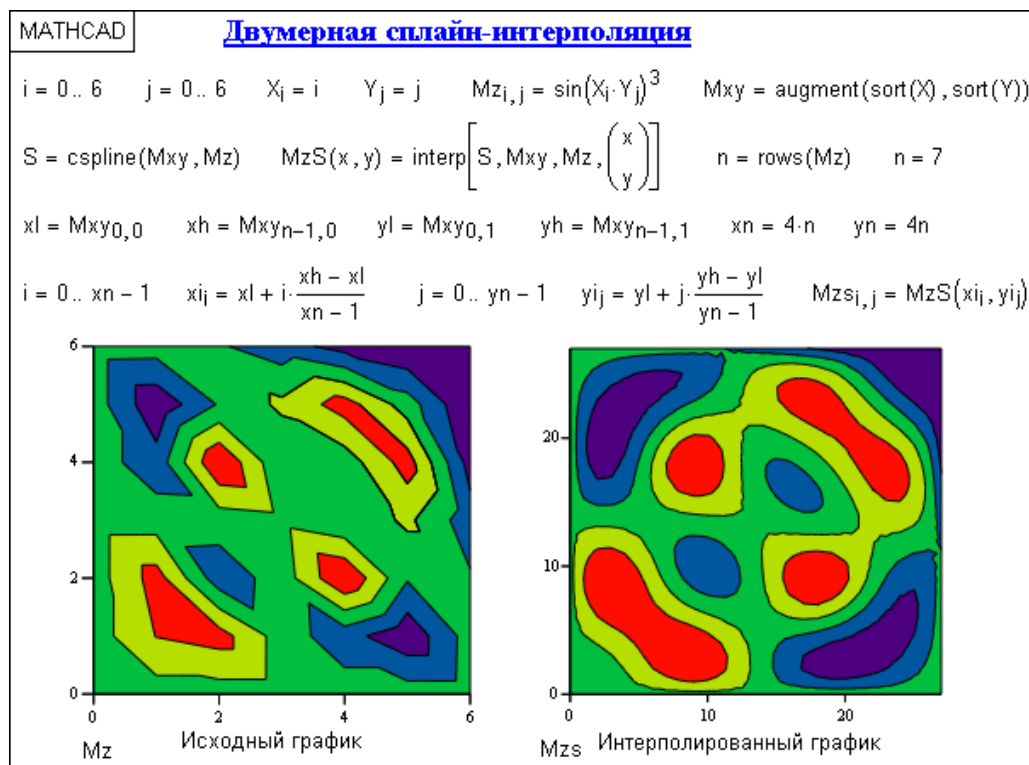


Рис. 14.3.2. Сплайн - интерполяция двумерных данных.

#### 14.4. СПЕКТРАЛЬНЫЙ МЕТОД ИНТЕРПОЛЯЦИИ

При дискретизации данных с равномерным шагом по аргументу наиболее точную интерполяцию финитных сигналов обеспечивает спектральный метод. При условии, естественно, что в спектре сигнала не содержится частотных составляющих, превышающих частоту Найквиста.

**Спектр дискретного сигнала.** Допустим, что для обработки задается произвольный аналоговый сигнал  $s(t)$ , имеющий фурье-образ  $S(f)$ . Равномерная дискретизация непрерывного сигнала  $s(t)$  с частотой  $F$  (шаг  $\Delta t = 1/F = \theta$ ) с математических

позиций означает умножение функции  $s(t)$  на гребневую (решетчатую) функцию  $\text{Ш}_\theta(t) = \sum_k \delta(t-k\Delta t)$ :

$$s_\theta(t) = s(t) \text{Ш}_\theta(t) = s(t) \sum_{k=-\infty}^{\infty} \delta(t-k\Delta t) = \sum_{k=-\infty}^{\infty} s(k\Delta t) \delta(t-k\Delta t). \quad (14.4.1)$$

С учетом известного преобразования Фурье гребневой функции  $\text{Ш}_\theta(t) \Leftrightarrow F \cdot \text{Ш}_F(f)$  фурье-образ дискретной функции  $s_\theta(t)$ :

$$S_F(f) = S(f) \otimes F \cdot \text{Ш}_F(f). \quad (14.4.2)$$

$$\text{Ш}_F(f) = \sum_{n=-\infty}^{\infty} \delta(f-nF). \quad (14.4.3)$$

Отсюда, для спектра дискретного сигнала имеем:

$$S_F(f) = F \cdot S(f) \otimes \sum_{n=-\infty}^{\infty} \delta(f-nF) = F \sum_{n=-\infty}^{\infty} S(f-nF). \quad (14.4.4)$$

Спектр дискретного сигнала представляет собой *непрерывную периодическую функцию* с периодом  $F$ , совпадающую с функцией  $F \cdot S(f)$  непрерывного сигнала  $s(t)$  в пределах центрального периода от  $-f_N$  до  $f_N$ , где  $f_N = 1/2\Delta t = F/2$  - частота Найквиста. Частота дискретизации сигнала должна быть минимум в два раза выше максимальной частотной составляющей в спектре сигнала ( $F = 1/\Delta t \geq 2f_{\max}$ ). Умножая функцию (14.4.2) на прямоугольную весовую функцию  $\Pi_F(f)$ , равную 1 в пределах главного частотного, получаем непрерывный спектр в бесконечных по частоте границах, равный спектру  $F \cdot S(f)$  в пределах главного частотного диапазона:

$$F \cdot S(f) = F \cdot [S(f) \otimes \text{Ш}_F(f)] \Pi_F(f). \quad (14.4.5)$$

Обратное преобразование Фурье этого спектра, с учетом коэффициента  $F$ , должно восстанавливать непрерывный сигнал, равный исходному аналоговому сигналу  $s(t)$ .

На рис. 14.4.1 приведен пример интерполяции и экстраполяции равномерных по аргументу дискретных данных в сравнении с сплайн-методом и методом по Лагранжу. Исходная аналоговая кривая дискретизирована корректно ( $f_{\max} < 1/2\Delta t$ ) и восстановленная по дискретным данным кривая  $fS(z)$  полностью ее повторяет. Близкие результаты к исходному сигналу дает также и сплайн-интерполяция, но доверять сплайн-экстраполяции, особенно по концевой части интервала задания данного сигнала, не приходится. Что касается интерполяции по Лагранжу, то можно видеть существенную погрешность интерполяции на концевых частях интервала сигнала и полную непригодность для задачи экстраполяции.

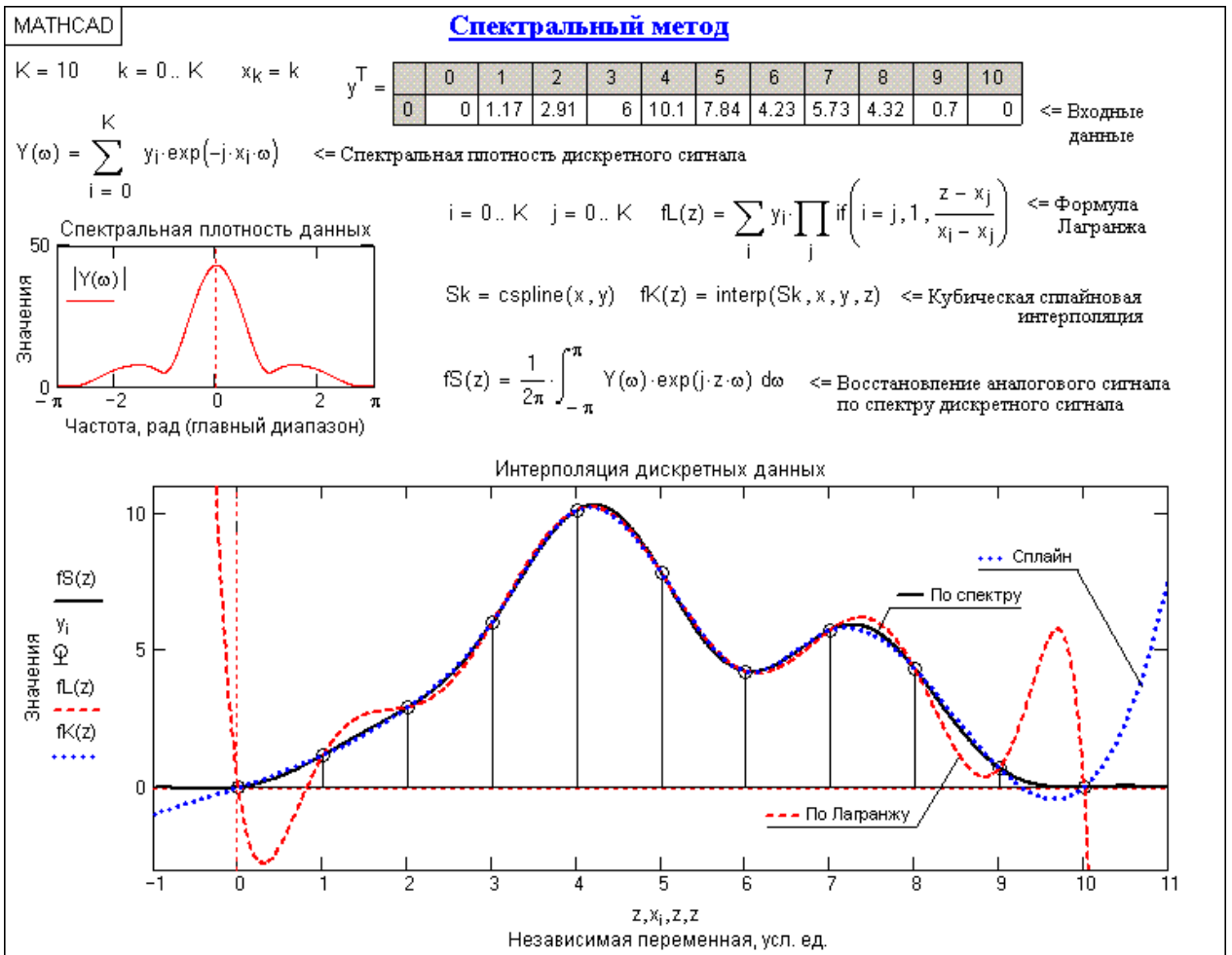


Рис. 14.4.1. Спектральный метод интерполяции и экстраполяции.

Вычисление спектра, учитывая информационную равноценность динамического и спектрального представления сигналов, может производиться в дискретном варианте с использованием быстрого преобразования Фурье.

При нарушении корректности дискретизации данных погрешности интерполяции возрастают практически во всех методах интерполяции. Это можно видеть на рис. 14.4.2, который полностью повторяет рис. 14.4.1 с изменением значения только одного, пятого отсчета (уменьшение с 7.84 до 2), что вызывает подъем высоких частот в спектре данных.

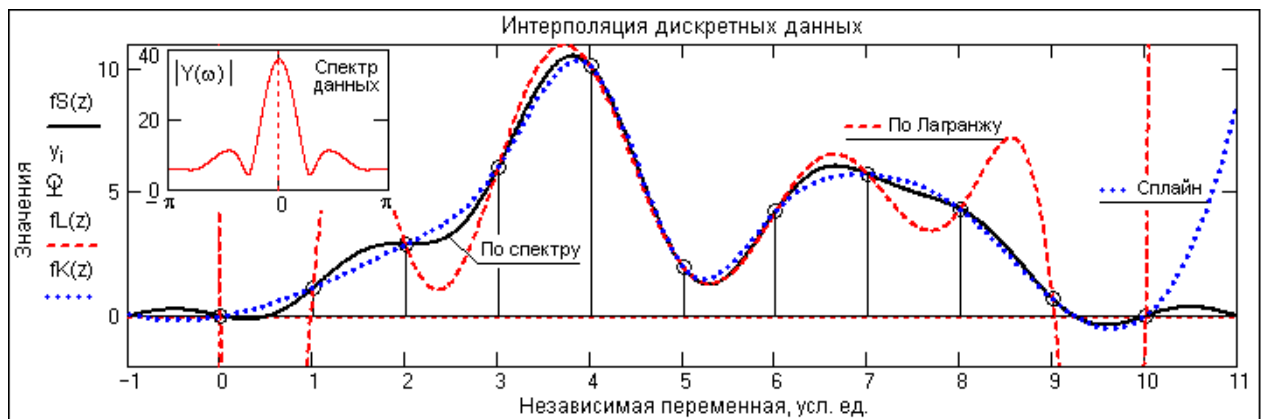


Рис. 14.4.2.

Следует учитывать, что при интерполяции данных, представляющих собой вырезки из сигнальных функций с определенной постоянной составляющей (сигнал не выходит на нулевые значения на конечных участках интервала задания), а равно и любых данных со скачками функций, при спектральном преобразовании на интерполированном сигнале в



окрестностях обрезания данных и скачков возникает явление Гиббса. Это можно видеть сравнением рисунков 14.4.1 и 14.4.3. Данные на рис. 14.4.1 в рисунке 14.4.3 подняты на 20 единиц постоянной составляющей.

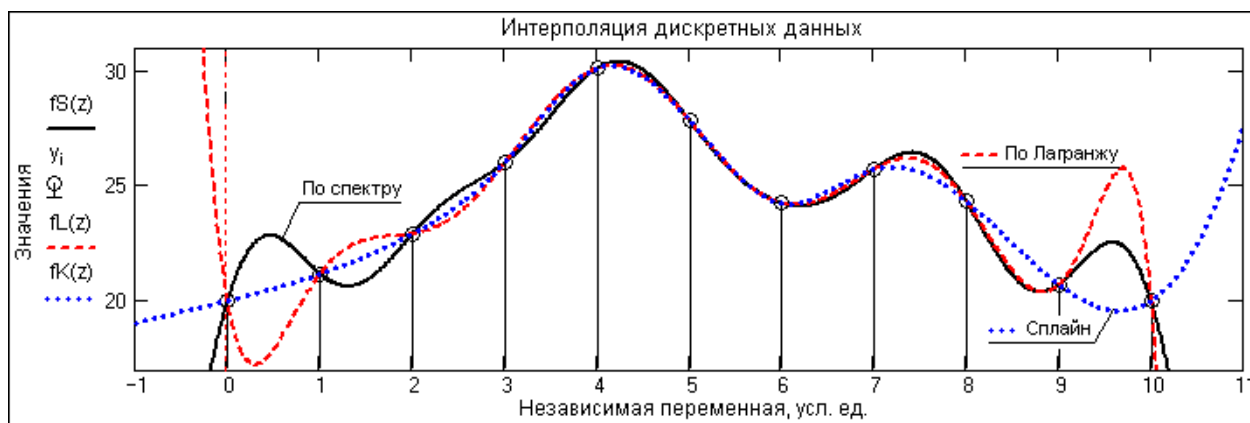


Рис. 14.4.3.

Для исключения этого эффекта можно рекомендовать перед интерполяцией производить определение линейного тренда данных по конечным значениям отсчетов и вычитать его из данных, с последующим восстановлением после интерполяции.

**Интерполяционный ряд Котельникова-Шеннона.** Произведем обратное преобразование обеих частей равенства (14.4.5). Умножение непрерывного и бесконечного спектра на  $\Pi$ -импульс в пределах главного диапазона отобразится в динамической области сверткой двух функций:

$$F \cdot s(t) = F \cdot s_0(t) \otimes \text{sinc}(\pi Ft).$$

$$s(t) = \text{sinc}(\pi Ft) \otimes \sum_{k=-\infty}^{\infty} s(k\Delta t) \delta(t-k\Delta t),$$

Отсюда, с учетом равенства  $\delta(t-k\Delta t) \otimes \text{sinc}(\pi Ft) = \text{sinc}[\pi F(t-k\Delta t)]$ , получаем:

$$s(t) = \sum_{k=-\infty}^{\infty} s(k\Delta t) \text{sinc}[\pi F(t-k\Delta t)]. \quad (14.4.6)$$

Эта формула носит название интерполяционного ряда Котельникова-Шеннона и, по существу, является разложением сигнала по системе ортогональных функций  $\text{sinc}(\pi F(t-k\Delta t)) = \text{sinc}(\pi(t/\Delta t - k))$ . С другой стороны, эта формула представляет собой свертку дискретной функции данных  $s(k\Delta t)$  с непрерывной функцией интегрального синуса. Для больших массивов дискретных данных точность восстановления сигнала обычно ограничивается интервалом задания функции интегрального синуса, по которому устанавливается интервал суммирования.

Из совокупности приведенных формул следует, что если для частоты дискретизации сигнала справедливо неравенство  $F \geq 2f_{\max}$ , где  $f_{\max}$  - наибольшая частота в спектре произвольной непрерывной функции  $s(t)$ , то функция  $s(t)$  может представляться в виде числовой последовательности дискретных значений  $s(k\Delta t)$ ,  $k = 0, 1, 2, \dots$ , и однозначно по этой последовательности восстанавливаться, в пределе - без потери точности. В этом и состоит сущность теоремы отсчетов Котельникова-Шеннона.

На рис. 14.4.4 приведен пример интерполяции входных данных, повторяющих данные рис. 14.4.1. Результаты интерполяции, как и следовало ожидать, абсолютно идентичны. Аналогичным образом влияют на результаты усечение и скачки функций (явление Гиббса).



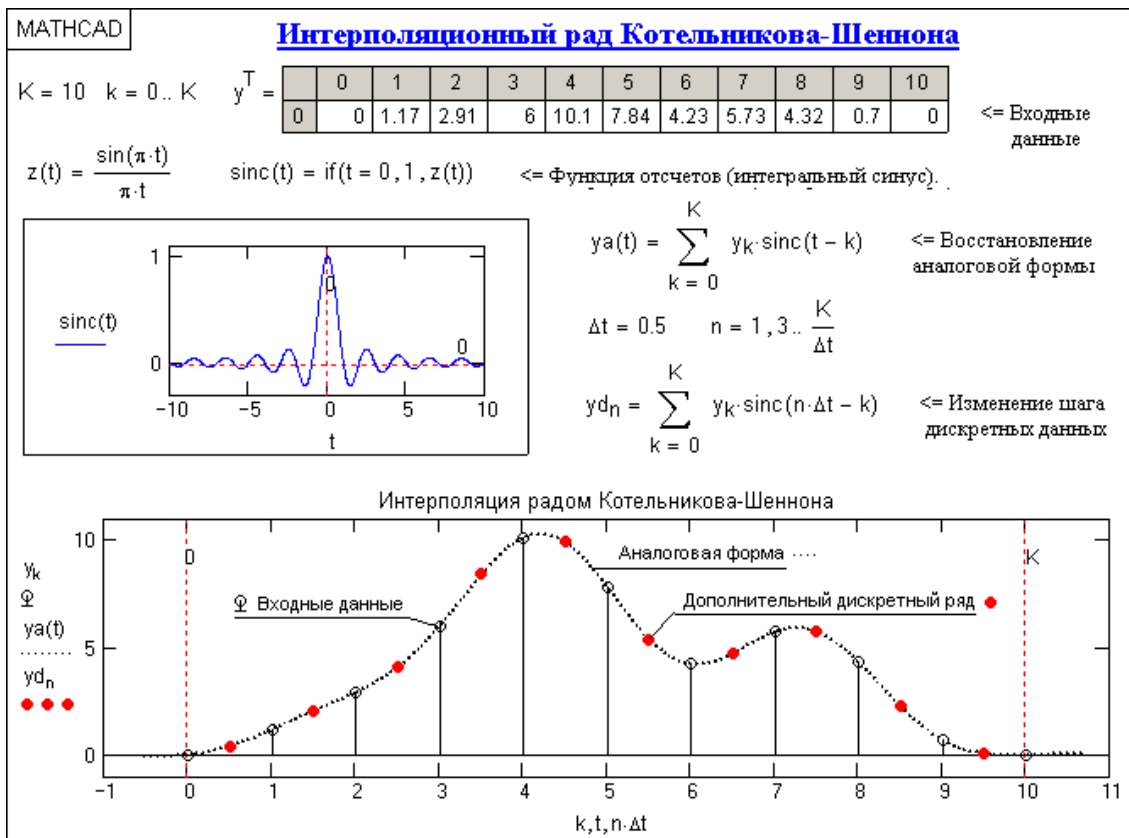


Рис. 14.4.4. Интерполяция по Котельникову-Шеннону.

#### 14.5. ДЕЦИМАЦИЯ И ИНТЕРПОЛЯЦИЯ ЦИФРОВЫХ СИГНАЛОВ

Применительно к цифровым сигналам децимация – уменьшение частоты дискретизации данных с сохранением в новом сигнале всей полезной информации. Интерполяция обратна децимации – увеличение частоты дискретизации без изменения информации. Цифровая децимация и интерполяция широко используется в современных системах обработки данных для сжатия и восстановления данных, для уменьшения объемов памяти хранения данных, для увеличения скорости передачи данных, и т.п.

Простой, но мало производительный подход – восстановить сигнал в аналоговой форме (ЦАП) и заново оцифровать его (АЦП) с новой частотой дискретизации. Цифровые методы позволяют выполнить эту операцию в более эффективной форме.

**Децимация с целым шагом.** Кратная компрессия частоты дискретизации снижает частоту дискретизации входного сигнала  $x(k)$  с  $f_d$  до  $f_d/M$  путем отбрасывания  $M-1$  отсчетов в каждой последовательной серии из  $M$ -отсчетов, т.е. из  $M$ -отсчетов оставляется только 1:

$$y(m) = \sum_{k=0}^K x(k) \delta(k-mM), \quad m = 0, 1, 2, \dots, K/M. \quad (14.5.1)$$

Естественно, что частота Найквиста  $f_N$  входного сигнала  $x(k)$  компрессора для выходного сигнала  $y(k)$  также уменьшается в  $M$  раз и становится равной  $f_N' = f_N/M$  для выходного сигнала. Для полного сохранения после компрессии полезной информации, содержащейся в сигнале  $x(k)$ , максимальная частота полезной информации во входном сигнале не должна превышать значения  $f_{\max} \leq f_N/2M$ . В противном случае децимация будет некорректной и в новом главном частотном диапазоне выходного сигнала произойдет искажение спектра полезной информации за счет сложения со спектрами боковых диапазонов. Пример корректной децимации сигнала с  $M=2$  и спектры входного и децимированного сигнала приведены на рис. 14.5.1.

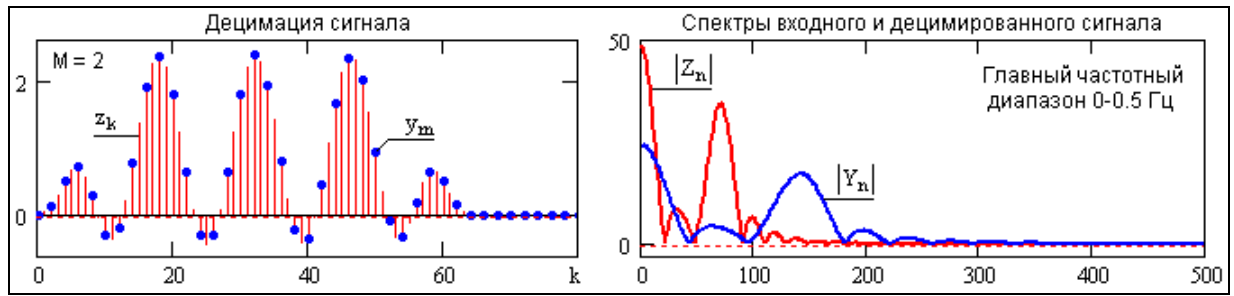


Рис. 14.5.1.

Входные сигналы кроме полезной информации могут содержать статистические шумы и помехи, распределенные по всему частотному диапазону. При децимации шумы и помехи в частотном диапазоне от  $f_N/M$  до  $f_N$  входного сигнала зеркально отражаются от  $f_N'$  нового главного частотного диапазона и их суммирование со спектром нового главного диапазона и полезного сигнала может приводить к увеличению уровня шумов и искажению информации. Для исключения этого эффекта перед конверсией сигнала необходимо выполнять его низкочастотную фильтрацию со срезом на частоте  $f_N/2M$ .

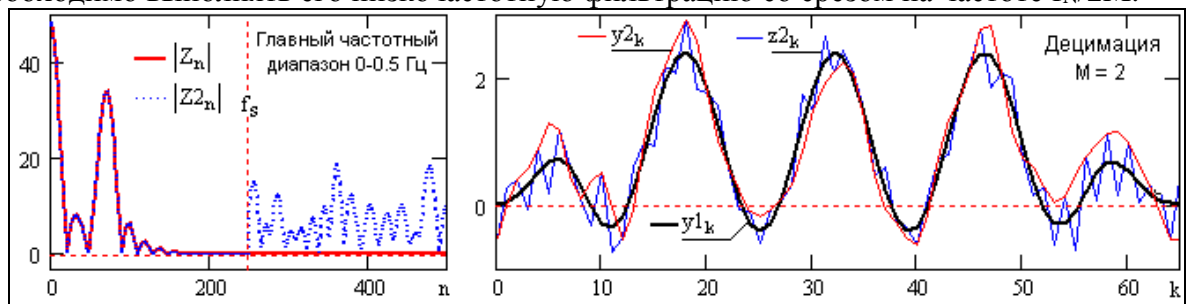


Рис. 14.5.2.

На рис. 14.5.2 приведен спектр  $Z_{2n}$  сигнала  $z_k$  с рисунка 14.5.1 (спектр  $Z_n$ ), в который для наглядности эффекта зеркального отражения условно введен только высокочастотный шум на интервале  $f_s - f_N$ , и соответствующий данному спектру сигнал  $z_{2k}$ . При децимации сигнала  $z_2$  с  $M=2$  сначала была выполнена его низкочастотная фильтрация с частотой среза  $f_s$ , что полностью сняло модельный шум, и при восстановлении сигнала (интерполяции) получен сигнал  $y_{1k}$ , полностью соответствующий сигналу  $z_k$ . При выполнении децимации без предварительной фильтрации восстанавливается сигнал  $y_{2k}$ , который по своей форме отличается как от сигнала  $y_{1k} = z_k$ , так и от входного сигнала  $z_{2k}$ .

**Интерполяция с целым шагом.** Экспандер частоты дискретизации увеличивает частоту дискретизации входного сигнала  $x_k$  с  $f_d$  до  $Lf_d$  путем введения  $(L-1)$  нулевых отсчетов после каждого отсчета входного сигнала. При этом форма спектра выходного сигнала  $y_k$  остается без изменения, но частотная шкала спектра сжимается в  $L$  раз и в границы главного диапазона спектра входного сигнала  $\pm f_N$  заходят боковые диапазоны спектра выходного сигнала (зеркальные частоты). Это наглядно можно видеть на рис. 14.5.3 сравнением спектров  $X_n$  для входного сигнала  $x_k$ , и  $Y_n$  для экспандированного сигнала  $x_k$  с  $L=2$ . Следовательно, фактическая частота Найквиста выходного сигнала становится равной  $f_N/L$ . Для исключения зеркальных частот и распределения энергии отсчетов  $x_k$  по  $L$  выходным интервалам экспандированный сигнал пропускается через фильтр нижних частот со срезом на частоте  $f_N/L$  и с коэффициентом  $L$  для компенсации распределения энергии отсчетов по интервалам  $L$ . Результат операции можно видеть на сигнале  $y_k$  по сравнению с исходным сигналом  $z_k$  (рис. 14.5.1), децимацией которого с  $M=2$  был получен сигнал  $x_k$ .

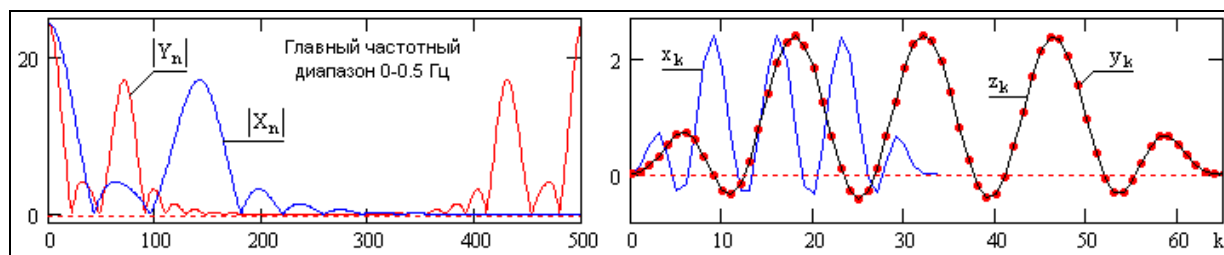


Рис. 14.5.3.

**Преобразование частоты дискретизации с нецелым шагом** на практике обычно выполняют представлением нецелого множителя максимально близким приближением рациональными числами вида  $L/M$ . Это позволяет выполнять преобразование частоты дискретизации последовательными операциями сначала интерполяции с шагом  $L$ , сохраняющей все частотные составляющие сигнала, и затем децимации с шагом  $M$ , при которой часть высокочастотных составляющих и шумов будет подавлена низкочастотной фильтрацией. Поскольку при этом низкочастотные фильтры экспандирования и децимации следуют друг за другом и работают на одной частоте дискретизации, то вместо двух фильтров можно применять один, имеющий минимальную частоту среза с коэффициентом усиления, равным  $L$ .

При программной обработке больших пакетов данных децимация и интерполяция может выполняться в спектральной области с использованием БПФ.

#### 14.6. МЕТОДИКА АППРОКСИМАЦИИ ЭМПИРИЧЕСКИХ ДАННЫХ

Эмпирические данные, как правило, задаются числовыми рядами значений двух величин: независимой ( $x_k$ ) и зависимой ( $y_k$ ), каждая из которых кроме определенной регулярной составляющей может содержать и случайные составляющие самой различной природы, обусловленные как статистической природой изучаемых процессов, так и внешними факторами процессов измерений и преобразования данных (шумы, помехи, ошибки измерений). Независимая переменная  $x_k$  обычно полагается детерминированной, а ее случайная составляющая "переносится" на зависимую переменную  $y_k$ . Полагается также, что значения случайной составляющей зависимой переменной распределены по некоторому вероятностному закону (например – нормальному).

При выполнении аппроксимации данных предполагается существование определенной детерминированной связи  $y(x)$  между регулярными составляющими этих двух числовых рядов на статистически значимом уровне, достаточном для ее выявления на уровне случайных составляющих. Задача выявления такой закономерности относится к числу неопределенных и неоднозначных, результат которой зависит от трех основных субъективных факторов:

- выбора меры близости зависимой переменной к искомой функции и метода построения приближения (параметров математической модели);
- выбора подходящего класса функции аппроксимации (степенной, тригонометрической и пр.), отвечающего физической природе моделируемого процесса;
- метода оптимизации порядка модельной функции или числа членов ряда аппроксимирующего выражения.

Отсюда следует, что оптимальная аппроксимация может быть обеспечена только достаточно гибкими интерактивными алгоритмами на основе многоэтапных итерационных процессов с возможностью коррекции на каждом этапе.

**Мера приближения.** Наиболее распространен критерий наилучшего приближения в виде минимума степенной разности между переменной  $y_k$  и аппроксимирующей функцией  $\varphi(x_k)$ :

$$\sum_k [y_k - \varphi(x_k)]^S \rightarrow \min, \quad (14.6.1)$$

где  $S > 0$  - положительное число.

**Квадратичная мера** реализуется при  $S = 2$  в методе наименьших квадратов (МНК) и обеспечивает максимальное правдоподобие функции приближения при нормальном распределении случайной составляющей зависимой переменной  $y_k$ . Несмещенной оценкой меры приближения в МНК является дисперсия остатков:

$$D = \{ \sum_k [y_k - \varphi(x_k)]^2 \} / (k-m), \quad (14.6.2)$$

где  $m$  – количество параметров в функции приближения,  $(k-m)$  – число степеней свободы. Однако эмпирические данные могут содержать выбросы и грубые ошибки, которые вызывают смещения вычисляемых параметров. Их влияние обычно исключается цензурированием данных: вычислением гистограммы разностей  $y_k - \varphi(x_k)$  после определения первого приближения функции аппроксимации и исключением "хвостовых" элементов гистограммы (до 2.5% от количества данных, или резко выделяющихся элементов данных на основании оценок вероятностей с использованием  $g$ - или  $t$ -распределений).

**Мера наименьших модулей** (метод Лагранжа) реализуется при  $S = 1$  и применяется при распределениях случайных составляющих зависимой переменной по законам, близким к закону Лапласа (двустороннее экспоненциальное распределение). Такая мера соответствует площади между графиками эмпирических данных и функции аппроксимации, и, по сравнению с квадратической, является более устойчивой, в том числе при наличии случайных составляющих с большими амплитудами (длинные "хвосты" разностных гистограмм). Оценки по модулю получили название "робастных" (robust – устойчивый).

Свойства квадратичной меры и меры наименьших модулей в определенной степени сочетаются при  $S = 3/2$ .

**Минимаксная мера** (мера Чебышева – минимизация максимального расхождения функции аппроксимации с данными) обеспечивает наилучшее приближение при равномерном распределении значений случайной составляющей, но не является устойчивой при наличии больших расхождений данных с функцией аппроксимации.

**Аппроксимирующая функция**, в принципе, может быть математической функцией любого типа, линейной комбинацией различных функций или функциональным рядом из степенных, тригонометрических и любых других функций. В основу ее построения желательно закладывать теоретические предположения о сущности изучаемого явления, хотя бы по таким свойствам, как область определения переменных и производных, асимптоты, минимумы и максимумы.

При полном отсутствии априорной информации о распределении случайной составляющей данных, на начальном этапе обычно используется квадратичная мера приближения, при этом существенное значение имеет количество задаваемых параметров функции аппроксимации, особенно при малом количестве данных. Как следует из (14.6.2), при прочих равных условиях целесообразно использовать функции с минимальным количеством задаваемых параметров, что обеспечивает большее число степеней свободы и меньшие значения дисперсии остатков.

Наибольшее распространение в практике аппроксимации при отсутствии теоретических аспектов изучаемых явлений получили функциональные ряды, для которых определяющее значение имеет порядок аппроксимирующей функции (модели).

**Порядок модели** ограничивает число членов функционального ряда аппроксимирующей функции определенным оптимальным количеством членов ряда, которое обеспечивает обоснованное расхождение с фактическими данными и минимизирующее отклонение от искомой регулярной составляющей данных.

Очевидно, что для функциональных рядов порядок модели (степень ряда для степенных рядов) определяет значение меры приближения. При повышении порядка модели минимум функции (14.6.1) стремится к нулю. Однако это означает, что при повышении порядка модели в функцию аппроксимации входит не только регулярная

составляющая данных, но все большая и большая доля случайных составляющих, в пределе до полного соответствия функции  $\varphi_k$  исходным данным  $y_k$ . Но повышение степени приближения к исходным данным при наличии в них случайных составляющих с какого-то определенного момента (порядка модели) не только не будет приближать функцию аппроксимации к регулярным составляющим данных, а наоборот – увеличивать расхождение. С этой точки зрения термин "меры приближения" (14.6.1) было бы целесообразнее заменить термином "мера аппроксимации" данных, а под мерой приближения понимать значение меры аппроксимации, при которой обеспечивается максимальная степень приближения функции аппроксимации к регулярной составляющей данных (минимум дисперсии разности функций аппроксимации и регулярной составляющей).

При разделении данных на значения регулярных составляющих  $s_k$  и случайных  $\sigma_k$ , для квадратичной меры можно записать:

$$\sum_k [y_k - \varphi(x_k)]^2 = \sum_k [s_k + \sigma_k - \varphi(x_k)]^2 = \sum_k [s_k - \varphi(x_k)]^2 + 2 \sum_k [s_k - \varphi(x_k)]\sigma_k + \sum_k \sigma_k^2.$$

При нулевом значении математического ожидания случайных величин  $\sigma_k$  значение второй суммы стремится к нулю, при этом для оптимальной аппроксимирующей функции:

$$\sum_k [s_k - \varphi(x_k)]^2 \rightarrow \min, \quad (14.6.3)$$

$$\sum_k [y_k - \varphi(x_k)]^2 \rightarrow \sum_k \sigma_k^2. \quad (14.6.4)$$

В пределе, при идеальной аппроксимации, выражение (14.6.3) стремится к нулю, а выражение (14.6.4) эквивалентно соотношению дисперсий:

$$\left\{ \sum_k [y_k - \varphi(x_k)]^2 \right\} / (k-m) \approx \sum_k \sigma_k^2 / k. \quad (14.6.5)$$

Отсюда следует, что при прочих равных условиях наилучшим является приближение, у которого мера приближения близка к дисперсии шума. Для "белых" шумов оценку их дисперсии в экспериментальных данных можно выполнять в спектральной области, если частота Найквиста данных минимум в 2 раза выше предельных частот регулярной составляющей.

При отсутствии информации о дисперсии шумов оптимальный порядок модели может определяться методом последовательных уточнений с последовательным нарастанием порядка модели и сравнением по критерию Фишера значимости различия дисперсии остатков каждого нового порядка с предыдущим. При увеличении порядка модели (начиная с 1-го) значимость различия дисперсий сначала является довольно высокой, постепенно уменьшается, и в области оптимальных порядков становится малозначимой. Это объясняется тем, что в этой области при небольших уменьшениях значения числителя выражения (14.6.2) одновременно, за счет увеличения порядка, сокращается число степеней свободы. После прохождения оптимальной зоны значения дисперсий остатков снова начинают увеличиваться с увеличением значимости различий.

**Оптимальная полиномиальная аппроксимация**

**Моделирование данных.**

$n = 0..5$   $K = 50$   $k = 0..K$

Модель нормального шума:  $q = \text{norm}(K + 1, 10, 3)$

$c_{1..5}$

$\leq$  Коэффициенты регулярной составляющей

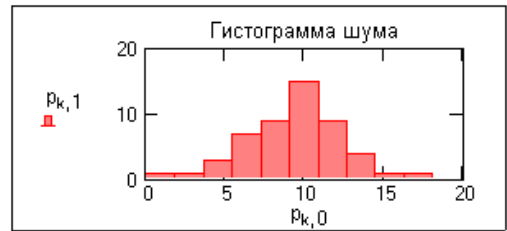
$p = \text{histogram}(10, q)$   $\text{mean}(q) = 9.277$   $Dq = \text{var}(q)$

5
4.705
-0.832
0.049
-0.001136
$9.021 \cdot 10^{-6}$

$x_k = k$   $\leq$  Вектор независимой переменной

$$s_k = \sum_{n=0}^5 c_n \cdot k^n \quad \leq \text{Регулярная составляющая}$$

$y_k = s_k + q_k - \text{mean}(q)$   $\leq$  Модель данных



**Обработка данных.**

$N = \text{rows}(y)$   $N = 51$

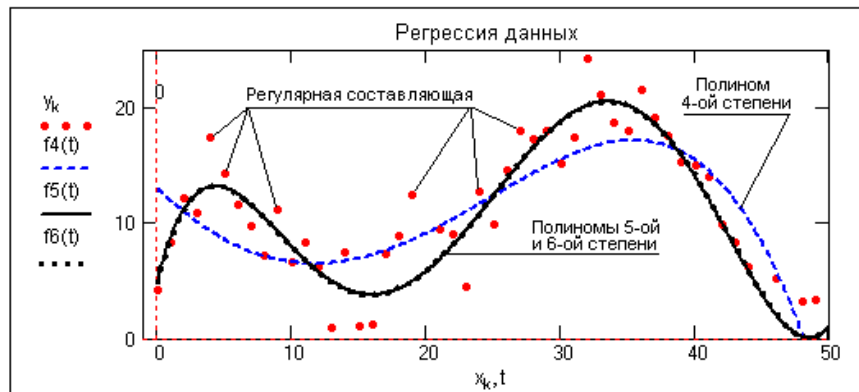
$s3 = \text{regress}(x, y, 3)$   $s4 = \text{regress}(x, y, 4)$   $s5 = \text{regress}(x, y, 5)$   $s6 = \text{regress}(x, y, 6)$   $s7 = \text{regress}(x, y, 7)$

$f3(t) = \text{interp}(s3, x, y, t)$   $f4(t) = \text{interp}(s4, x, y, t)$   $f5(t) = \text{interp}(s5, x, y, t)$   $f6(t) = \text{interp}(s6, x, y, t)$   $f7(t) = \text{interp}(s7, x, y, t)$

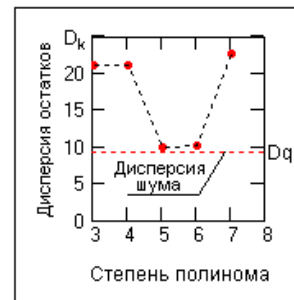
$k5 = \text{submatrix}(s5, 3, \text{length}(s5) - 1, 0, 0)$   $k5^T = \begin{pmatrix} 5.002 & 4.444 & -0.765 & 0.044 & -1.014 \times 10^{-3} & 7.934 \times 10^{-6} \end{pmatrix}$

$$d(N, m, f, x, y) = \frac{1}{N - m} \cdot \sum_{k=0}^{N-1} (f(x_k) - y_k)^2 \quad D_3 = d(N, 3, f3, x, y) \quad D_4 = d(N, 4, f4, x, y) \quad D_5 = d(N, 5, f5, x, y)$$

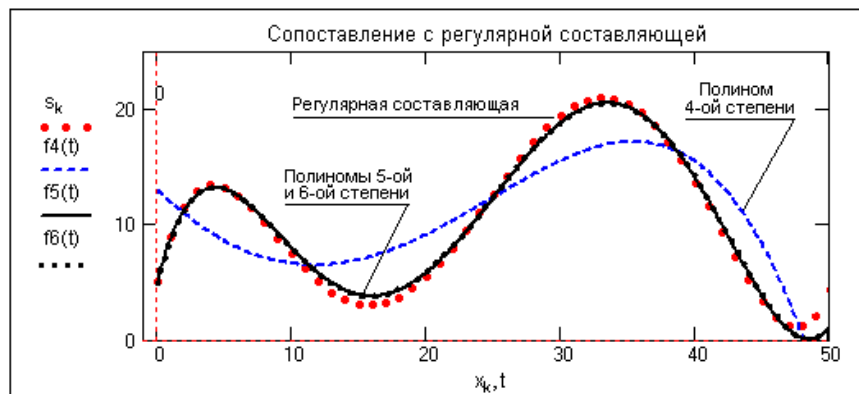
$$D_6 = d(N, 6, f6, x, y) \quad D_7 = d(N, 7, f7, x, y)$$



**Дисперсия приближения**



$Ds_3 = d(N, 3, f3, x, s)$   $Ds_4 = d(N, 4, f4, x, s)$   $Ds_5 = d(N, 5, f5, x, s)$   $Ds_6 = d(N, 6, f6, x, s)$   $Ds_7 = d(N, 7, f7, x, s)$



**Дисперсия расхождения с регулярной составляющей**

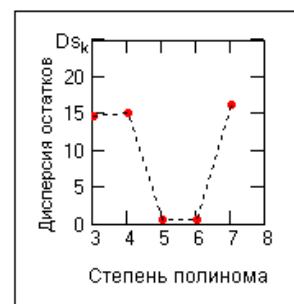


Рис. 14.6.1.

Оптимальный порядок модели при нормальном распределении шума может устанавливаться и непосредственно по минимуму дисперсии остатков. Это можно наглядно видеть на примере, приведенном на рис. 14.6.1.

Одномерная полиномиальная аппроксимация данных в векторе Y полиномом с произвольной степенью n и с произвольными координатами отсчетов в векторе X в Mathcad выполняется функциями:

- $\text{regress}(X, Y, n)$  – вычисляет вектор S для функции  $\text{interp}(\dots)$ , в составе которого находятся коэффициенты  $c_i$  полинома n-й степени;
- $\text{interp}(S, X, Y, x)$  – возвращает значения функции аппроксимации по координатам x. Функция  $\text{interp}(\dots)$  реализует вычисления по формуле:

$$f(x) = c_0 + c_1 \cdot x^1 + c_2 \cdot x^2 + \dots + c_n \cdot x^n \equiv \sum_i c_i \cdot x^i.$$

- Значения коэффициентов  $c_i$  могут быть извлечены из вектора  $S$  функцией  $\blacktriangleright$  `submatrix(S, 3, length(S), 0, 0)`.

**Оценка качества приближения.** Для оценки качества математической модели эмпирической зависимости используется коэффициент детерминации (Adjusted  $R^2$ ):

$$\text{Adjusted } R^2 = D_\varphi / D_y = 1 - D_o / D_y,$$

где:  $D_\varphi$  - дисперсия функции приближения,  $D_y$  - дисперсия данных,  $D_o$  - дисперсия остатков. Чем выше качество аппроксимации, тем ближе к 1 значение коэффициента детерминации.

## Тема 15. РЕГРЕССИЯ

Аппроксимация данных с учетом их статистических параметров относится к задачам *регрессии*. Они обычно возникают при обработке экспериментальных данных, полученных в результате измерений процессов или физических явлений, статистических по своей природе (как, например, измерения в радиометрии и ядерной геофизике), или на высоком уровне помех (шумов). Задачей регрессионного анализа является подбор математических формул, наилучшим образом описывающих экспериментальные данные.

Термин "регрессия" появился при исследовании соотношения роста родителей и их детей, в которых было установлено, что рост "регрессирует" к среднему, т.е. высокие родители имеют более низких детей, а низкие родители – более высоких.

В качестве основной математической системы для примеров будем использовать систему Mathcad.

### 15.1. ПОСТАНОВКА ЗАДАЧИ РЕГРЕССИИ

Математическая постановка задачи регрессии заключается в следующем. Зависимость величины (числового значения) определенного свойства случайного процесса или физического явления  $Y$  от другого переменного свойства или параметра  $X$ , которое в общем случае также может относиться к случайной величине, зарегистрирована на множестве точек  $x_k$  множеством значений  $y_k$ , при этом в каждой точке зарегистрированные значения  $y_k$  и  $x_k$  отображают действительные значения  $Y(x_k)$  со случайной погрешностью  $\sigma_k$ , распределенной, как правило, по нормальному закону. По совокупности значений  $y_k$  требуется подобрать такую функцию  $f(x_k, a_0, a_1, \dots, a_n)$ , которой зависимость  $Y(x)$  отображалась бы с минимальной погрешностью. Отсюда следует условие приближения:

$$y_k = f(x_k, a_0, a_1, \dots, a_n) + \sigma_k.$$

Функцию  $f(x_k, a_0, a_1, \dots, a_n)$  называют регрессией величины  $y$  на величину  $x$ . Регрессионный анализ предусматривает задание вида функции  $f(x_k, a_0, a_1, \dots, a_n)$  и определение численных значений ее параметров  $a_0, a_1, \dots, a_n$ , обеспечивающих наименьшую погрешность приближения к множеству значений  $y_k$ . Как правило, при регрессионном анализе погрешность приближения вычисляется *методом наименьших квадратов* (МНК). Для этого выполняется минимизация функции квадратов остаточных ошибок:

$$\sigma(a_0, a_1, \dots, a_n) = \sum_k [f(x_k, a_0, a_1, \dots, a_n) - y_k]^2.$$

Для определения параметров  $a_0, a_1, \dots, a_n$  функция остаточных ошибок дифференцируется по всем параметрам, полученные уравнения частных производных приравниваются нулю и решаются в совокупности относительно всех значений параметров. Виды регрессии обычно называются по типу аппроксимирующих функций: полиномиальная, экспоненциальная, логарифмическая и т.п.

### 15.2. ЛИНЕЙНАЯ РЕГРЕССИЯ



**Общий принцип.** Простейший способ аппроксимации по МНК произвольных данных  $s_k$  - с помощью полинома первой степени, т.е. функции вида  $y(t) = a+bt$ , которую обычно называют линией регрессии. С учетом дискретности данных по точкам  $t_k$ , для функции остаточных ошибок имеем:

$$\sigma(a, b) = \sum_k [(a+b t_k) - s_k]^2.$$

Для вычисления оценок коэффициентов дифференцируем функцию остаточных ошибок по аргументам  $a$  и  $b$ , приравняем полученные уравнения нулю и формируем два нормальных уравнения системы:

$$\sum_k 2((a+b t_k)-s_k) \equiv a \sum_k 1 + b \sum_k t_k - \sum_k s_k = 0,$$

$$\sum_k 2((a+b t_k)-s_k) t_k \equiv a \sum_k t_k + b \sum_k t_k^2 - \sum_k s_k t_k = 0,$$

Решение данной системы уравнений в явной форме для  $K$ -отсчетов:

$$b = [\sum_k t_k s_k - \sum_k t_k \sum_k s_k] / [\sum_k t_k^2 - (\sum_k t_k)^2] = (\overline{s_k t_k} - \overline{s_k} \overline{t_k}) / (\overline{t_k^2} - \overline{t_k}^2).$$

$$a = [\sum_k s_k - b \sum_k t_k] / K = \overline{s_k} - b \overline{t_k}$$

Полученные значения коэффициентов используем в уравнении регрессии  $y(t) = a+bt$ . Прямая  $(s - \overline{s_k}) = b (t - \overline{t_k})$  называется линией регрессии  $s$  по  $t$ . Для получения линии регрессии  $t$  по  $s$ ,  $(t - \overline{t_k}) = b (s - \overline{s_k})$ , аргумент  $b$  в этой формуле заменяется на значение  $b = (\overline{s_k t_k} - \overline{s_k} \overline{t_k}) / (\overline{s_k^2} - \overline{s_k}^2)$ .

По аналогичной методике вычисляются коэффициенты и любых других видов регрессии, отличаясь только громоздкостью соответствующих выражений.

**Реализация в Mathcad.** Линейная регрессия в системе Mathcad выполняется по векторам аргумента  $X$  и отсчетов  $Y$  функциями:

- $\text{intercept}(X, Y)$  – вычисляет параметр  $a$ , смещение линии регрессии по вертикали;
- $\text{slope}(X, Y)$  – вычисляет параметр  $b$ , угловой коэффициент линии регрессии.

Расположение отсчетов по аргументу  $X$  произвольное. Функцией  $\text{corr}(X, Y)$  дополнительно можно вычислить коэффициент корреляции Пирсона. Чем он ближе к 1, тем точнее обрабатываемые данные соответствуют линейной зависимости.

Пример выполнения линейной регрессии приведен на рис. 15.2.1.

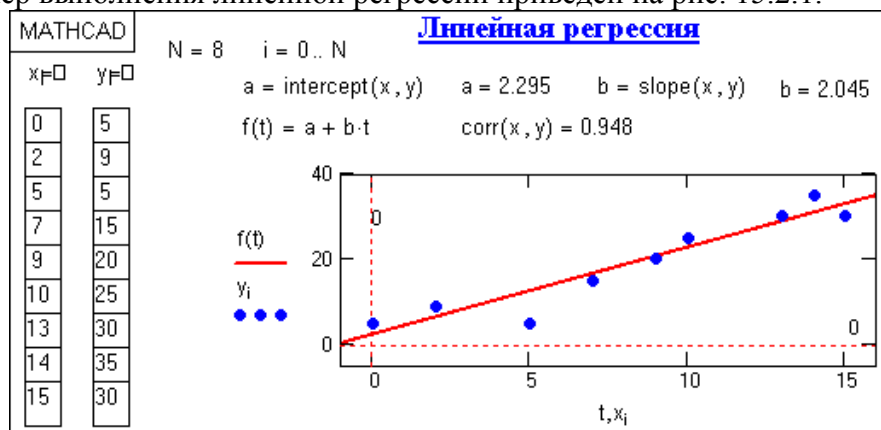


Рис. 15.2.1.

### 15.3. ПОЛИНОМИАЛЬНАЯ РЕГРЕССИЯ [25]

**Одномерная полиномиальная регрессия** с произвольной степенью  $n$  полинома и с произвольными координатами отсчетов в Mathcad выполняется функциями:

- $\text{regress}(X, Y, n)$  – вычисляет вектор  $S$  для функции  $\text{interp}(\dots)$ , в составе которого находятся коэффициенты  $k_i$  полинома  $n$ -й степени;
- $\text{interp}(S, X, Y, x)$  – возвращает значения функции аппроксимации по координатам  $x$ . Функция  $\text{interp}(\dots)$  реализует вычисления по формуле:



$$f(x) = k_0 + k_1 x^1 + k_2 x^2 + \dots + k_n x^n \equiv \sum_i k_i x^i.$$

Значения коэффициентов  $k_i$  могут быть извлечены из вектора S функцией `submatrix(S, 3, length(S), 0, 0)`.

На рис. 15.3.1 приведен пример полиномиальной регрессии с использованием полиномов 2, 3 и 8-й степени. Степень полинома обычно устанавливают не более 4-6 с последовательным повышением степени, контролируя среднеквадратическое отклонение функции аппроксимации от фактических данных. Нетрудно заметить, что по мере повышения степени полинома функция аппроксимации приближается к фактическим данным, а при степени полинома, равной количеству отсчетов минус 1, вообще превращается в функцию интерполяции данных, что не соответствует задачам регрессии.

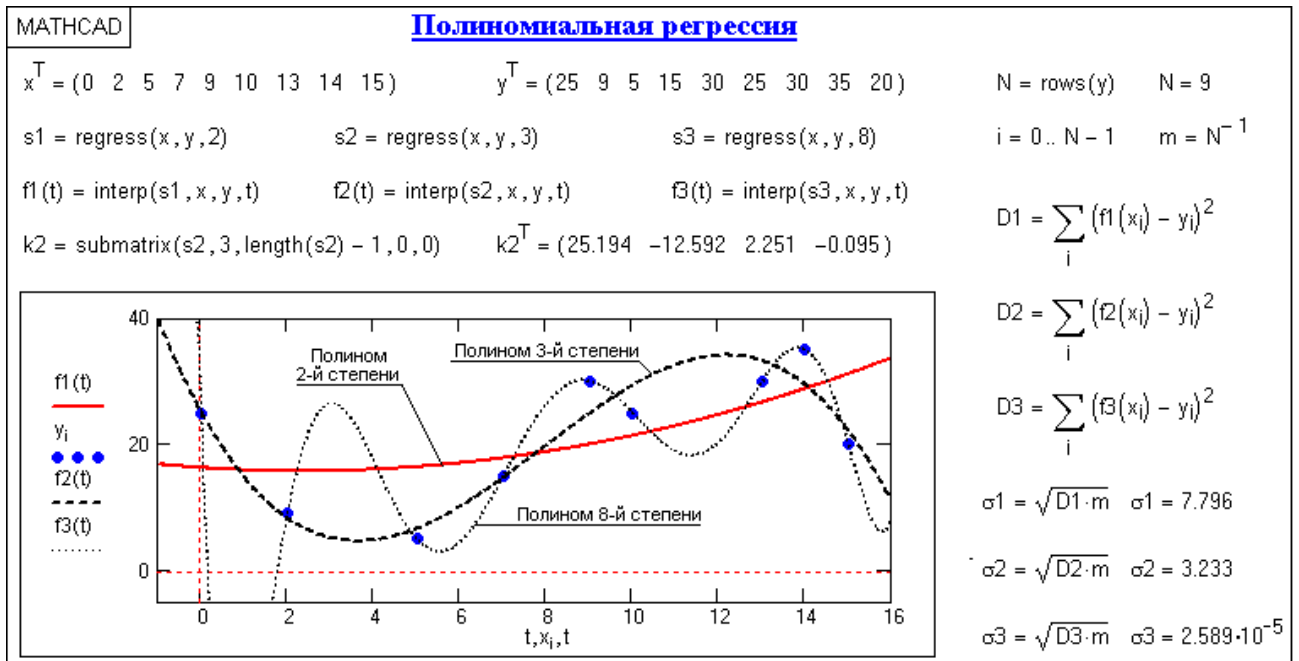
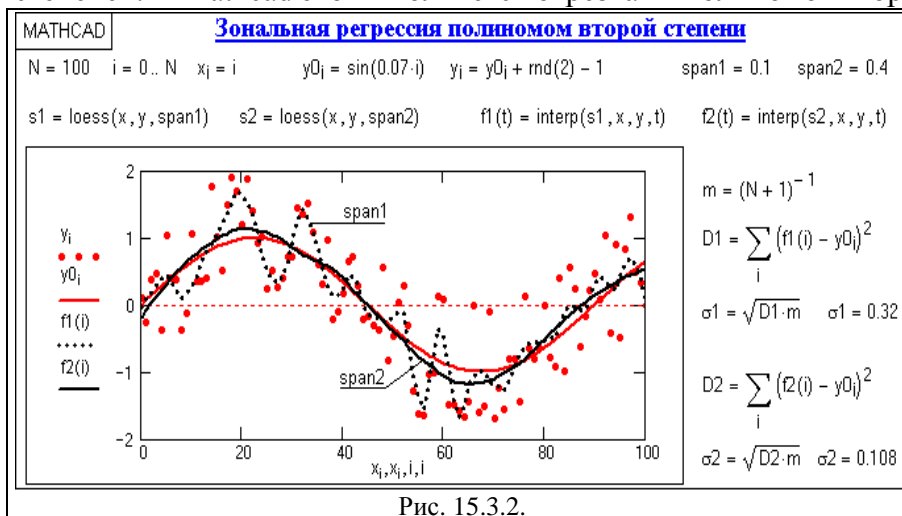


Рис. 15.3.1. Одномерная полиномиальная регрессия.

**Зональная регрессия.** Функция `regress` по всей совокупности точек создает один аппроксимирующий полином. При больших координатных интервалах с большим количеством отсчетов и достаточно сложной динамике изменения данных рекомендуется применять последовательную локальную регрессию отрезками полиномов малых степеней. В Mathcad это выполняется отрезками полиномов второй степени функцией



`loess(X, Y, span)`, которая формирует специальный вектор S для функции `interp(S, X, Y, x)`. Аргумент `span > 0` в этой функции (порядка 0.1-2) определяет размер локальной области и подбирается с учетом характера данных и необходимой степени их сглаживания (чем больше `span`, тем

больше степень сглаживания данных).

На рис. 15.3.2 приведен пример вычисления регрессии модельной кривой (отрезка

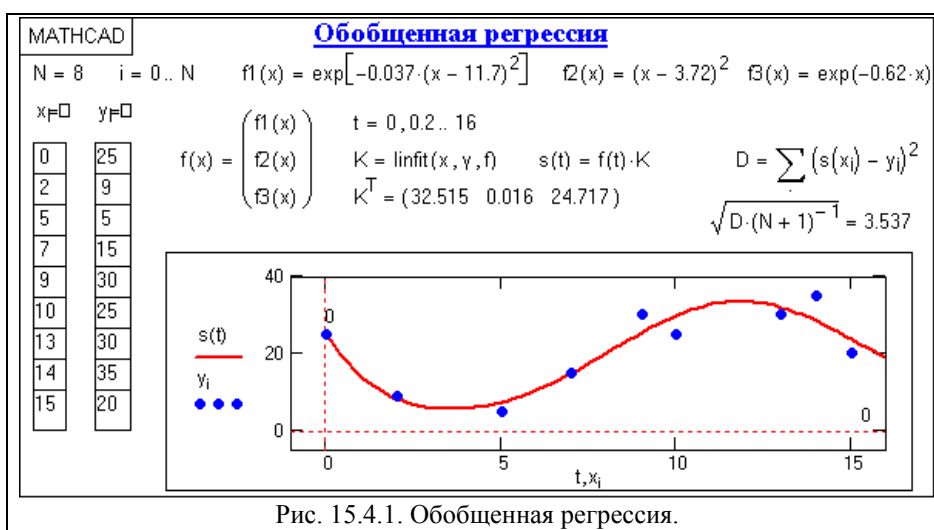
синусоиды) в сумме с шумами. Вычисления выполнены для двух значений  $\text{span}$  с определением среднеквадратического приближения к базовой кривой. При моделировании каких-либо случайных процессов и сигналов на высоком уровне шумов по минимуму среднеквадратического приближения может определяться оптимальное значение параметра  $\text{span}$ .

#### 15.4. НЕЛИНЕЙНАЯ РЕГРЕССИЯ

**Линейное суммирование произвольных функций.** В Mathcad имеется возможность выполнения регрессии с приближением к функции общего вида в виде весовой суммы функций  $f_n(x)$ :

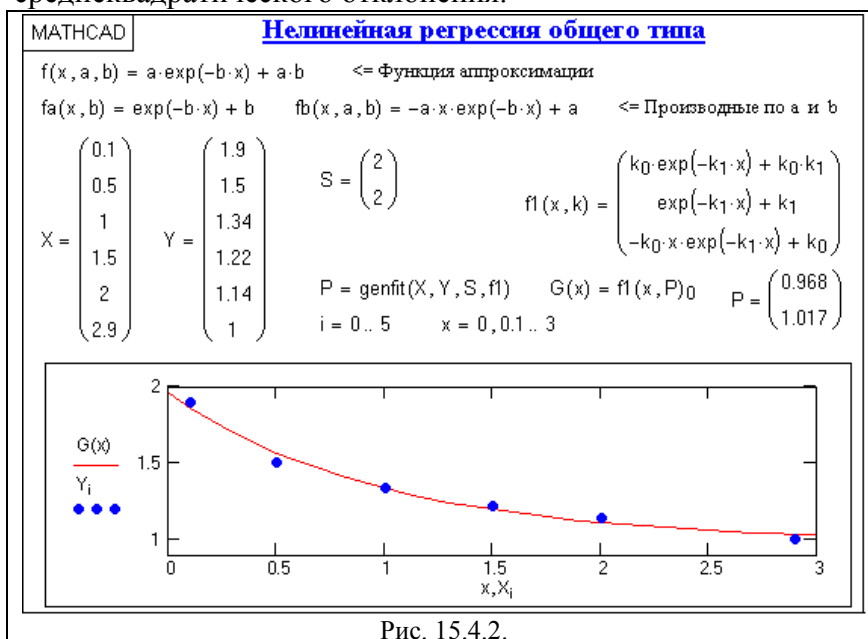
$$f(x, K_n) = K_1 f_1(x) + K_2 f_2(x) + \dots + K_N f_N(x),$$

при этом сами функции  $f_n(x)$  могут быть любого, в том числе нелинейного типа. С одной стороны, это резко повышает возможности аналитического отображения функций регрессии. Но, с другой стороны, это требует от пользователя определенных навыков аппроксимации экспериментальных данных комбинациями достаточно простых функций.



Реализуется обобщенная регрессия по векторам  $X$ ,  $Y$  и  $f$  функцией  $\text{linfit}(X, Y, f)$ , которая вычисляет значения коэффициентов  $K_n$ . Вектор  $f$  должен содержать символьную запись функций  $f_n(x)$ . Координаты  $x_k$  в векторе  $X$  могут

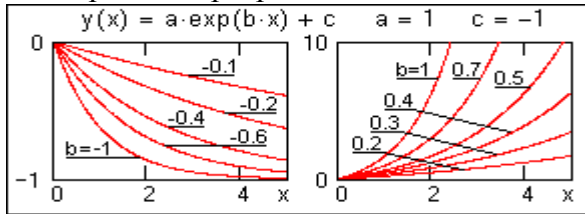
быть любыми, но расположенными в порядке возрастания значений  $x$  (с соответствующими отсчетами значений  $y_k$  в векторе  $Y$ ). Пример выполнения регрессии приведен на рис. 15.4.1. Числовые параметры функций  $f_1$ - $f_3$  подбирались по минимуму среднеквадратического отклонения.



**Регрессия общего типа.** Второй вид нелинейной регрессии реализуется путем подбора параметров  $k_i$  к заданной функции аппроксимации с использованием функции  $\text{genfit}(X, Y, S, F)$ , которая возвращает коэффициенты  $k_i$ , обеспечивающие минимальную среднюю квадратическую погрешность приближения функции регрессии к входным

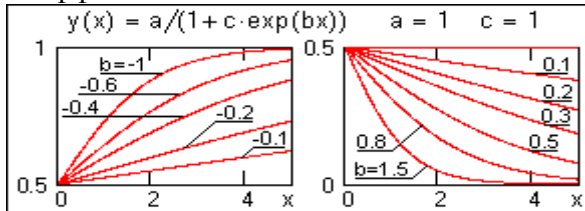
данным (векторы X и Y координат и отсчетов). Символьное выражение функции регрессии и символьные выражения ее производных по параметрам  $k_i$  записываются в вектор F. Вектор S содержит начальные значения коэффициентов  $k_i$  для решения системы нелинейных уравнений итерационным методом. Пример использования метода приведен на рис. 15.4.2.

**Типовые функции регрессии Mathcad.** Для простых типовых формул аппроксимации предусмотрен ряд функций регрессии, в которых параметры функций подбираются программой Mathcad самостоятельно. К ним относятся следующие функции:

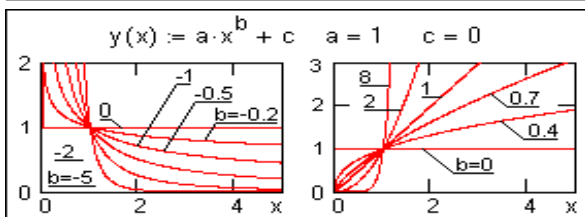


→  $\text{expfit}(X, Y, S)$  – возвращает вектор, содержащий коэффициенты  $a$ ,  $b$  и  $c$  экспоненциальной функции  $y(x) = a \cdot \exp(b \cdot x) + c$ . В вектор S вводятся начальные значения коэффициентов  $a$ ,  $b$  и  $c$  первого приближения. Для ориентировки по форме

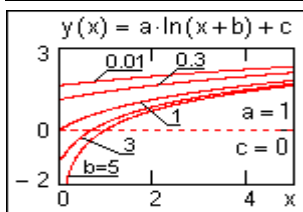
аппроксимационных функций и задания соответствующих начальных значений коэффициентов на рисунках слева приводится вид функций при постоянных значениях коэффициентов  $a$  и  $c$ .



→  $\text{lgsfit}(X, Y, S)$  – то же, для выражения  $y(x) = a / (1 + c \cdot \exp(b \cdot x))$ .

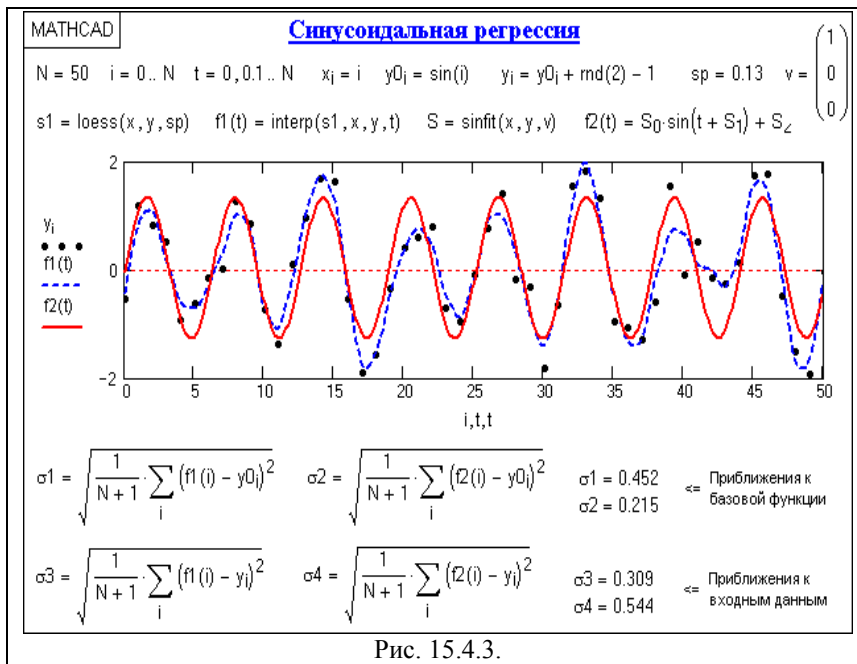


→  $\text{pwrfit}(X, Y, S)$  – то же, для выражения  $y(x) = a \cdot x^b + c$ .



→  $\text{logfit}(X, Y)$  – то же, для выражения  $y(x) = a \ln(x + b) + c$ . Задания начального приближения не требуется.

→  $\text{medfit}(X, Y)$  – то же, для выражения  $y(x) = a + b \cdot x$ , т.е. для функции линейной регрессии. Задания начального приближения также не требуется. График – прямая линия.

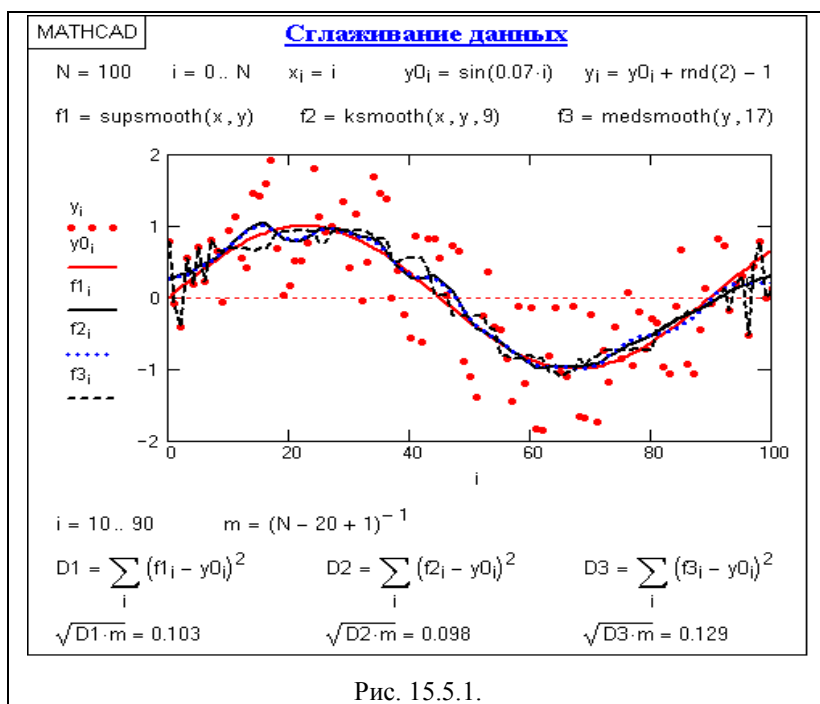


На рис. 15.4.3 приведен пример реализации синусоидальной регрессии модельного массива данных по базовой синусоиде в сопоставлении с зональной регрессией полиномом второй степени. Как можно видеть из сопоставления методов по средним квадратическим приближениям к базовой кривой и к исходным данным, известность функции математического ожидания для статистических данных с ее использованием в качестве базовой для функции регрессии дает возможность с более высокой точностью определять параметры регрессии в целом по всей совокупности данных, хотя при этом кривая регрессии не отражает локальных особенностей фактических отсчетов данной реализации. Это имеет место и для всех других методов с заданием функций регрессии.

### 15.5. СГЛАЖИВАНИЕ ДАННЫХ

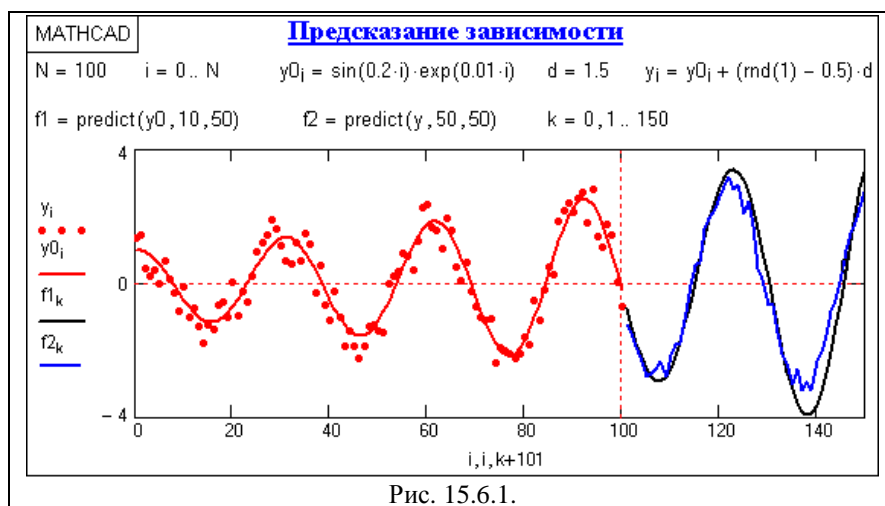
Сглаживание данных, как искаженных помехами, так и статистических по своей природе, можно считать частным случаем регрессии без определения символьной формы ее функции. В Mathcad для сглаживания применяются следующие функции:

- $\text{supsmooth}(X, Y)$  – возвращает вектор линейно сглаженных данных  $Y$ , метод наименьших квадратов по  $k$  отсчетам с адаптивным выбором значения  $k$  с учетом динамики изменения данных. Значения вектора  $X$  должны идти в порядке возрастания.
- $\text{ksmooth}(X, Y, b)$  – вычисляет вектор сглаженных данных на основе распределения Гаусса. Параметр  $b$  задает ширину окна сглаживания и должен быть в несколько раз больше интервала между отсчетами по оси  $x$ .
- $\text{medsmooth}(Y, b)$  – вычисляет вектор сглаженных данных по методу скользящей медианы с шириной окна  $b$ , которое должно быть нечетным числом.



Сопоставление методов сглаживания приведено на рис. 15.5.1. Как можно видеть на этом рисунке, качество сглаживания функциями  $\text{supsmooth}(X, Y)$  и  $\text{ksmooth}(X, Y, b)$  практически идентично (при соответствующем выборе параметра  $b$ ). Медианный способ уступает по своим возможностям двум другим. Можно заметить также, что на конечных точках интервала задания данных качество сглаживания ухудшается, особенно в медианном способе, который вообще не может выполнять свои функции на конечных интервалах длиной  $b/2$ .

## 15.6. ПРЕДСКАЗАНИЕ ЗАВИСИМОСТЕЙ



Функция Mathcad  $\text{predict}(Y, n, K)$ , где  $n$  — степень полинома аппроксимации вектора равномерно распределенных данных  $Y$ , позволяет вычислить вектор  $K$  точек предсказания (экстраполяции) поведения произвольного сигнала за пределами его задания (по возрастанию

координат  $x$ ). Предсказание тем точнее, чем более гладкую форму имеет заданный сигнал.

Пример использования функции приведен на рис. 15.6.1 для гладкой и статистически зашумленной сигнальной кривой. Степень аппроксимирующего полинома определяет глубину использования входных данных и может быть достаточно небольшой для гладких и монотонных сигналов. Ошибка прогнозирования увеличивается по мере удаления от заданных данных.

## Тема 16. МЕДИАННЫЕ ФИЛЬТРЫ

Медианные фильтры достаточно часто применяются на практике как средство предварительной обработки цифровых данных. Специфической особенностью фильтров является явно выраженная избирательность по отношению к элементам массива, представляющим собой немонотонную составляющую последовательности чисел в пределах окна (апертуры) фильтра, и резко выделяющихся на фоне соседних отсчетов. В то же время на монотонную составляющую последовательности медианный фильтр не

действует, оставляя её без изменений. Благодаря этой особенности, медианные фильтры при оптимально выбранной апертуре могут, например, сохранять без искажений резкие границы объектов, эффективно подавляя некоррелированные или слабо коррелированные помехи и малоразмерные детали. Это свойство позволяет применять медианную фильтрацию для устранения аномальных значений в массивах данных, уменьшения выбросов и импульсных помех. Характерной особенностью медианного фильтра является его нелинейность. Во многих случаях применение медианного фильтра оказывается более эффективным по сравнению с линейными фильтрами, поскольку процедуры линейной обработки являются оптимальными при равномерном или гауссовом распределении помех, что в реальных сигналах может быть далеко не так. В случаях, когда перепады значений сигналов велики по сравнению с дисперсией аддитивного белого шума, медианный фильтр дает меньшее значение среднеквадратической ошибки по сравнению с оптимальными линейными фильтрами. Особенно эффективным медианный фильтр оказывается при очистке сигналов от импульсных шумов при обработке изображений, акустических сигналов, передаче кодовых сигналов и т.п. Однако детальное исследование свойств медианных фильтров как средства фильтрации сигналов различного типа являются довольно редкими.

### 16.1. МЕДИАННАЯ ФИЛЬТРАЦИЯ ОДНОМЕРНЫХ СИГНАЛОВ

**Принцип фильтрации.** Медианы давно использовались и изучались в статистике как альтернатива средним арифметическим значениям отсчетов в оценке выборочных средних значений. Медианой числовой последовательности  $x_1, x_2, \dots, x_n$  при нечетном  $n$  является средний по значению член ряда, получающегося при упорядочивании этой последовательности по возрастанию (или убыванию). Для четных  $n$  медиану обычно определяют как среднее арифметическое двух средних отсчетов упорядоченной последовательности.

Медианный фильтр представляет собой оконный фильтр, последовательно скользящий по массиву сигнала, и возвращающий на каждом шаге один из элементов, попавших в окно (апертуру) фильтра. Выходной сигнал  $u_k$  скользящего медианного фильтра шириной  $2n+1$  для текущего отсчета  $k$  формируется из входного временного ряда  $\dots, x_{k-1}, x_k, x_{k+1}, \dots$  в соответствии с формулой:

$$u_k = \text{med}(x_{k-n}, x_{k-n+1}, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_{k+n-1}, x_{k+n}), \quad (16.1.1)$$

где  $\text{med}(x_1, \dots, x_m, \dots, x_{2n+1}) = x_{n+1}$ ,  $x_m$  – элементы вариационного ряда, т.е. ранжированные в порядке возрастания значений  $x_m$ :  $x_{(1)} = \min(x_1, x_2, \dots, x_{2n+1}) \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(2n+1)} = \max(x_1, x_2, \dots, x_{2n+1})$ .

Таким образом, медианная фильтрация осуществляет замену значений отсчетов в центре апертуры медианным значением исходных отсчетов внутри апертуры фильтра. На практике апертура фильтра для упрощения алгоритмов обработки данных, как правило, устанавливается с нечетным числом отсчетов, что и будет приниматься при рассмотрении в дальнейшем без дополнительных пояснений.

**Одномерные фильтры.** Медианная фильтрация реализуется в виде процедуры локальной обработки отсчетов в скользящем окне, которое включает определенное число отсчетов сигнала. Для каждого положения окна выделенные в нем отсчеты ранжируются по возрастанию или убыванию значений. Средний по своему положению отчет в ранжированном списке называется медианой рассматриваемой группы отсчетов. Этим отсчетом заменяется центральный отчет в окне для обрабатываемого сигнала. В силу этого медианный фильтр относится к числу нелинейных фильтров, заменяющим медианным значением аномальные точки и выбросы независимо от их амплитудных значений, и является устойчивым по определению, способным аннулировать даже бесконечно большие отсчеты.

Алгоритм медианной фильтрации обладает явно выраженной избирательностью к элементам массива с немонотонной составляющей последовательности чисел в пределах

апертуры и наиболее эффективно исключает из сигналов одиночные выбросы, отрицательные и положительные, попадающие на края ранжированного списка. С учетом ранжирования в списке медианные фильтры хорошо подавляют шумы и помехи, протяженность которых составляет менее половины окна. Стабильной точкой является последовательность (в одномерном случае) или массив (в двумерном случае), которые не изменяются при медианной фильтрации. В одномерном случае стабильными точками медианных фильтров являются "локально-монотонные" последовательности, которые медианный фильтр оставляет без изменений. Исключение составляют некоторые периодические двоичные последовательности.

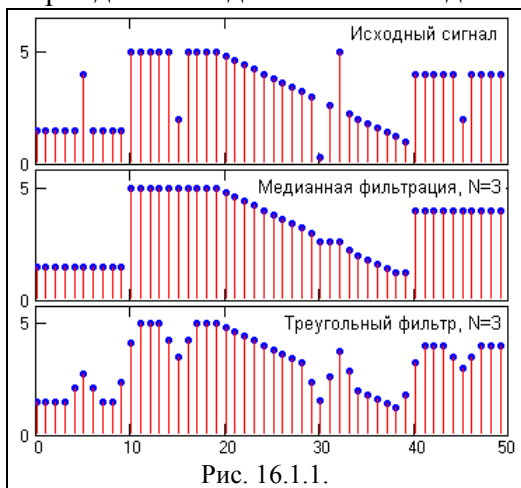


Рис. 16.1.1.

Благодаря этой особенности, медианные фильтры при оптимально выбранной апертуре могут сохранять без искажений резкие границы объектов, подавляя некоррелированные и слабо коррелированные помехи и малоразмерные детали. При аналогичных условиях алгоритмы линейной фильтрации неизбежно «смазывает» резкие границы и контуры объектов. На рис. 16.1.1 приведен пример обработки сигнала с импульсными шумами медианным и треугольным фильтрами с одинаковыми размерами окна  $N=3$ . Преимущество медианного фильтра очевидно.

В качестве начальных и конечных условий фильтрации обычно принимаются концевые значения сигналов, либо медиана находится только для тех точек, которые вписываются в пределы апертуры.

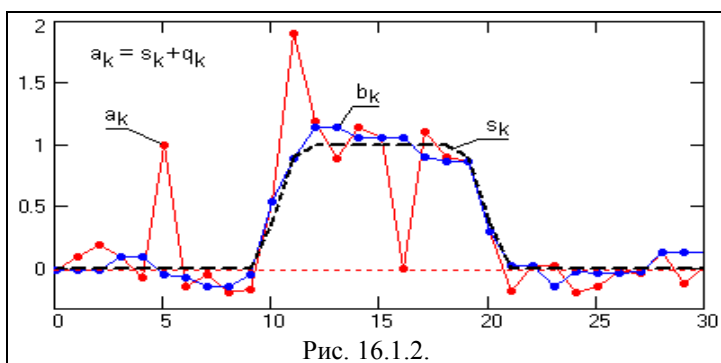


Рис. 16.1.2.

На рис. 16.1.2 приведен пример медианной фильтрации модельного сигнала  $a_k$ , составленного из детерминированного сигнала  $s_k$  в сумме со случайным сигналом  $q_k$ , имеющим

равномерное распределение с одиночными импульсными выбросами. Окно фильтра равно 5. Результат фильтрации – отсчеты  $b_k$ .

**Подавление статистических шумов** медианными фильтрами в связи с их нелинейностью обычно рассматривается только на качественном уровне. Нельзя также четко разграничить влияние медианных фильтров на сигнал и шум.

Если значения элементов последовательности чисел  $\{x_i\}$  в апертуре фильтра являются независимыми одинаково распределенными (НОР) случайными величинами со средним значением  $m$

$$x = m + z,$$

то математическое ожидание  $M\{z\} = 0$  и, следовательно,  $M\{x\} = m$ .

Пусть  $F(x)$  и  $f(x)=F'(x)$  обозначают функции распределения и плотности вероятностей величин  $x$ . Согласно теории вероятностей, распределение  $y = \text{med}(x_1, \dots, x_n)$  для больших  $n$  является приблизительно нормальным  $N(m_t, \sigma_n)$ , где  $m_t$  - теоретическая медиана, определяемая из условия  $F(m_t) = 0.5$ , при этом дисперсия распределения:

$$\sigma_n^2 = 1/(n 4f^2(m_t)). \quad (16.1.2)$$

Приведенные результаты справедливы как для одномерной, так и для двумерной фильтрации, если  $n$  выбирать равным числу точек в апертуре фильтра. Если  $f(x)$



симметрична относительно  $m$ , то распределение медиан также будет симметрично относительно  $m$  и, таким образом, справедлива формула:

$$M\{\text{med}(x_1, \dots, x_n)\} = M\{x_i\} = m.$$

Если случайные величины  $x$  являются НОР и равномерно распределены на отрезке  $[0, 1]$ , то можно найти точное значение дисперсии медианы по формуле:

$$\sigma_n^2 = 1/(4(n+2)) = 3_x/(n+2).$$

Если случайные величины  $x$  являются независимыми, одинаково распределенными с нормальным распределением  $N(m, \sigma)$ , то  $m_t = m$ . Модифицированная формула дисперсии медианы для малых нечетных значений  $n$ :

$$\sigma_g^2 \approx \pi\sigma^2/(2n-2+\pi). \quad (16.1.2')$$

Значение дисперсии шумов для случайных величин в скользящем  $n$ -окне арифметического усреднения (фильтр МНК первого порядка) имеет значение  $\sigma^2/n$ . Это означает, что для нормального белого шума при равных значениях  $n$  окон медианного фильтра и фильтра скользящего усреднения, дисперсия шумов на выходе медианного фильтра приблизительно на 57% больше, чем у фильтра скользящего среднего. Чтобы медианный фильтр давал ту же дисперсию, что и скользящее усреднение, его апертура должна быть на 57% больше. При этом следует иметь в виду, что искажение полезных сигналов, особенно при наличии в них скачков и крутых перепадов, даже при большей апертуре медианного фильтра может оказаться меньше, чем у фильтров скользящего среднего.

Положение изменяется, если плотность распределения случайных величин существенно отличается от нормального и имеет длинные хвосты, которые и ликвидируются медианным фильтром, что обеспечивает оптимальную и наиболее правдоподобную оценку текущих значений сигнала по минимуму среднеквадратического приближения. Так, при экспоненциальном (по модулю) распределении плотности шумов

$$f(x) = (\sqrt{2}/\sigma) \exp(-\sqrt{2}|x-m|/\sigma)$$

дисперсия шумов после медианного фильтра на 50% меньше, чем после фильтра скользящего среднего.

Предельным случаем таких распределений является импульсный шум, случайный по амплитудам и месту появления, который и подавляется медианными фильтрами с наибольшей эффективностью.

**Импульсные и точечные шумы.** При регистрации, обработке и обмене данными в современных измерительно-вычислительных и информационных системах потоки сигналов кроме полезного сигнала  $s(t-\tau_0)$  и флуктуационных шумов  $q(t)$  содержат, как правило, импульсные потоки  $g(t) = \sum_k \delta(t-\tau_k)$  различной интенсивности с регулярной или хаотической структурой

$$x(t) = s(t-\tau_0) + g(t) + q(t). \quad (16.1.3)$$

Под импульсным шумом понимается искажение сигналов большими импульсными выбросами произвольной полярности и малой длительности. Причиной появления импульсных потоков могут быть как внешние импульсные электромагнитные помехи, так и наводки, сбои и помехи в работе самих систем. Совокупность статистически распределенного шума и потока квазидетерминированных импульсов представляет собой комбинированную помеху. Радикальный метод борьбы с комбинированной помехой - применение помехоустойчивых кодов. Однако это приводит к снижению скорости и усложнению систем приема-передачи данных. Простым, но достаточно эффективным альтернативным методом очистки сигналов в таких условиях является двухэтапный алгоритм обработки сигналов  $x(t)$ , где на первом этапе производится устранение из потока  $x(t)$  шумовых импульсов, а на втором - очистка сигнала частотными фильтрами от статистических шумов. Для сигналов, искаженных действием импульсных шумов, отсутствует строгая в математическом смысле постановка и решение задачи фильтрации.



Известны лишь эвристические алгоритмы, наиболее приемлемым из которых является алгоритм медианной фильтрации.

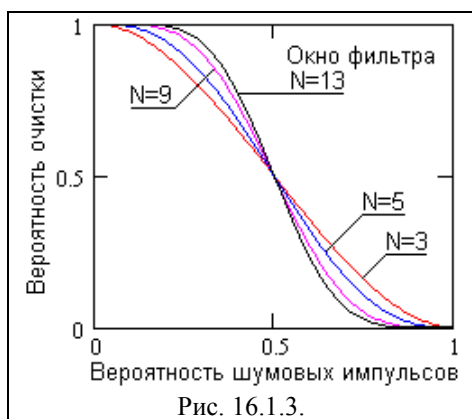
Допустим, что шум  $q(t)$  представляет собой статистический процесс с нулевым математическим ожиданием, полезный сигнал  $s(t-\tau_0)$  имеет неизвестное временное положение  $\tau_0 \in [0, T]$ , а поток шумовых импульсов  $g(t)$  имеет вид:

$$g(t) = \sum_{k=1}^K \varepsilon_k a_k g(t-\tau_k), \quad (16.1.4)$$

где  $a_k$  - амплитуда импульсов в потоке,  $\tau_k$  - неизвестное временное положение импульсов,  $\varepsilon_k=1$  с вероятностью  $p_k$  и  $\varepsilon_k=0$  с вероятностью  $1-p_k$ . Такое задание импульсной помехи соответствует потоку Бернулли /44/.

При применении к потоку  $x(t)$  скользящей медианной фильтрации с окном  $N$  отсчетов ( $N$  - нечетное) медианный фильтр полностью устраняет одиночные импульсы, удаленные друг от друга минимум на половину апертуры фильтра, и подавляет импульсные помехи, если количество импульсов в пределах апертуры не превосходит  $(N-1)/2$ . В этом случае, при  $p_k = p$  для всех импульсов помехи, вероятность подавления помех может быть определена по выражению /31/:

$$R(p) = \sum_{m=0}^{(N-1)/2} C_N^m p^m (1-p)^{N-m}. \quad (16.1.5)$$



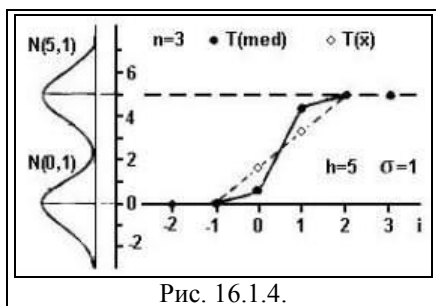
На рис. 16.1.3 приведены результаты расчетов вероятности подавления импульсной помехи медианным фильтром. При  $p < 0.5$  результаты статистического моделирования процесса показывают хорошее соответствие расчетным значениям. Для интенсивных импульсных шумовых потоков при  $p > 0.5$  медианная фильтрация становится мало эффективной, т.к. происходит не подавление, а усиление и трансформация его в поток импульсов другой структуры (со случайной длительностью).

Если вероятность ошибки не очень велика, то медианная фильтрация даже с достаточно малой апертурой значительно уменьшит число ошибок. Эффективность

исключения шумовых импульсов повышается с увеличением апертуры фильтра, но одновременно может увеличиваться и искажение полезного сигнала.

**Перепад плюс шум.** Рассмотрим фильтрацию перепадов при наличии аддитивного белого шума, т. е. фильтрацию последовательностей, или изображений, с

$$x = s + z,$$



где  $s$  - детерминированный сигнал, равный 0 по одну сторону от перепада и  $h$  — по другую, а  $z$  - случайные значения белого шума. Предположим, что случайные значения шума  $z$  распределены по нормальному закону  $N(0, \sigma)$ . Для начала рассмотрим одномерную фильтрацию и будем считать, что перепад происходит в точке  $i = 1$ , таким образом, что для  $i \leq 0$  величина  $x_i$  есть  $N(0, \sigma)$ , а для  $i \geq 1$  величина  $x_i$  есть  $N(h, \sigma)$ .

На рис. 16.1.4 показана последовательность значений математического ожидания медиан и скользящего среднего вблизи перепада высотой  $h = 5$  при  $N = 3$ . Значения скользящего среднего следуют по наклонной линии, что свидетельствует о смазывании перепада. Поведение математического ожидания значений медианы также свидетельствует о некотором смазывании, хотя в гораздо меньше, чем для скользящего среднего.

Если воспользоваться мерой среднеквадратичной ошибки (СКО), усредненной по  $N$  точкам вблизи перепада, и вычислить значения СКО в зависимости от значений  $h$ , то нетрудно зафиксировать, что при малых значениях  $h < 2$  СКО для скользящего среднего немного меньше, чем для медианы, но при  $h > 3$  СКО медианы значительно меньше, чем СКО среднего. Этот результат показывает, что скользящая медиана значительно лучше, чем скользящее среднее, для перепадов большой высоты. Похожие результаты можно получить и для апертуры  $N=5$ , и для двумерной фильтрации с апертурами  $3 \times 3$  и  $5 \times 5$ . Таким образом, математические ожидания медианы для малых  $h$  близки к математическим ожиданиям для соответствующих средних, но для больших  $h$  они асимптотически ограничены. Объясняется это тем, что при больших  $h$  (скажем,  $h > 4$ ) переменные  $x$  со средним значением 0 (для данного примера) будут резко отделены от переменных  $x$  со средним  $h$ .

Использованная мера точности может характеризовать только резкость поперек перепада и ничего не говорит о гладкости фильтрованного изображения вдоль перепада. Скользящее усреднение дает сигналы, гладкие вдоль перепада, тогда как при обработке с помощью медианного фильтра протяженные перепады оказываются слегка изрезанными.

**Ковариационные функции** при белом шуме на входе. Нормализованные функции автокорреляции выходных сигналов медианных и усредняющих фильтров подобны друг другу. Сходство функций корреляции объясняется относительно высокой корреляцией между медианой и средним, которая достигает 0.8 при больших  $N$ .

Приближенная формула функции автоковариации для последовательности, подвергнутой медианной фильтрации определяется выражением:

$$K(\tau) = \sigma^2 / (N + (\pi/2) - 1) \sum_{j=-(N-1)}^{N-1} (1 - |j|/N) \arcsin(\rho(j+\tau)). \quad (16.1.6)$$

Скользящая медиана почти не сглаживает процессы, ведущие себя на больших интервалах, как функции вида  $x_i = (-1)^i y$ . Скользящее усреднение оказывает большое сглаживающее действие на подобный процесс, так как регулярные флуктуации значений  $x$  полностью уничтожаются. В целом можно ожидать, что приближенные формулы ковариационных функций скользящих медиан будут полезны только для последовательностей, на которые медианные фильтры действуют так же, как и скользящее усреднение. В случае с сильно осциллирующими последовательностями и последовательностями перепадов большой пользы от них ждать не следует.

**Преобразование статистики шумов.** Медианная фильтрация является нелинейной операцией над входным процессом, которая наряду с исключением импульсных помех изменяет и распределение статистических шумов  $q(t)$ , что может быть нежелательным для построения последующих фильтров. Аналитический расчет преобразования статистики шумов затруднителен из-за слабой разработанности соответствующего математического аппарата.

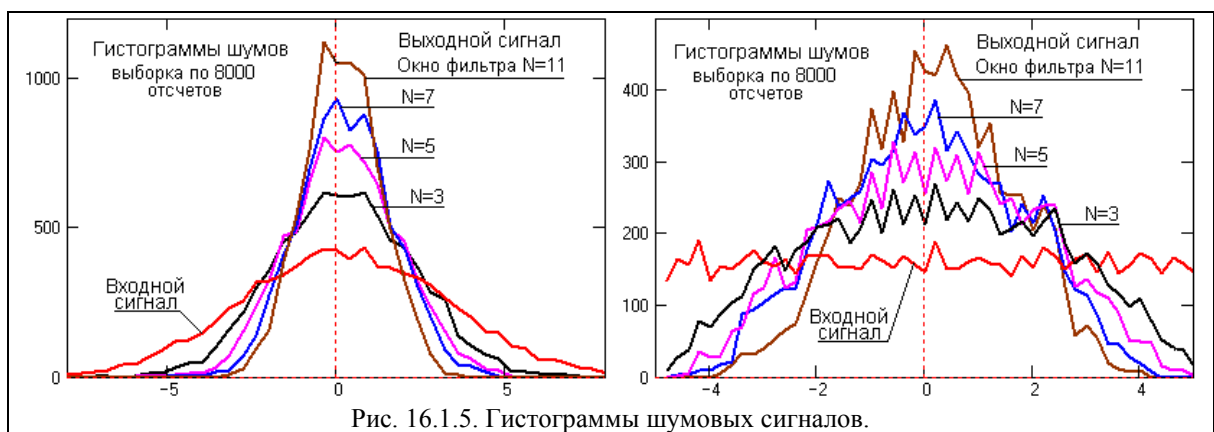
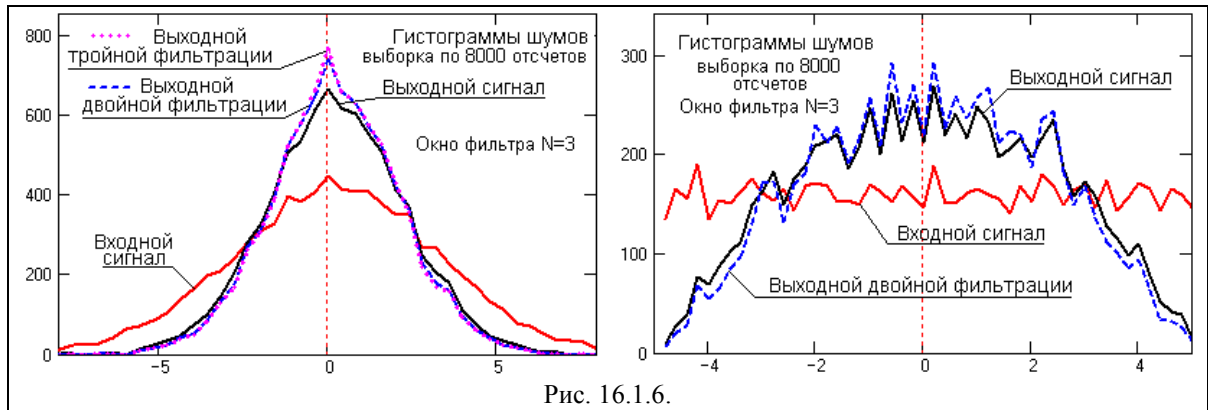


Рис. 16.1.5. Гистограммы шумовых сигналов.

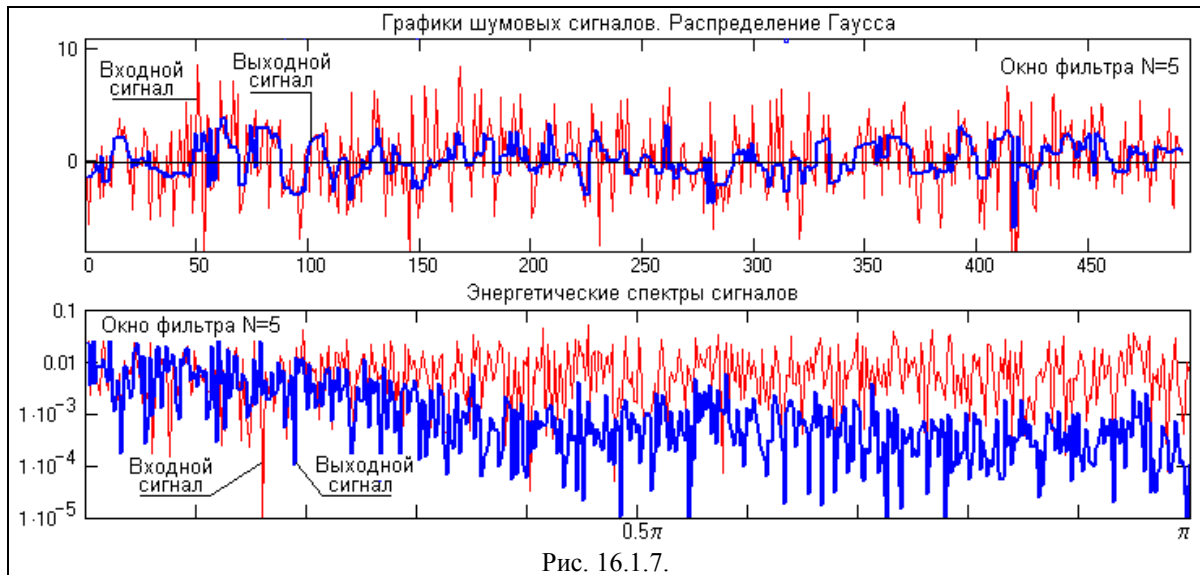
На рис. 16.1.5 приведены примеры медианной фильтрации модельных шумовых

сигналов с гауссовым и равномерным распределением при различной ширине окна фильтра. Как следует из этих графиков, при фильтрации происходит преимущественное подавление шумовых сигналов с большими отклонениями от среднего значения с уменьшением стандарта (СКО - среднеквадратического отклонения) распределения. Уменьшение стандарта тем больше, чем больше окно фильтра. Этим же определяется и преобразование формы распределения выходного равномерного шума (а равно и других распределений шумов) к гауссовой по мере увеличения размера окна фильтра.



На рис. 16.1.6 приведен пример изменения гистограмм шума при выполнении дву- и трехкратной последовательной фильтрации. Как видно из графиков, основной эффект фильтрации достигается на первом цикле.

Уменьшение количества больших шумовых отклонений от среднего значения шума приводит также к изменению спектра шума и к определенному подавлению его высокочастотных составляющих, которых больше в "хвостах" шумовых распределений. Это можно видеть на рис. 16.1.7 на спектрах плотности мощности входного и выходного сигналов.



Нелинейность медианной фильтрации (замена больших отклонений средними по рангу в окне) приводит к повышению низкочастотных составляющих спектра шума. Этот эффект наглядно виден на рис. 16.1.8, где приводятся сглаженные значения отношения модулей спектров выходного модельного шумового сигнала к входному, т.е. эквивалент коэффициента передачи фильтром шумовых сигналов. На коэффициент передачи фильтром полезных низкочастотных сигналов это не отражается, он остается равным 1, но может приводить к ухудшению отношения сигнал/шум.

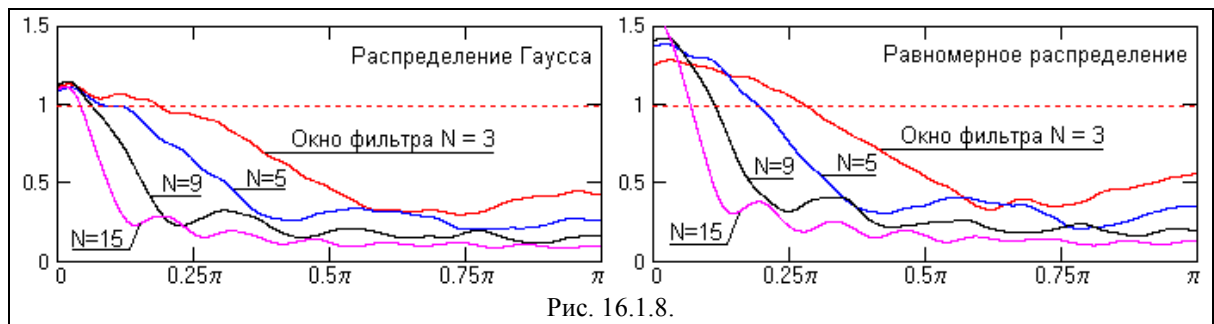


Рис. 16.1.8.

Медианный фильтр можно применять и по прямо противоположному назначению – обнаружению в сигналах и выделению квазидетерминированных помех.

**Частотные свойства фильтра.** Для описания линейных фильтров используют импульсную реакцию на единичный импульс, на ступенчатую функцию, и частотные передаточные функции в главном частотном диапазоне. Так как медианный фильтр ликвидирует единичные импульсы и сохраняет перепады, то можно говорить, что импульсная реакция фильтра равна нулю, а отклик на ступенчатую функцию равен 1. Что касается частотной характеристики фильтра, то, в силу нелинейности фильтра, ее нельзя представить какой-либо детерминированной функцией апертуры и частоты. В какой-то мере можно говорить о реакции фильтра на косинусоидальные функции, которая также существенно различается для низких и высоких частот главного частотного диапазона и фазы гармоник в апертуре фильтра, что можно видеть на рис. 16.1.9.

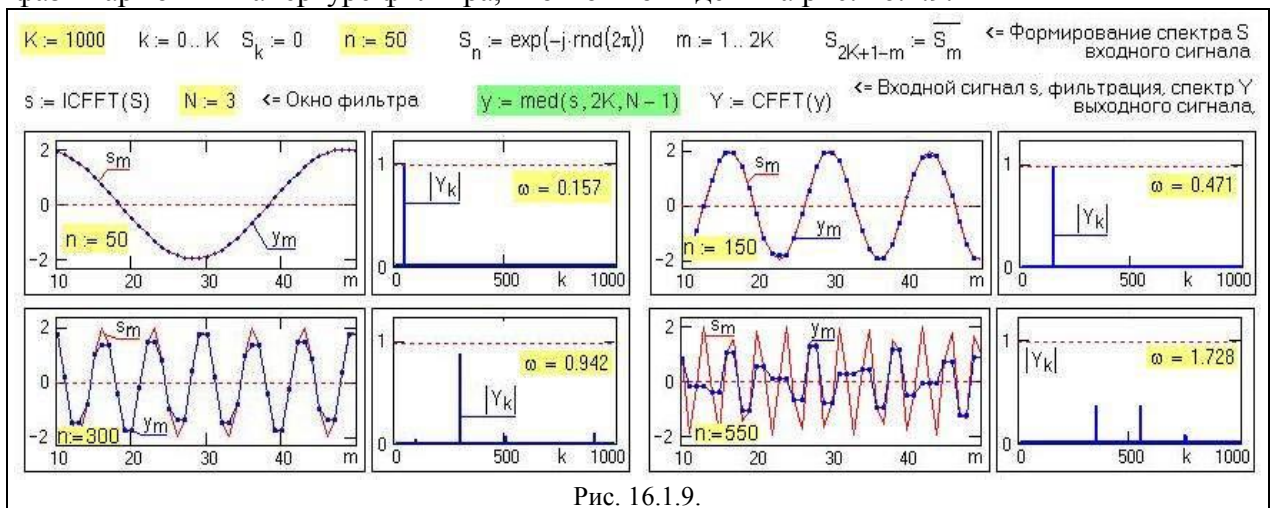


Рис. 16.1.9.

На рисунке приведено моделирование однотональных гармоник со случайной начальной фазой. Математические модели сигналов задавались в главном диапазоне спектральной области ( $0-2\pi$ , количество точек дискретизации спектра - 2000). Модуль гармоники устанавливался равным 1, при этом модуль спектра выходного сигнала после фильтрации, по-существу, отображает передаточную функцию фильтра. Окно медианного фильтра равно 3.

Как показывает моделирование, для низких частот, когда период гармоники много больше окна апертуры фильтров, скользящая медиана и скользящее среднее имеют сходные характеристики, коэффициент передачи  $K_{\Pi}$  однотональных сигналов равен 1. По мере роста частоты гармоники и в зависимости от фазы сигнала в апертуре фильтра начинается искажение сигнала на экстремальных значениях (занижение экстремальных значений), и значение  $K_{\Pi}$  начинает уменьшаться. Когда значение апертуры медианного фильтра становится соизмеримым с периодом сигнала, в спектре выходного сигнала появляются "ложные" гармоники, вызванные интерференцией частоты входного сигнала с частотой его дискретизации (нижние графики на рисунке 16.1.9).



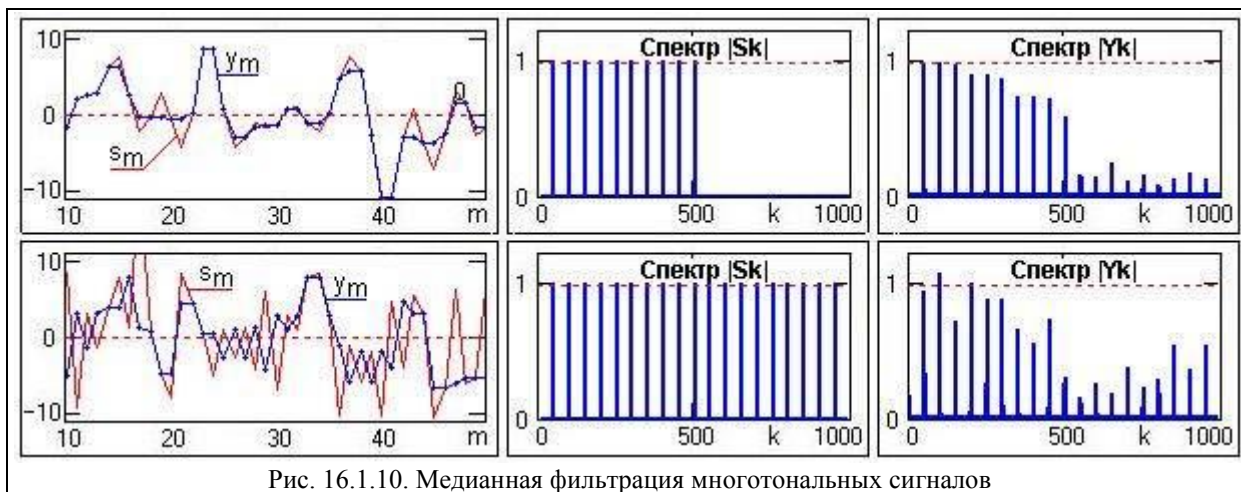


Рис. 16.1.10. Медианная фильтрация многотональных сигналов

Для многотональных входных сигналов начинается также интерференция частот гармоник между собой, что приводит к появлению многочисленных ложных высокочастотных гармоник (верхние графики на рис. 16.1.10), а при наличии во входном сигнале высокочастотных гармоник искажаются также и коэффициенты передачи низкочастотных гармоник (нижние графики на рисунке), т.е. частотные отклики для одиночных гармонических функций не соответствуют передаточным характеристикам для произвольных сигналов, являющихся суммой косинусоидальных функций, т.к. передаточные функции становятся резко нерегулярными в силу интерференции разных частот.

Картина частотной интерференции зависит также от фазы гармоник, что усиливает нерегулярность конечных результатов и наглядно видно на рис. 16.1.11 при различных случайных реализациях фазы гармоник. При увеличении размеров апертуры фильтров нерегулярность передачи фильтров увеличивается.

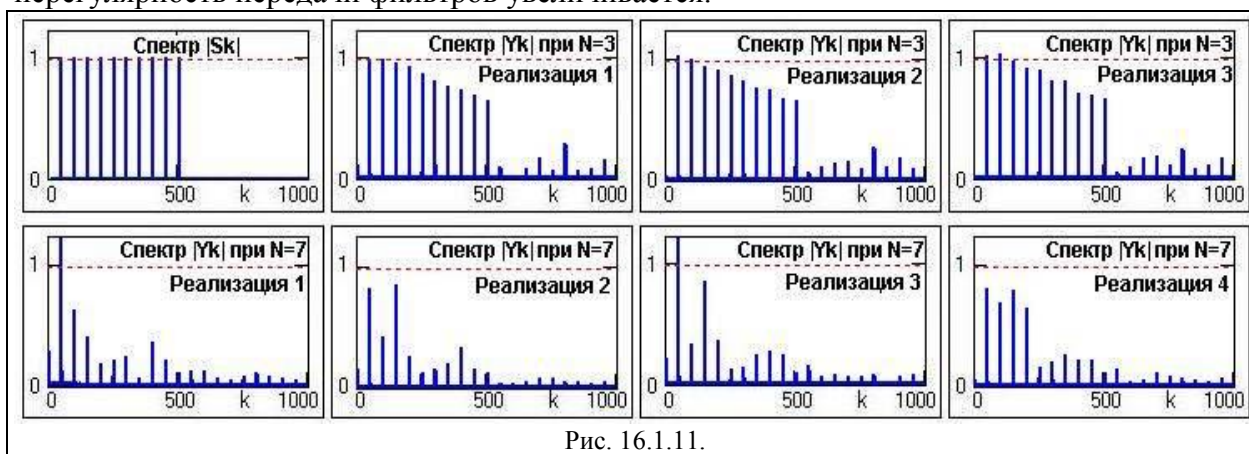


Рис. 16.1.11.

#### Разновидности медианных фильтров.

**Взвешенно-медианные фильтры** применяют, если желательно придать больший вес центральным точкам. Это достигается путем повторения  $k_i$  раз каждого набора отсчетов в апертуре фильтра. Так, например, при  $N=3$  и  $k_1=k_2=2$ ,  $k_0=3$  вычисление взвешенной медианы входного числового ряда производится по формуле:

$$y_i = \text{med} (x_{i-1}, x_{i-1}, x_0, x_0, x_0, x_1, x_1).$$

Такая растянутая последовательность также сохраняет перепады сигнала и в определенных условиях позволяет увеличить подавление дисперсии статистических шумов в сигнале. Ни один из весовых коэффициентов  $k_i$  не должен быть значительно больше всех других.

**Итерационные медианные фильтры** выполняются последовательным повторением медианной фильтрации. Если апертура единичной медианной фильтрации сохраняет перепады в сигнале, то они сохраняются при итеративном применении фильтра

вплоть до тех пор, пока не прекратятся изменения в фильтруемом сигнале, при этом конечный результат существенно отличается от итеративного применения скользящего среднего, где в пределе получается постоянная числовая последовательность. При использовании итерационных фильтров можно изменять апертуру фильтра при каждом шаге итерации.

#### **Достоинства медианных фильтров.**

- Простая структура фильтра, как для аппаратной, так и для программной реализации.
- Фильтр не изменяет ступенчатые и пилообразные функции.
- Фильтр хорошо подавляет одиночные импульсные помехи и случайные шумовые выбросы отсчетов.

#### **Недостатки медианных фильтров.**

- Медианная фильтрация нелинейна, так как медиана суммы двух произвольных последовательностей не равна сумме их медиан, что в ряде случаев может усложнять математический анализ сигналов.
- Фильтр вызывает уплощение вершин треугольных функций.
- Подавление белого и гауссового шума менее эффективно, чем у линейных фильтров. Слабая эффективность наблюдается также при фильтрации флюктуационного шума.
- При увеличении размеров окна фильтра происходит размытие крутых изменений сигнала и скачков.

Недостатки метода можно уменьшить, если применять медианную фильтрацию с адаптивным изменением размера окна фильтра в зависимости от динамики сигнала и характера шумов (адаптивная медианная фильтрация). В качестве критерия размера окна можно использовать, например, величину отклонения значений соседних отсчетов относительно центрального ранжированного отсчета  $/i/$ . При уменьшении этой величины ниже определенного порога размер окна увеличивается.

## **16.2. МЕДИАННАЯ ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ**

**Шумы в изображениях.** Никакая система регистрации не обеспечивает идеального качества изображений исследуемых объектов. Изображения в процессе формирования их системами (фотографическими, голографическими, телевизионными) обычно подвергаются воздействию различных случайных помех или шумов. Фундаментальной проблемой в области обработки изображений является эффективное удаление шума при сохранении важных для последующего распознавания деталей изображения. Сложность решения данной задачи существенно зависит от характера шумов. В отличие от детерминированных искажений, которые описываются функциональными преобразованиями исходного изображения, для описания случайных воздействий используют модели аддитивного, импульсного и мультипликативного шумов.

Наиболее распространенным видом помех является случайный аддитивный шум, статистически независимый от сигнала. Модель аддитивного шума используется тогда, когда сигнал на выходе системы или на каком-либо этапе преобразования может рассматриваться как сумма полезного сигнала и некоторого случайного сигнала. Модель аддитивного шума хорошо описывает действие зернистости фотопленки, флюктуационный шум в радиотехнических системах, шум квантования в аналого-цифровых преобразователях и т.п.

Аддитивный гауссов шум характеризуется добавлением к каждому пикселю изображения значений с нормальным распределением и с нулевым средним значением. Такой шум обычно появляется на этапе формирования цифровых изображений. Основную информацию в изображениях несут контуры объектов. Классические линейные фильтры способны эффективно удалить статистический шум, но степень размытости мелких деталей на изображении может превысить допустимые значения. Для решения этой

проблемы используются нелинейные методы, например алгоритмы на основе анизотропной диффузии Перрона и Малика, билатеральные и трилатеральные фильтры. Суть таких методов заключается в использовании локальных оценок, адекватных определению контура на изображении, и сглаживания таких участков в наименьшей степени.

Импульсный шум характеризуется заменой части пикселей на изображении значениями фиксированной или случайной величины. На изображении такие помехи выглядят изолированными контрастными точками. Импульсный шум характерен для устройств ввода изображений с телевизионной камеры, систем передачи изображений по радиоканалам, а также для цифровых систем передачи и хранения изображений. Для удаления импульсного шума используется специальный класс нелинейных фильтров, построенных на основе ранговой статистики. Общей идеей таких фильтров является детектирование позиции импульса и замена его оценочным значением, при сохранении остальных пикселей изображения неизменными.

**Двумерные фильтры.** Медианная фильтрация изображений наиболее эффективна, если шум на изображении имеет импульсный характер и представляет собой ограниченный набор пиковых значений на фоне нулей. В результате применения медианного фильтра наклонные участки и резкие перепады значений яркости на изображениях не изменяются. Это очень полезное свойство именно для изображений, на которых контуры несут основную информацию.

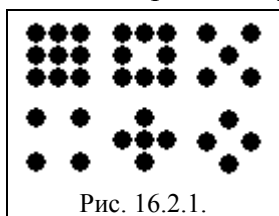


Рис. 16.2.1.

При медианной фильтрации зашумленных изображений степень сглаживания контуров объектов напрямую зависит от размеров апертуры фильтра и формы маски. Примеры формы масок с минимальной апертурой приведены на рис. 16.2.1. При малых размерах апертуры лучше сохраняются контрастные детали изображения, но в меньшей степени подавляются импульсные шумы. При больших размерах апертуры наблюдается обратная картина. Оптимальный выбор формы сглаживающей апертуры зависит от специфики решаемой задачи и формы объектов. Особое значение это имеет для задачи сохранения перепадов (резких границ яркости) в изображениях.

Под изображением перепада понимаем изображение, в котором точки по одну сторону от некоторой линии имеют одинаковое значение  $a$ , а все точки по другую сторону от этой линии - значение  $b$ ,  $b \neq a$ . Если апертура фильтра симметрична относительно начала координат, то медианный фильтр сохраняет любое изображение перепада. Это выполняется для всех апертур с нечетным количеством отсчетов, т.е. кроме апертур (квадратные рамки, кольца), которые не содержат начала координат. Тем не менее квадратные рамки и кольца будут лишь незначительно изменять перепад.

Для упрощения дальнейшего рассмотрения ограничимся примером фильтра с квадратной маской размером  $N \times N$ , при  $N=3$ . Скользящий фильтр просматривает отсчеты изображения слева-направо и сверху-вниз, при этом входную двумерную последовательность также представим в виде последовательного числового ряда отсчетов  $\{x(n)\}$  слева-направо сверху-вниз. Из этой последовательности в каждой текущей точке маска фильтра выделяет массив  $w(n)$ , как  $W$ -элементный вектор, который в данном случае содержит все элементы из окна  $3 \times 3$ , центрированные вокруг  $x(n)$ , и сам центральный элемент, если это предусмотрено типом маски:

$x_1(n)$	$x_2(n)$	$x_3(n)$
$x_4(n)$	$x(n)$	$x_5(n)$
$x_6(n)$	$x_7(n)$	$x_8(n)$

Рис. 16.2.2.

$$w(n) = [x_1(n), x_2(n), \dots, x_w(n)]. \quad (16.2.1)$$

В этом случае значения  $x_i$  соответствует отображению слева-направо и сверху-вниз окна  $3 \times 3$  в одномерный вектор, как показано на

рис. 16.2.2.

Элементы данного вектора, как и для одномерного медианного фильтра, также

могут быть упорядочены в ряд по возрастанию или убыванию своих значений:

$$r(n) = [r_1(n), r_2(n), \dots, r_w(n)], \quad (16.2.2)$$

определено значение медианы  $y(n) = \text{med}(r(n))$ , и центральный отсчет маски заменен значением медианы. Если по типу маски центральный отсчет не входит в число ряда 16.2.1, то медианное значение находится в виде среднего значения двух центральных отсчетов ряда 16.2.2.

Приведенные выражения не объясняют способа нахождения выходного сигнала вблизи конечных и пограничных точек в конечных последовательностях и изображениях. Один из простых приемов состоит в том, что нужно находить медиану только тех точек внутри изображения, которые попадают в пределы апертуры. Поэтому для точек, расположенных рядом с границами, медианы будут определены, исходя из меньшего числа точек.

На рис. 16.2.3 приведен пример очистки зашумленного изображения медианным фильтром Черненко /2i/. Зашумление изображения по площади составляло 15%, для очистки фильтр применен последовательно 3 раза.

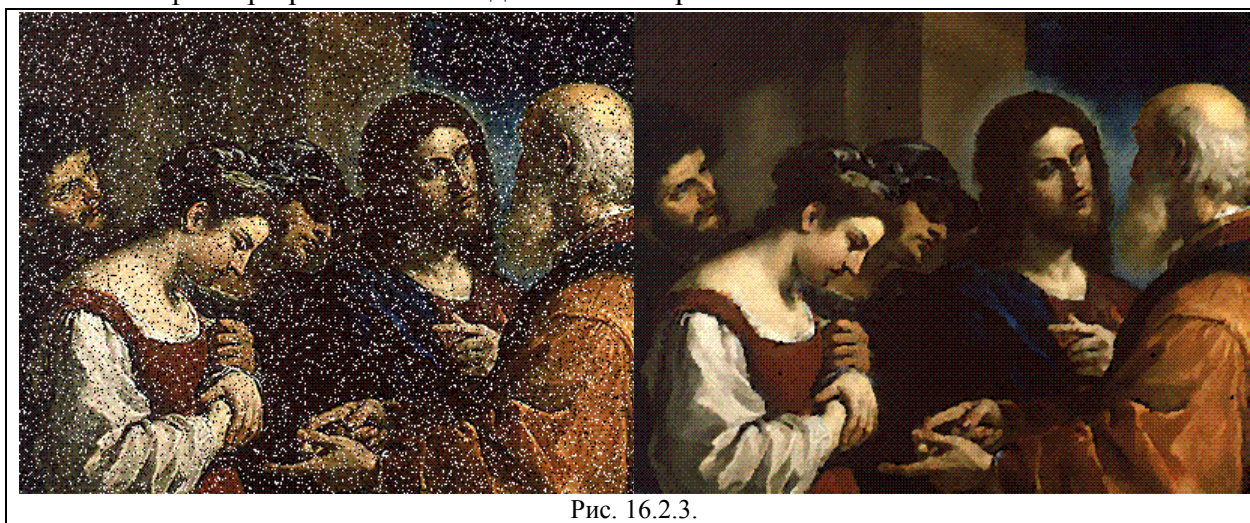


Рис. 16.2.3.

Медианная фильтрация может выполняться и в рекурсивном варианте, при котором значения сверху и слева от центрального отсчета в маске (в данном случае  $x_1(n)$ - $x_4(n)$  на рис. 16.2.2) в ряде 16.2.1 заменяются на уже вычисленные в предыдущих циклах значения  $y_1(n)$ - $y_4(n)$ .

**Адаптивные двумерные фильтры.** Противоречие по зависимости степени подавления шумов и искажения сигнала от апертуры фильтра в некоторой степени сглаживается при применении фильтров с динамическим размером маски, с адаптацией размеров апертуры под характер изображения. В адаптивных фильтрах большие апертуры используются в монотонных областях обрабатываемого сигнала (лучшее подавление шумов), а малые – вблизи неоднородностей, сохраняя их особенности, при этом размер скользящего окна фильтра устанавливается в зависимости от распределения яркости пикселей в маске фильтра. В их основе лежит, как правило, анализ яркости окрестностей центральной точки маски фильтра.

Простейшие алгоритмы динамического изменения апертуры фильтра, симметричного по обеим осям, обычно работают по заданному на основании эмпирических данных пороговому коэффициенту яркости  $S_{\text{порог}} = [0, 1]$ . В каждом текущем положении маски на изображении итерационный процесс начинается с апертуры минимального размера. Величины отклонения яркости соседних пикселей  $A(r, n)$ , попавших в окно размером  $(n \times n)$ , относительно яркости центрального отсчета  $A(r)$  вычисляются по формуле:

$$S_n(r) = |A(r,n)/A(r) - 1|. \quad (16.2.3)$$

Критерий, согласно которому производится увеличение размера маски с



центральным отсчетом  $g$  и выполняется следующая итерация, имеет вид:

$$\max[S_n(r)] < S_{\text{порог}}. \quad (16.2.4)$$

Максимальный размер маски (количество итераций), как правило, ограничивается. Для неквадратных масок, имеющих размеры  $(n \times m)$ , итерации могут вычисляться с раздельным увеличением параметров  $n$  и  $m$ , а также с изменением формы масок в процессе итераций.

**Фильтры на основе ранговой статистики.** В последние два десятилетия в цифровой обработке изображений активно развиваются нелинейные алгоритмы на основе ранговой статистики для восстановления изображений, поврежденных различными моделями шумов. Подобные алгоритмы позволяют избежать дополнительного искажения изображения при удалении шума, а также значительно улучшить результаты работы фильтров на изображениях с высокой степенью зашумленности.

Сущность ранговой статистики обычно заключается в том, что ряд 16.2.1 не включает центральный отсчет маски фильтра, и по ряду 16.2.2 производится вычисление значения  $m(n)$ . При  $N=3$  по рис. 16.2.2:

$$m(n) = (x_4(n) + x_5(n))/2. \quad (16.2.5)$$

Вычисление выходного значения фильтра, которым заменяется центральный отсчет, выполняется по формуле:

$$y(n) = \alpha x(n) + (1-\alpha) m(n). \quad (16.2.6)$$

Значение коэффициента доверия  $\alpha$  связывается определенной зависимостью со статистикой отсчетов в окне фильтра (например, полной дисперсией отсчетов, дисперсией разностей  $x(n) - x_i(n)$  или  $m(n) - x_i(n)$ , дисперсией положительных и отрицательных разностей  $x(n) - x_i(n)$  или  $m(n) - x_i(n)$ , и т.п.). По существу, значение коэффициента  $\alpha$  должно задавать степень поврежденности центрального отсчета и, соответственно, степень заимствования для его исправления значения из отсчетов  $m(n)$ . Выбор статистической функции и характер зависимости от нее коэффициента  $\alpha$  может быть достаточно многообразным и зависит как от размеров апертуры фильтра, так и от характера изображений и шумов.

## Тема 17. ОБРАБОТКА ИЗОБРАЖЕНИЙ

Размах исследований в области цифровой обработки изображений стремительно нарастает. Это определяется тем, что обработка изображений — это обработка многомерных сигналов, а большинство сигналов в реальном мире является многомерными.

Изображение в математическом представлении - двумерный сигнал, несущий огромное количество информации. Цветное изображение размером  $500 \times 500$  элементов - это массив в несколько сотен тысяч байтов. Обработать такую информацию можно лишь рациональной организацией вычислений. Для конкретных задач обработки изображений можно применять эффективные способы обработки с учетом особенностей и ограничений этой конкретной задачи. Но если говорить об обработке изображений для решения широкого класса задач, то необходимо выделить набор стандартных операций, из которых можно строить алгоритмы для решения произвольных задач. К их числу относятся линейные преобразования, двумерная свертка и двумерное дискретное преобразование Фурье.

Но при обработке изображений широкое использование находят и нелинейные преобразования. Особенность изображений состоит в том, что отдельные элементы изображения находятся в определенной связи с соседними элементами. Поэтому большинство алгоритмов преобразования изображений носит локальный характер, т. е. обрабатывают изображения по группам элементов, располагающихся в окрестности вокруг данного. Линейные преобразования удовлетворяют свойству локальности и

допускают построение алгоритмов, вычислительная сложность которых мало зависит от размеров охватываемой окрестности. Такие же свойства требуются и от нелинейных преобразований изображений. К классу таких преобразований относятся алгоритмы, которые называют алгоритмами ранговой фильтрации, основанными на вычислении локальных ранговых статистик изображений. При вычислении ранговых статистик и производных от них возможны упрощения, связанные с информационной избыточностью изображений. Наиболее известный алгоритм этого класса — алгоритм медианной фильтрации. Другими примерами ранговых алгоритмов могут служить алгоритмы экстремальной фильтрации, которые заменяют анализируемый элемент изображения максимумом или минимумом по окрестности. Еще одно свойство ранговых алгоритмов - локальная адаптация к характеристикам обрабатываемого изображения и потенциальные возможности их использования не только для сглаживания и очистки от шумов, но и для выделения признаков при автоматическом распознавании изображений.

При обработке изображений широко используются методы обработки одномерных сигналов, если возможно их обобщение на многомерные сигналы. При этом, приходится учитывать, что математические методы описания многомерных систем не отличаются завершённостью. Многомерные системы обладают большим числом степеней свободы, и их проектирование приобретает гибкость, не свойственную одномерным системам. В то же время, многомерные полиномы не разлагаются на простые множители, что усложняет анализ и синтез многомерных систем.

### 17.1. ОСНОВНЫЕ ПОНЯТИЯ

**Графическое представление изображений.** Для представления графической информации на двумерной плоскости (экране монитора) применяются два подхода: растровый и векторный.

При векторном подходе графическая информация описывается как совокупность абстрактных геометрических объектов - прямые, отрезки, кривые, прямоугольники и т.п. Векторное описание предполагает априорные знания о структуре изображения.

Растровая графика оперирует с произвольными изображениями в виде растров. Растр (raster) - это описание изображения на плоскости путем разбиения (дискретизации) его на одинаковые элементы по регулярной сетке и присвоение каждому элементу своего цветового и любых других атрибутов. Самый простой растр – прямоугольный, самый экономичный по количеству отсчетов для передачи изображений - гексагональный. С математических позиций растр – это кусочно-постоянная аппроксимация на плоскости непрерывной функции изображения.

Элемент растра называют пикселем (pixel). Стандартная идентификация пикселей:

$$f(i, j) = (A(i, j), C(i, j)), \quad (17.1.1)$$

где  $A(i, j) \subset R^2$  - область пикселя,  $C(i, j) \in C$  - атрибут пикселя (как правило, цвет). Чаще всего используются два вида атрибутов:

$C(i, j) = I(i, j)$  - интенсивность (яркость) пикселя;

$C(i, j) = \{R(i, j), G(i, j), B(i, j)\}$  - цветовые атрибуты в цветовой модели RGB.

В матричной форме:

$$M_{ij} = (A_{ij}, C_{ij}).$$

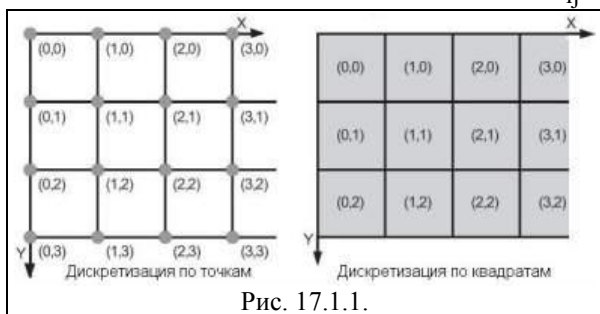


Рис. 17.1.1.

При дискретизации непрерывных изображений значения  $A_{ij}$  могут определяться двояко, либо как значения точек  $A_{ij} = (i, j)$ , для которых определены атрибуты  $C_{ij}$ , либо как значения квадратов  $A_{ij} = (i, i+1) \times (j, j+1)$  или любой другой формы, с определением  $C_{ij}$  по средним значениям в пределах этой формы (рис. 17.1.1).

На практике, как правило, X и Y - ограниченные наборы неотрицательных целых чисел квадратного или прямоугольного растра с аспектвым отношением (aspect ratio) ширины к высоте растра, которое записывается в виде, например "4:3".

**Представление цвета в машинной графике.** Понятие цвета базируется на восприятии глазами человека электромагнитных волн в определенном диапазоне частот. Воспринимаемый нами дневной свет имеет длины волн  $\lambda$  от 400 нм (фиолетовый) до 700 нм (красный). Описанием светового потока может служить его спектральная функция  $I(\lambda)$ . Свет называется монохроматическим, если его спектр имеет только одну определенную длину волны.

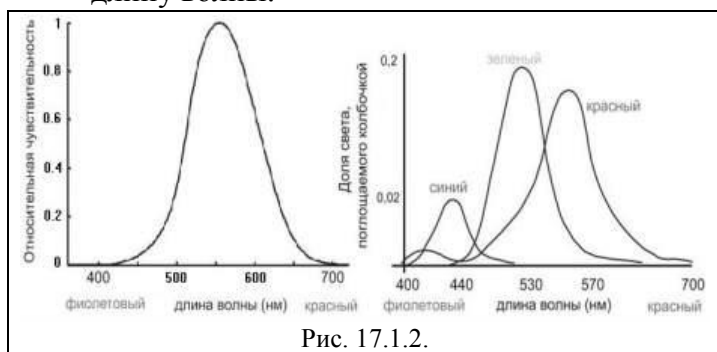


Рис. 17.1.2.

На сетчатке глаза находятся два типа рецепторов: палочки и колбочки. Спектральная чувствительность палочек (рис. 17.1.2) прямо пропорциональна яркости падающего света. Колбочки разделяются на три вида, каждый из которых имеет определенную чувствительность в ограниченных диапазонах с максимумами к красному, зеленому и синему цветам, и резко теряют

свою чувствительность в темноте. Восприимчивость глаза к синему цвету значительно ниже, чем к двум другим. Важным свойством восприятия света человеком является линейность при сложении цветов с разными длинами волн.

**Цветовая модель RGB** (Red, Green, Blue - красный, зеленый, голубой) в машинной графике в настоящее время является самой распространенной. В этой модели спектральная функция представляется как сумма кривых чувствительности для каждого типа колбочек с неотрицательными весовыми коэффициентами (с нормировкой от 0 до 1), которые так и обозначаются - R, G и B. Модель характеризуется свойством аддитивности для получения новых цветов. К примеру, кодировка спектральных функций:

- черного цвета:  $f_{\text{black}} = 0$ ,  $(R,G,B) = (0,0,0)$ ;
- фиолетового цвета  $f_{\text{violet}} = f_{\text{red}} + f_{\text{blue}}$ ,  $(R,G,B) = (1,0,1)$ ;
- белого цвета  $f_{\text{white}} = f_{\text{red}} + f_{\text{green}} + f_{\text{blue}}$ ,  $(R,G,B) = (1,1,1)$ .

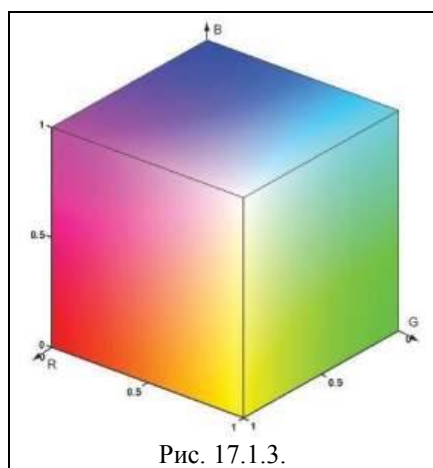


Рис. 17.1.3.

Трехмерное пространство цветов модели RGB приведено на рис. 17.1.3. В силу особенностей восприятия света рецепторами не все цвета, видимые человеком, представимы в этой модели. Однако доля воспроизводимых цветов значительно больше, чем доля не представимых в этой модели.

**Цветовая система CIE XYZ.** Международный стандарт представления цвета CIE (CIE - Commission Internationale de l'Eclairage) был принят в 1931 году Международной комиссией по освещению. В нем определяются три базисные функции  $\rho_x(\lambda)$ ,  $\rho_y(\lambda)$ ,  $\rho_z(\lambda)$ , зависящие от длины волны, линейные комбинации которых с неотрицательными коэффициентами (X, Y и Z)

позволяют получить все видимые человеком цвета. Этими функциями учитывается относительное восприятие интенсивности света рецепторами глаза. В трехмерном пространстве цветовая система CIE образует конус в первом квадранте и применяется для высококачественного отображения цветных изображений.

## 17.2. ГЕОМЕТРИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ РАСТРОВЫХ ИЗОБРАЖЕНИЙ

**Области и этапы преобразований.** Изображения можно разделить на текстурные и детальные. В текстурных изображениях все отсчёты (элементы) несут информацию

(изображение на экране телевизора). Детальным называется изображение, на котором можно выделить мешающие объекты, фон и полезные объекты.

Существуют три основные группы алгоритмов обработки изображений на компьютерах:

1. Первичная (предварительная) обработка изображений с целью реставрации, очистки от случайных шумов, улучшения качества, коррекции геометрических искажений оптических систем (расфокусировка, аберрации и пр.).

2. Описание изображений, распознавание образов. Выполняется для определения параметров деталей изображения и включает: нахождение однородных по уровню освещённости и цвету областей изображения, выделение признаков формы изображений, определение координат особых точек объектов и пр.

3. Эффективное кодирование для уменьшения объема при передаче и хранении.

Большинство методов первичной обработки основаны на использовании линейных пространственно-инвариантных (ЛПИ) фильтров. Линейные алгоритмы выполняются с помощью двумерных аналогов одномерных КИХ и БИХ фильтров. Их можно применять, например, при реализации фильтров для снижения уровня шума на изображениях.

КИХ фильтры реализуются методом свёртки. Преимуществом двумерных КИХ фильтров является наглядность, простота и абсолютная устойчивость. БИХ фильтры реализуются с помощью разностных уравнений и z-преобразований. Они более скоростные по сравнению с КИХ фильтрами, но могут оказаться неустойчивыми. Синтез двумерных БИХ фильтров отличается от синтеза одномерных, так как для двумерной функции в явном виде не удаётся выделить полюса.

Для реставрации изображений и улучшения их качества могут потребоваться и нелинейные методы. Так, например, чтобы подавить шум и при этом сохранить контурную часть изображений, приходится применять нелинейные или линейные пространственно-неинвариантные (ЛПНИ) фильтры, которые реализуются ранговыми алгоритмами. Все ранговые нелинейные фильтры основаны на быстрых алгоритмах вычисления локальных гистограмм.

Одним из таких методов является медианная фильтрация. Применение медианных фильтров эффективно для подавления некоторых видов шума и периодических помех без одновременного искажения сигнала, например, для подавления пачек шумовых выбросов, включая выпадение строк. Метод может применяться также при решении задач, связанных с распознаванием, например, для выделения тонких линий и небольших изолированных объектов.

Алгоритмы описания изображений и распознавания образов, как правило, нелинейны и носят эвристический характер. Признаками объектов обычно являются площадь изображения объекта, периметр контура изображения, отношение площади к квадрату периметра изображения. Форму объекта можно охарактеризовать радиусом вписанной в изображение или описанной вокруг изображения объекта окружности, длиной минимального и максимального радиус-вектора от “центра масс” изображения.

**Дискретизация.** Преобразования изображений в компьютере и хранение обработанных данных выполняются в дискретном виде. Для получения дискретного представления из непрерывных аналоговых изображений реального мира применяется дискретизация (sampling). Практически ее осуществляют устройства ввода (цифровой фотоаппарат, сканер или другие). Для визуального восприятия обработанных изображений на устройствах вывода (дисплей, плоттер и др.) осуществляется реконструкция аналогового изображения по его дискретизированному представлению.

В простейшем случае черно-белых изображений мы имеем двумерный массив  $s_a(x, y)$ . Для цветных изображений в модели RGB, учитывая свойство аддитивности при сложении цветов, каждый слой R, G и B также может рассматриваться и обрабатываться, как двумерный массив, с последующим суммированием результатов.

Из способов обобщения одномерной периодической дискретизации на двумерный случай наиболее простым является периодическая дискретизация в прямоугольных координатах:

$$s(n,m) = s_a(n\Delta x, m\Delta y),$$

где  $\Delta x$  и  $\Delta y$  - горизонтальный и вертикальный интервалы дискретизации двумерного непрерывного сигнала  $s_a(x,y)$  с непрерывными координатами  $x$  и  $y$ . Ниже значения  $\Delta x$  и  $\Delta y$ , как и в одномерном случае, принимаются равными 1.

Дискретизация двумерного сигнала также приводит к периодизации его спектра и наоборот. Сохраняется и условие информационной равноценности координатного и частотного представлений дискретного сигнала при равном количестве точек дискретизации в главных диапазонах сигнала. Для прямоугольной дискретизации прямое и обратное преобразование Фурье определяются выражениями:

$$S(k,l) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} s(n,m) \exp(-jn2\pi k/N - jm2\pi l/M), \quad (17.2.1)$$

$$S(k,l) = \sum_{n=0}^{N-1} \exp(-jn2\pi k/N) \sum_{m=0}^{M-1} s(n,m) \exp(-jm2\pi l/M), \quad (17.2.1')$$

$$s(n,m) = \frac{1}{NM} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} S(k,l) \exp(-jn2\pi k/N - jm2\pi l/M). \quad (17.2.2)$$

$$s(n,m) = \frac{1}{NM} \sum_{k=0}^{N-1} \exp(-jn2\pi k/N) \sum_{l=0}^{M-1} S(k,l) \exp(-jm2\pi l/M). \quad (17.2.2')$$

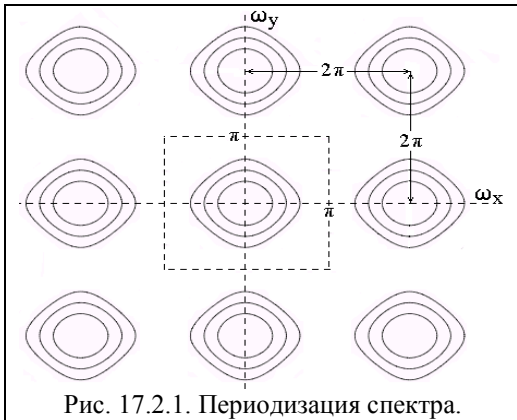


Рис. 17.2.1. Периодизация спектра.

Эти выражения показывают, что двумерное ДПФ по прямоугольному растру дискретизации данных может вычисляться с помощью одномерных последовательных ДПФ. Вторые суммы выражений (17.2.1') и (17.2.2') являются одномерными ДПФ сечений функций  $s(n,m)$  и  $S(k,l)$  по линиям  $n$  и  $k$  соответственно, а первые - одномерными ДПФ вычисленных функций в сечениях по  $m$  и  $l$ . Другими словами, исходные матрицы значений  $s(n,m)$  и  $S(k,l)$  пересчитываются сначала в промежуточные матрицы с ДПФ по строкам (или по столбцам), а промежуточные - в окончательные с ДПФ по столбцам (или соответственно по строкам).

Для того чтобы периодическое повторение спектра (рис. 17.2.1), вызванное дискретизацией аналогового сигнала с частотой  $F_x=1/\Delta x$  и  $F_y=1/\Delta y$ , не изменяло спектр в главном частотном диапазоне (по отношению к спектру исходного аналогового сигнала), необходимо и достаточно, чтобы максимальные частотные составляющие  $f_{max}$  в спектре аналогового сигнала как по строкам, так и по столбцам, не превышали частоты Найквиста ( $f_{max,x} \leq f_N = F_x/2$ ,  $f_{max,y} \leq f_M = F_y/2$ ). Это означает, что частота дискретизации сигнала должна быть минимум в два раза выше максимальной частотной составляющей в спектре сигнала:

$$F_x \geq 2f_{max,x}, \quad F_y \geq 2f_{max,y}, \quad (17.2.3)$$

что обеспечивает выход спектральных функций на нулевые значения на концах главного диапазона спектра.

**Интерполяционный ряд восстановления двумерного сигнала.** Если непрерывный сигнал  $s_a(x,y)$  является сигналом с ограниченным спектром, а периоды дискретизации выбраны достаточно малыми и спектры соседних периодов не перекрываются:

$$S_a(\Omega_x, \Omega_y) = 0 \quad \text{при } |\Omega_x| \geq \pi/\Delta x, \quad |\Omega_y| \geq \pi/\Delta x,$$

то, как и в одномерном случае, сигнал  $s_a(x,y)$  может быть восстановлен по дискретному сигналу с использованием двумерного аналога ряда Котельникова-Шеннона:

$$s_a(x,y) = \sum_n \sum_m s(n,m) \frac{\sin \pi(x/\Delta x - n)}{\pi(x/\Delta x - n)} \frac{\sin \pi(y/\Delta y - m)}{\pi(y/\Delta y - m)}. \quad (17.2.4)$$

**Частотные искажения изображений и их устранение.** Сигнал с неограниченным спектром также может быть дискретизирован, однако в этом случае имеет место наложение спектров в смежных периодах, при этом высокие частоты, большие частот Найквиста, будут "маскироваться", как и в одномерном случае, под низкие частоты главного периода. Эффект "отражения" от границ периода дает еще более сложную картину вследствие интерференции частот, отраженных по разным координатам. Аналогичный эффект, известный как алиасинг (aliasing) будет наблюдаться и при недостаточной частоте дискретизации изображений. Особенно наглядно этот эффект можно наблюдать на резких контрастных изменениях яркости.

Для борьбы с подобными явлениями применяют префильтрацию (антиалиасинг) – предварительную свертку аналогового изображения с весовой функцией фильтра, отсекающей высокочастотные компоненты, которые могут привести к алиасингу. В двумерном случае фильтрация описывается следующим образом:

$$z(x, y) = h(x', y') \otimes \otimes s(x-x', y-y'). \quad (17.2.5)$$

Следует заметить, что аналоговые изображения существуют только в оптическом диапазоне, например, в виде светового отображения экране, фотобумаге или фотопленке, но не могут существовать в памяти компьютера. Поэтому физическое выполнение префильтрации возможно только при регистрации изображения путем его расфокусировки, что, как правило, не применяется. Первичная информация всегда должна регистрироваться с максимальной полнотой и точностью, а очистка первичной информации от излишних подробностей и избыточности – дело последующей обработки данных. Поэтому применительно к уравнению 17.2.5 двумерная префильтрация, в ее практическом исполнении, может представлять собой только фильтрацию изображений, дискретизированных с большим запасом по главному частотному диапазону (с излишней разрешающей способностью), и применяется, как правило, при передискретизации на более крупный шаг, например, при сжатии изображений. Префильтрация может встраиваться также в алгоритмы построения изображений.



Рис. 17.2.2.

На практике при этом обычно используются двумерные аналоги весовых функций, применяемых при обработке одномерных сигналов (Ганна, Хемминга, Гаусса и др.). Как правило, фильтры  $h(x', y')$  имеют либо радиальную, либо осевую симметрию. На рис. 17.2.2

приведен пример исходного изображения и изображения после выполнения префильтрации.





Рис. 17.2.3. Примеры весовых функций.

На рис. 17.2.3 и ниже, в таблице 17.2.1 приведены примеры наиболее распространенных одномерных фильтров для антиалисинга. Они могут выполняться и в виде аналоговых фильтров, и применяться, например, при передаче телевизионных строк изображений в аналоговой форме по радиоканалам (антиалисинг по горизонтали). В принципе, подобная же операция может выполняться и по столбцам

(дубль- изображение), и после суммирования изображения будет выполнена полная операция антиалисинга, но такой метод относится больше к области специальных научных исследований.

Таблица 17.2.1.

Основные весовые функции

Временное окно	Весовая функция	Фурье-образ
Естественное (Π)	$\Pi(t) = 1,  t  \leq \tau; \Pi(t) = 0,  t  > \tau$	$\Pi(\omega) = 2\tau \operatorname{sinc}[\omega\tau]$
Бартлетта (Δ)	$b(t) = 1 -  t /\tau$	$B(\omega) = \tau \operatorname{sinc}^2(\omega\tau/2)$
Хеннинга, Ганна	$p(t) = 0.5[1 + \cos(\pi t/\tau)]$	$0.5\Pi(\omega) + 0.25\Pi(\omega + \pi/\tau) + 0.25\Pi(\omega - \pi/\tau)$
Хемминга	$p(t) = 0.54 + 0.46 \cos(\pi t/\tau)$	$0.54\Pi(\omega) + 0.23\Pi(\omega + \pi/\tau) + 0.23\Pi(\omega - \pi/\tau)$
Карре (2-е окно)	$p(t) = b(t) \operatorname{sinc}(\pi t/\tau)$	$\tau \cdot B(\omega) * \Pi(\omega), \Pi(\omega) = 1$ при $ \omega  < \pi/\tau$
Лапласа-Гаусса	$p(t) = \exp[-\beta^2(t/\tau)^2/2]$	$[(\tau/\beta) \sqrt{2\pi} \exp(-\tau^2\omega^2/(2\beta^2))] \otimes \Pi(\omega)$

Двумерные аналоги одномерных фильтров  $f_1(x)$  строятся в двух вариантах симметрии: или как функция от радиуса:

$$f_2(x, y) = f_1(\sqrt{x^2 + y^2}),$$

или как произведение:

$$f_2(x, y) = f_1(x) \times f_1(y).$$

Первый вариант - более корректный, но второй обладает свойством сепарабельности, т.е. двумерную свертку можно выполнять двумя одномерными свертками последовательно по строкам с  $f_1(x)$  и по столбцам с  $f_1(y)$ .

**Передискретизация изображения** или ресамплинг (resampling) – это изменение частоты дискретизации цифрового сигнала. Применительно к цифровым изображениям это означает изменение размеров изображения.

Существуют различные алгоритмы ресамплинга изображений. Например, для увеличения изображения в 2 раза методом билинейной интерполяции (bilinear interpolation) промежуточные столбцы и строки получают линейной интерполяцией значений соседних столбцов и строк. Можно каждую точку нового изображения получить как взвешенную сумму большего числа точек исходного изображения (бикубическая и другие виды интерполяции). Наиболее качественный ресамплинг получается при использовании алгоритмов, учитывающих не только временную, но и частотную область сигнала.

Рассмотрим алгоритм ресамплинга с максимальным сохранением частотной информации изображения. Работу алгоритма будем рассматривать на одномерных сигналах, так как двумерное изображение можно сначала растянуть или сжать по горизонтали (по строкам) а потом – по вертикали (по столбцам), и свести ресамплинг двумерного изображения к ресамплингу одномерных сигналов.

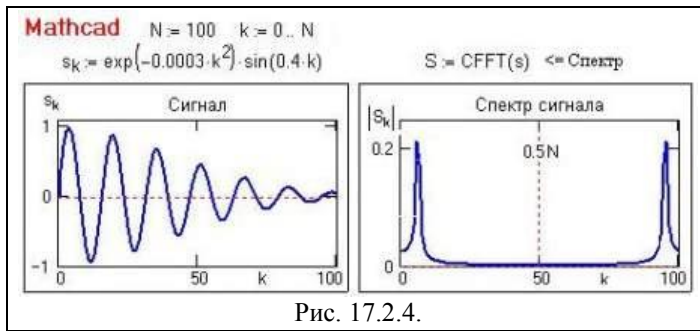


Рис. 17.2.4.

равно количеству отсчетов сигнала) и приводится в главном диапазоне БПФ ( $0-2\pi$ , частота Найквиста  $\omega_N = \pi/\Delta t = \pi$ , или  $0.5N$  по нумерации отсчетов спектра при шаге по спектру  $\Delta f = 1/T$  или  $\Delta\omega = 2\pi/T$ ). Для выполнения растяжения необходимо выполнить 2 шага.

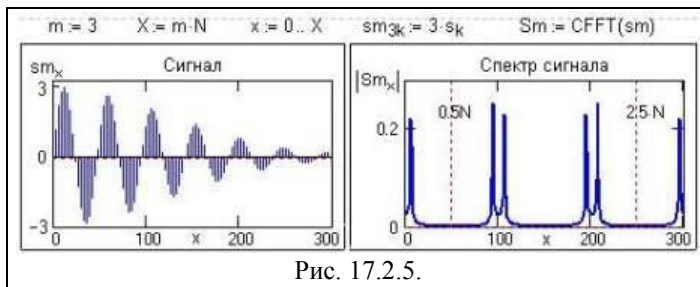


Рис. 17.2.5.

теперь располагается в  $m$ -раз больше интервалов дискретизации ( $mN$ ), и новый шаг дискретизации будет равен  $\Delta x = \Delta t/m$ . Соответственно, новая частота Найквиста для этого сигнала равна  $m\pi/\Delta t = m\pi$ . Но физическая величина шага по спектру в единицах частоты обратна физической величине интервала задания сигнала ( $\Delta f = 1/T$ ), и, следовательно, БПФ по  $mN$  точкам сигнала вычислит  $mN$  точек спектра в главном диапазоне БПФ  $0-2\pi m$  с шагом спектра исходного сигнала, в котором будут присутствовать  $m$ -периодов спектра исходного сигнала (один главный и  $m-1$  боковых).

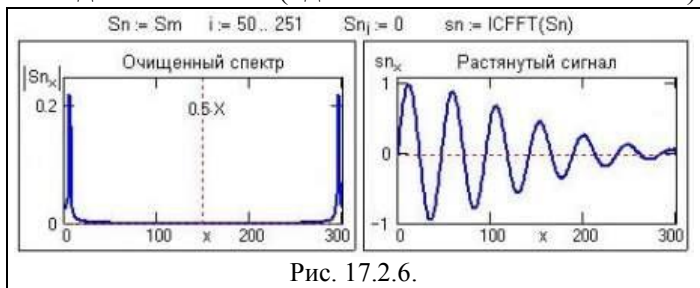


Рис. 17.2.6.

раз длиннее исходного с полным сохранением всей частотной информации.

По аналогичному принципу может быть построен и алгоритм сжатия (децимации) сигнала в  $n$  раз, при этом порядок шагов изменяется на обратный. При сжатии сигнала выполняется увеличение шага дискретизации сигнала и, соответственно, уменьшение частоты Найквиста, при этом отрезаемые высокие частоты (шумы и незначимые высокочастотные части спектра сигнала) будут отражаться от границы главного диапазона и суммироваться с основной информацией, создавая искажения. Для исключения этого явления сначала проводят низкочастотную фильтрацию сигнала с частотой среза, равной новой частоте Найквиста (антиалиасинг), и только затем децимацию сигнала путем прореживания.

При выполнении ресемплинга только во временной области алгоритмы растяжения и сжатия объединяют, как правило, в единый последовательный процесс с заданием изменения шага дискретизации в виде отношения  $m/n$ , что позволяет задавать целые значения  $m$  и  $n$  при дробных значениях изменения шага дискретизации. Это существенно упрощает алгоритмы и повышает эффективность и качество их работы. Так например, при растяжении сигнала в 1.5 раза при  $m/n = 3/2$  сигнал сначала растягивается в 3 раза

Допустим, мы имеем одномерный сигнал (рис. 17.2.4), заданный на интервале  $0-T$  и дискретизированный с шагом  $\Delta t=1$  ( $N$  интервалов). Нужно «растянуть» сигнал в  $m$  раз. Спектр сигнала, показанный на рисунке, вычисляется быстрым преобразованием Фурье (БПФ, количество отсчетов спектра

Первый шаг – интерполяция нулями, увеличивающая длину сигнала в  $m$  раз. (рис. 17.2.5). Нужно умножить все отсчеты исходного сигнала на  $m$ , а потом после каждого отсчета сигнала вставить  $m-1$  нулевое значение. На интервале  $0-T$ , значение которого остается без изменения,

Второй шаг – отфильтровывание боковых диапазонов спектра с помощью НЧ-фильтра или во временной, или в спектральной области. На рис. 17.2.6 выполнены очистка спектра и обратное преобразование Фурье, в результате чего получен сигнал в  $m$



(простое и равномерное дополнение нулями всех отсчетов, затем выполняется НЧ-фильтрация, после чего сигнал прореживается в два раза. Антиалиасингового фильтра не требуется, т.к. частота его среза перекрывается частотой первого НЧ-фильтра. При обратной операции сжатия (например,  $m/n = 2/3$ ), аналогично используется только антиалиасинговый фильтр.

### 17.3. ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ

Под фильтрацией изображений понимают операцию, имеющую своим результатом изображение того же размера, полученное из исходного по некоторым правилам. Обычно интенсивность (цвет) каждого пикселя результирующего изображения обусловлен интенсивностями (цветами) пикселей, расположенных в некоторой его окрестности в исходном изображении.

Правила фильтрации могут быть самыми разнообразными. Фильтрация изображений является одной из самых фундаментальных операций компьютерного зрения, распознавания образов и обработки изображений. С той или иной фильтрации исходных изображений начинается работа подавляющего большинства методов обработки изображений.

**Линейные фильтры** имеют очень простое математическое описание. Будем считать, что задано исходное полутоновое изображение  $A$ , и обозначим интенсивности его пикселей  $A(x, y)$ . Линейный фильтр определяется вещественнозначной функцией  $h$  (ядром фильтра), заданной на растре. Сама фильтрация производится при помощи операции дискретной свертки (взвешенного суммирования):

$$B(x, y) = h(i, j) \otimes \otimes A(x, y) = \sum_i \sum_j h(i, j) A(x-i, y-j). \quad (17.3.1)$$

Результатом служит изображение  $B$ . Обычно ядро фильтра отлично от нуля только в некоторой окрестности  $N$  точки  $(0, 0)$ . За пределами этой окрестности  $h(i, j)$  равно нулю, или очень близко к нему и им можно пренебречь. Суммирование производится по  $(i, j) \in N$ , и значение каждого пикселя  $B(x, y)$  определяется пикселями изображения  $A$ , которые лежат в окне  $N$ , центрированном в точке  $(x, y)$  (обозначение - множество  $N(x, y)$ ). Ядро фильтра, заданное на прямоугольной окрестности  $N$ , может рассматриваться как матрица  $m$  на  $n$ , где длины сторон являются нечетными числами. При задании ядра матрицей ее следует центрировать. Если пиксель  $(x, y)$  находится в окрестности краев изображения, то координаты  $A(x-i, y-j)$  для определенных  $(i, j)$  могут соответствовать несуществующим пикселям  $A$  за пределами изображения. Данную проблему можно разрешить несколькими способами.

- Не проводить фильтрацию для таких пикселей, обрезав изображение  $B$  по краям, или применив для их значений исходные значения изображения  $A$ .
- Не включать отсутствующий пиксель в суммирование, распределив его вес  $h(i, j)$  равномерно среди других пикселей окрестности  $N(x, y)$ .
- Доопределить значения пикселей за границами изображения при помощи экстраполяции.
- Доопределить значения пикселей за границами изображения, при помощи зеркального продолжения изображения.

Выбор способа производится с учетом конкретного фильтра и особенностей изображения.

**Сглаживающие фильтры.** Простейший прямоугольный сглаживающий фильтр радиуса  $r$  задается при помощи матрицы размера  $(2r+1) \times (2r+1)$ , все значения которой равны  $1/(2r+1)^2$ , а сумма значений равна единице. Это двумерный аналог низкочастотного одномерного П-образного фильтра скользящего среднего. При фильтрации с таким ядром значение пикселя заменяется усредненным значением пикселей в квадрате со стороной  $2r+1$  вокруг него. Пример маски фильтра  $3 \times 3$ :

$$M_1^{low} = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Одним из применений фильтров является шумоподавление. Шум меняется независимо от пикселя к пикселю и, при условии, что математическое ожидание значения шума равно нулю, шумы соседних пикселей при суммировании будут компенсировать друг друга. Чем больше окно фильтрации, тем меньше будет усредненная интенсивность шума, однако при этом будет происходить и соответствующее размытие значащих деталей изображения. Образом белой точки на черном фоне при фильтрации (реакция на единичный импульс) будет равномерно серый квадрат.

Шумоподавление при помощи прямоугольного фильтра имеет существенный недостаток: все пиксели в маске фильтра на любом расстоянии от обрабатываемого оказывают на результат одинаковый эффект. Несколько лучший результат получается при модификации фильтра с увеличением веса центральной точки:

$$M_2^{low} = \frac{1}{10} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Более эффективное шумоподавление можно осуществить, если влияние пикселей на результат будет уменьшаться с увеличением расстояния от обрабатываемого. Этим свойством обладает гауссовский фильтр с ядром:  $h(i, j) = (1/2\pi\sigma^2) \exp(-(i^2+j^2)/2\sigma^2)$ . Гауссовский фильтр имеет ненулевое ядро бесконечного размера. Однако значения ядра фильтра очень быстро убывает к нулю при удалении от точки (0, 0), и потому на практике можно ограничиться сверткой с окном небольшого размера вокруг (0, 0), например, взяв радиус окна равным  $3\sigma$ .

Гауссовская фильтрация также является сглаживающей. Однако, в отличие от прямоугольного фильтра, образом точки при гауссовой фильтрации будет симметричное размытое пятно, с убыванием яркости от середины к краям. Степень размытия изображений определяются параметром  $\sigma$ .

**Контрастоповышающие фильтры.** Если сглаживающие фильтры снижают локальную контрастность изображения, размывая его, то контрастоповышающие фильтры производят обратный эффект и, по существу, являются фильтрами высоких пространственных частот. Ядро контрастоповышающего фильтра в точке (0, 0) имеет значение, большее 1, при общей сумме значений, равной 1. Например, контрастоповышающими фильтрами являются фильтры с ядром, задаваемым матрицами:

$$M_1^{contr} = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \quad M_2^{contr} = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{pmatrix}.$$

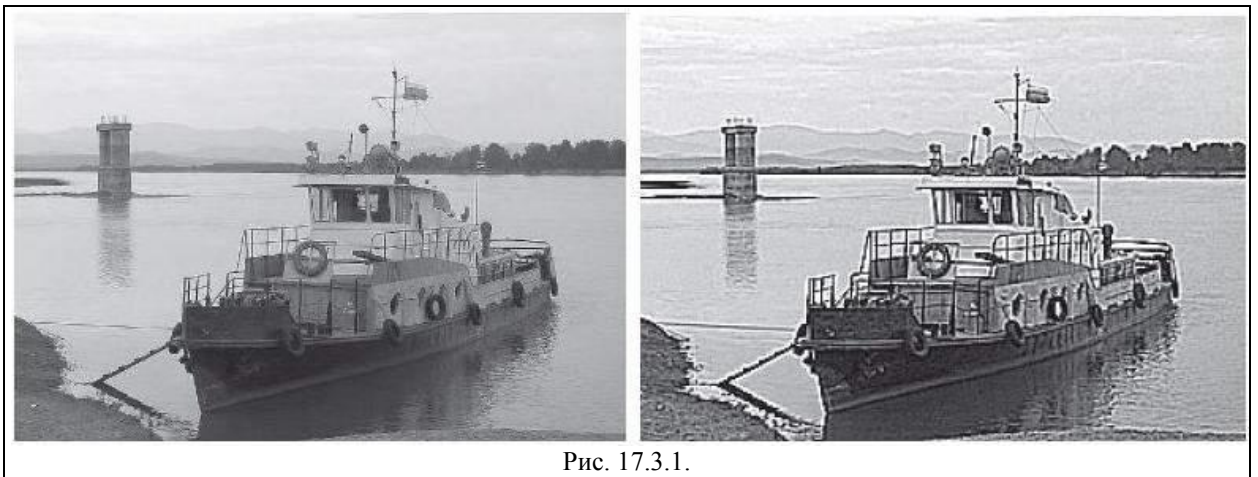


Рис. 17.3.1.

Пример применения фильтра  $M_1^{contr}$  приведен на рис. 17.3.1. Эффект повышения

контраста достигается за счет того, что фильтр подчеркивает разницу между интенсивностями соседних пикселей, удаляя эти интенсивности друг от друга. Этот эффект будет тем сильнее, чем больше значение центрального члена ядра. Характерным артефактом линейной контрастоповышающей фильтрации являются заметные светлые и менее заметные темные ореолы вокруг границ.

**Разностные фильтры** – это линейные фильтры, задаваемые дискретными аппроксимациями дифференциальных операторов (по методу конечных разностей). Данные фильтры играют важнейшую роль во многих приложениях, например, для задач поиска границ на изображении.

Простейшим дифференциальным оператором является взятие производной по x-координате  $d/dx$ , который определен для непрерывных функций. Распространенными вариантами аналогичных операторов для дискретных изображений являются фильтры Прюита (Prewitt) и Собеля (Sobel):

$$M_1^{prewitt} = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}, \quad M_1^{sobel} = \frac{1}{4} \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}.$$

Фильтры, приближающие оператор производной по y-координате  $d/dy$ , получаются путем транспонирования матриц.

Простейший алгоритм вычисления нормы градиента по трем смежным точкам:

$$G(x, y) = \sqrt{(A_{x+1,y} - A_{x,y})^2 + (A_{x,y+1} - A_{x,y})^2}.$$

Применяется также упрощенная формула вычислений:

$$G(x, y) \cong |A_{x+1,y} - A_{x,y}| + |A_{x,y+1} - A_{x,y}|.$$

Вычисление нормы градиента по четырем смежным точкам (оператор Робертса):

$$G(x, y) = \sqrt{(A_{x+1,y+1} - A_{x,y})^2 + (A_{x+1,y} - A_{x,y+1})^2}.$$

$$G(x, y) \cong |A_{x+1,y+1} - A_{x,y}| + |A_{x+1,y} - A_{x,y+1}|.$$

В алгоритме Собеля используется восемь отсчетов яркости в окрестностях центральной точки:

$$G(x, y) = \sqrt{G_{x,y}^2 + G_{y,x}^2}, \quad G(x, y) \cong |G_{x,y} + G_{y,x}|,$$

$$G_{x,y} = [A_{x-1,y-1} + 2A_{x-1,y} + A_{x-1,y+1}] - [A_{x+1,y-1} + 2A_{x+1,y} + A_{x+1,y+1}],$$

$$G_{y,x} = [A_{x-1,y-1} + 2A_{x,y-1} + A_{x+1,y-1}] - [A_{x-1,y+1} + 2A_{x,y+1} + A_{x+1,y+1}].$$

Наряду с более точным определением нормы градиента алгоритм Собеля позволяет определять и направление вектора градиента в плоскости анализа изображения в виде угла  $\varphi$  между вектором градиента и направлением строк матрицы:

$$\varphi(x, y) = \text{arctg}(G_{y,x}/G_{x,y}).$$

В отличие от сглаживающих и контрастоповышающих фильтров, не меняющих среднюю интенсивность изображения, в результате применения разностных операторов получается, как правило, изображение со средним значением пикселя близким к нулю. Вертикальным перепадам (границам) исходного изображения соответствуют пиксели с большими по модулю значениями на результирующем изображении. Поэтому разностные фильтры называют также фильтрами выделения границы объектов.

Аналогично вышеприведенным фильтрам, по методу конечных разностей можно составить фильтры для других дифференциальных операторов. В частности, важный для многих приложений дифференциальный оператор Лапласа (лапласиан)  $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$  можно приблизить для дискретных изображений фильтром с матрицей (один из

вариантов):

$$M_{\Delta} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

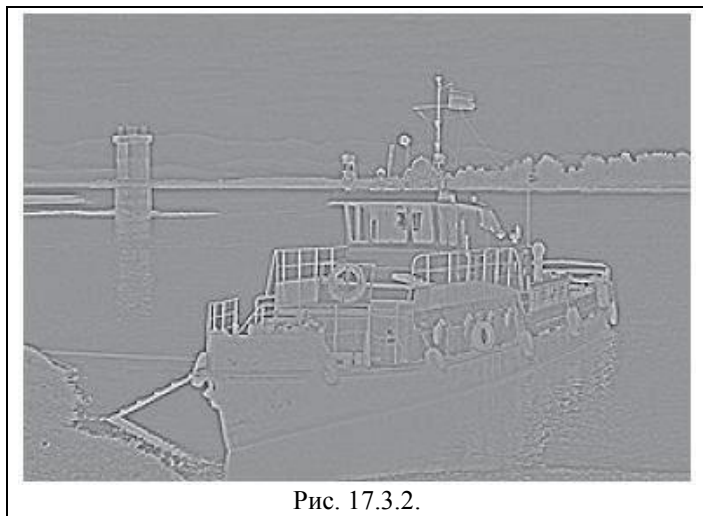


Рис. 17.3.2.

Как видно на рис. 17.3.2, в результате применения дискретного лапласиана большие по модулю значения соответствуют как вертикальным, так и горизонтальным перепадам яркости. Фильтр является, таким образом, фильтром, находящим границы любой ориентации. Нахождение границ на изображении может производиться путем применения этого фильтра и взятия всех пикселей, модуль значения которых превосходит некоторый порог.

Однако такой алгоритм имеет существенные недостатки. Главный из них

- неопределенность в выборе величины порога. Для разных частей изображения приемлемый результат обычно получается при существенно разных пороговых значениях. Кроме того, разностные фильтры очень чувствительны к шумам изображения.

**Двумерная циклическая свертка.** Как и для одномерных сигналов, двумерная свертка может выполняться в области пространственных частот с использованием алгоритмов быстрого преобразования Фурье и перемножением двумерных спектров изображения и ядра фильтра. Она также является циклической, и выполняется обычно в скользящем варианте. С учетом цикличности, для вычисления постоянного шаблона спектра ядра размеры маски фильтра ядра удваиваются по осям и дополняются нулями, и эти же размеры маски используются для выделения скользящего по изображению окна, в пределах которого и выполняется БПФ. Реализация КИХ фильтра с помощью БПФ особенно эффективна, если фильтр имеет большую опорную область.

**Нелинейные фильтры.** В цифровой обработке изображений широко применяются нелинейные алгоритмы на основе ранговой статистики для восстановления изображений, поврежденных различными моделями шумов. Они позволяют избежать дополнительного искажения изображения при удалении шума, а также значительно улучшить результаты работы фильтров на изображениях с высокой степенью зашумленности.

Введем понятие  $M$ -окрестности элемента изображения  $A(x, y)$ , который является для этой окрестности центральным. В простейшем случае  $M$ -окрестность содержит  $N$ -пикселей – точки, попадающие в маску фильтра, включая (или не включая) центральный. Значения этих  $N$ -элементов можно расположить в вариационном ряду  $V(r)$ , ранжированном по возрастанию (или убыванию), и вычислить определенные моменты этого ряда, например, среднее значение яркости  $m_N$  и дисперсии  $d_N$ . Вычисление выходного значения фильтра, которым заменяется центральный отсчет, выполняется по формуле:

$$B(x, y) = \alpha A(x, y) + (1-\alpha)m_N. \quad (17.3.2)$$

Значение коэффициента  $\alpha = [0, 1]$  связывается определенной зависимостью со статистикой отсчетов в окне фильтра, например:

$$\alpha = d_N / (d_N + k d_S), \quad (17.3.3)$$

где  $d_S$  – дисперсия шумов по изображению в целом или по  $S$ -окрестности при  $S > M$  и  $M \in S$ ,  $k$  – константа доверия дисперсии  $S$ -окрестностей. Как следует из этой формулы, при  $k=1$  и  $d_N \approx d_S$  имеет место  $\alpha \approx 0.5$ , а значение  $B(x, y) = (A(x, y) + m_N)/2$ , т.е. складываются в равной степени от значений центрального отсчета и среднего значения пикселей его  $M$ -окрестности. При увеличении

значений  $d_N$  происходит увеличение вклада в результат значения центрального отсчета, при уменьшении – значения  $m_N$ . Весомость вклада средних значений по  $M$ -окрестности можно изменять значением коэффициента  $k$ .

Выбор статистической функции и характер зависимости от нее коэффициента  $\alpha$  может быть достаточно многообразным (например, по дисперсиям разностей отсчетов в  $M$ -окрестности с центральным отсчетом), и зависит как от размеров апертуры фильтра, так и от характера изображений и шумов. По существу, значение коэффициента  $\alpha$  должно задавать степень поврежденности центрального отсчета и, соответственно, функцию заимствования для его исправления отсчетов из  $M$ -окрестности.

Наиболее простыми и распространенными типами нелинейных фильтров для обработки изображений являются пороговые и медианные фильтры.

**Пороговая фильтрация** задается, например, следующим образом:

$$B(x, y) = \begin{cases} A(x, y), & A(x, y) - m_N \leq p \\ m_N, & A(x, y) - m_N > p \end{cases}$$

Величина  $p$  является порогом фильтрации. Если величина центральной точки фильтра превышает среднее значение отсчетов  $m_N$  в ее  $M$ -окрестности на величину порога, то она заменяется средним значением. Значение порога может быть как константой, так и функционально зависимым от величины центральной точки.

**Медианная фильтрация** определяется следующим образом:

$$B(x, y) = \text{med} \{M(x, y)\},$$

т.е. результат фильтрации есть медианное значение пикселей окрестности, форма которой определяется маской фильтра. Медианная фильтрация способна эффективно удалять из изображения помехи, независимо воздействующие на отдельные пиксели. Например, такими помехами являются "битые" пиксели при цифровой съемке, "снеговой" шум, когда часть пикселей заменяется на пиксели с максимальной интенсивностью, и т.п. Преимущество медианной фильтрации заключается в том, что "горячий" пиксель на темном фоне будет заменен темным, а не "размазан" по окрестности.

Медианная фильтрация обладает выраженной избирательностью по отношению к элементам массива, представляющим собой немонотонную составляющую последовательности чисел в пределах апертуры фильтра. В то же время монотонную составляющую последовательности медианный фильтр оставляет без изменений. Благодаря этой особенности, медианные фильтры при оптимально выбранной апертуре сохраняют без искажений резкие границы объектов, подавляя некоррелированные или слабо коррелированные помехи и малоразмерные детали.

**Фильтры экстремумов** определяются по правилам:

$$B_{\min}(x, y) = \min \{M(x, y)\},$$

$$B_{\max}(x, y) = \max \{M(x, y)\},$$

т.е. результат фильтрации есть минимальное и максимальное значения пикселей в маске фильтра. Применяются такие фильтры, как правило, для бинарных изображений.

#### 17.4. СЖАТИЕ ИЗОБРАЖЕНИЙ

Типичное изображение с разрешением порядка  $3000 \times 2000$  при 24 бит на пиксель для передачи цвета имеет объем 17 мегабайт. Для профессиональных устройств размер получаемого раstra изображений может быть значительно больше, глубина цвета до 48 бит на пиксель, а размер одного изображения может быть больше 200 мегабайт. Поэтому весьма актуальными являются алгоритмы сжатия изображений для уменьшения объема данных, представляющих изображение.

Существуют два основных класса алгоритмов:

1. Сжатие без потерь  $A$  (lossless compression), если существует такой обратный алгоритм  $A^{-1}$ , что для любого  $h$  - изображения  $A[h] = h_1$  имеем  $A^{-1}[h_1] = h$ . Сжатие без потерь применяется в таких графических форматах представления изображений, как: GIF, PCX, PNG, TGA, TIFF, и применяется при обработке особо ценной первичной



информации (медицинские изображения, аэро- и космоснимки и т.п.), когда даже малейшие искажения нежелательны

2. Сжатие с потерями (lossy compression), если оно не обеспечивает возможность точного восстановления исходного изображения. Парный к А алгоритм примерного восстановления изображения будем обозначать как А\*. Пара (А, А\*) подбирается так, чтобы обеспечить большие коэффициенты сжатия при сохранении визуального качества. Сжатие с потерями применяется в графических форматах: JPEG, JPEG2000 и т.д.

Все алгоритмы и утверждения относятся как к изображениям, так и к произвольным последовательностям, элементы которых могут принимать конечное количество значений. При этом следует учитывать, что не существует идеальных алгоритмов, сжимающих без потерь любой набор данных.

**Алгоритмы кодирования длины повторения (RLE)** базируются на простом принципе: замене повторяющихся групп элементов исходной последовательности на пару (количество, элемент), либо только на количество.

Битовый уровень. Будем рассматривать исходные данные на уровне последовательности битов, например, представляющих черно-белое изображение. Поряд обычно идет несколько 0 или 1, и кодировать можно количество идущих подряд одинаковых цифр. Но количество повторений также надо кодировать битами. Можно считать, что каждое число повторений изменяется от 0 до 7 (3-х битовый код), чередуя последовательность кодов единиц и нулей. Например, последовательности 1111111111 можно сопоставить числа 7 0 4, т.е. 7 единиц, 0 нулей, 4 единицы, при этом имеем новый год – 111 000 100. Чем больше длина последовательностей одинаковых бит, тем больше эффект. Так, последовательность из 21 единицы, 21 нуля, 3 единицы и 7 нулей закодируется так: 111 000 111 000 111 111 000 111 000 111 011 111, т.е. из исходной последовательности длиной 51 бит, имеем последовательность длиной 36 бит.

Байтовый уровень. Предположим, что на вход подается полутоновое изображение, где на значение интенсивности пикселя отводится 1 байт, при этом ожидание длинной цепочки одинаковых битов существенно снижается.

Будем разбивать входной поток на байты (код от 0 до 255) и кодировать повторяющиеся байты парой (количество, буква). Одиночный байт можно не изменять. Так, байты AABBBCCDAA кодируем (2A) (3B) (C) (D) (2A).

Однако модификации данного алгоритма редко используются сами по себе (например, в формате PCX), т.к. подкласс последовательностей, на которых алгоритм эффективен, относительно узок. Чаще они используются как одна из стадий конвейера сжатия.

**Словарные алгоритмы** вместо кодирования только по одному элементу входящей последовательности производят кодирование цепочки элементов. При этом используется словарь цепочек (созданный по входной последовательности) для кодирования новых.

Алгоритм LZ77 был одним из первых, использующих словарь. В качестве словаря используются N последних уже закодированных элементов последовательности. В процессе сжатия словарь-подпоследовательность "скользит" по входящей последовательности. Цепочка элементов на выходе кодируется следующим образом: положение совпадающей части обрабатываемой цепочки элементов в словаре - смещение (относительно текущей позиции), длина, первый элемент, следующий за совпавшей частью цепочки. Длина цепочки совпадения ограничивается сверху числом n. Соответственно, задача состоит в нахождении наибольшей цепочки из словаря, совпадающей с обрабатываемой последовательностью. Если же совпадений нет, то записывается нулевое смещение, единичная длина и первый элемент незакодированной последовательности.

Описанная выше схема кодирования приводит к понятию скользящего окна (sliding window), состоящего из двух частей:

- подпоследовательность уже закодированных элементов длины N-словарь - буфер

поиска (search buffer);

- подпоследовательность длины  $n$  из цепочки элементов, для которой будет произведена попытка поиска совпадения - буфер предварительного просмотра (look-ahead buffer).

Декодирование сжатой последовательности является расшифровкой записанных кодов: каждой записи сопоставляется цепочка из словаря и явно записанного элемента, после чего производится сдвиг словаря. Словарь воссоздается по мере работы алгоритма декодирования.

Данный алгоритм является родоначальником целого семейства алгоритмов. К его достоинствам можно отнести приличную степень сжатия на достаточно больших последовательностях и быструю распаковку. К недостаткам относят медленную скорость сжатия и меньшую, чем у альтернативных алгоритмов, степень сжатия.

Алгоритм LZW. Словарь в данном алгоритме представляет собой таблицу, которая заполняется цепочками элементов по мере работы алгоритма. В процессе сжатия отыскивается наиболее длинная цепочка, уже записанная в словарь. Каждый раз, когда новая цепочка элементов не найдена в словаре, она добавляется в словарь, при этом записывается код цепочки. В теории на размер таблицы не накладывается ограничений, однако ограничение на размер позволяет улучшить степень сжатия, т.к. накапливаются ненужные (не встречающиеся) цепочки. Чем больше вхождений имеет таблица, тем больше информации нужно выделять для хранения кодов.

Декодирование заключается в прямой расшифровке кодов, т.е. в построении словаря и вывода соответствующих цепочек. Словарь инициализируется так же, как и в кодере. К достоинствам алгоритма можно отнести высокую степень сжатия и достаточно высокую скорость, как сжатия, так и декодирования.

**Алгоритмы статистического кодирования** ставят в соответствие каждому элементу последовательности код так, чтобы его длина соответствовала вероятности появления элемента. Сжатие происходит за счет замены элементов исходной последовательности, имеющих одинаковые длины (каждый элемент занимает одинаковое количество бит), на элементы разной длины, пропорциональной отрицательному логарифму от вероятности, т.е. элементы, встречающиеся чаще, чем остальные, имеют код меньшей длины.

Алгоритм Хаффмена использует префиксный код переменной длины, обладающий особым свойством: менее короткие коды не совпадают с префиксом (начальной частью) более длинных. Такой код позволяет осуществлять взаимно-однозначное кодирование. Процесс сжатия заключается в замене каждого элемента входной последовательности его кодом. Построение набора кодов обычно осуществляется с помощью так называемых *кодовых деревьев*.

Алгоритм Хаффмена является двухпроходным. Первый проход по изображению создает таблицу весов элементов, а во время второго происходит кодирование. Существуют реализации алгоритма с фиксированной таблицей. Часто бывает, что априорное распределение вероятностей элементов алфавита неизвестно, т.к. не доступна вся последовательность сразу, при этом используются адаптивные модификации алгоритма Хаффмена.

**Сжатие изображений с потерями.** Объем информации, нужной для хранения изображений, обычно велик. Классические алгоритмы, являясь алгоритмами общего назначения, не учитывают, что сжимаемая информация есть изображение - двумерный объект, и не обеспечивают достаточной степени сжатия.

Сжатие с потерями основывается на особенностях восприятия человеком изображения: наибольшей чувствительности в определенном диапазоне волн цвета, способности воспринимать изображение как единое целое, не замечая мелких искажений. Главный класс изображений, на который ориентированы алгоритмы сжатия с потерями - фотографии, изображения с плавными цветовыми переходами.

**Оценка потерь в изображениях.** Существует множество мер для оценки потерь в изображениях после их восстановления (декодирования) из сжатых, однако для всех из них можно подобрать такие два изображения, что их мера отличия будет достаточно большой, но на глаз различия будут почти незаметными. И наоборот - можно подобрать изображения, сильно различающиеся на глаз, но имеющие небольшую меру отличия.

Стандартной числовой мерой потерь обычно является среднеквадратическое отклонение (СКО) значений пикселей восстановленного изображения от исходного. Тем не менее, самой важной "мерой" оценки потерь является мнение наблюдателя. Чем меньше различий (а лучше - их отсутствие) обнаруживает наблюдатель, тем выше качество алгоритма сжатия. Алгоритмы сжатия с потерями часто предоставляют пользователю возможность выбирать объем "теряемых" данных, т.е. право выбора между качеством и размером сжатого изображения. Естественно, что чем лучше визуальное качество при большем коэффициенте сжатия, тем алгоритм лучше.

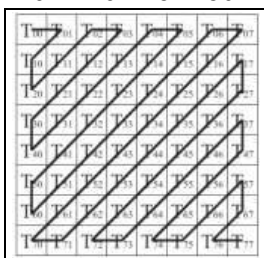
**Преобразование Фурье.** В общем случае изображение можно рассматривать как функцию двух переменных, определенную в точках конечного растра. Множество таких функций на точках фиксированного конечного растра образуют конечномерное евклидово пространство, и к ним может быть применено дискретное преобразование Фурье, т.е. спектральное представление изображения. Оно обеспечивает:

- Некоррелированность и независимость коэффициентов спектра, т.е. точность представления одного коэффициента не зависит от любого другого.
- "Уплотнение" энергии (energy compaction). Преобразование сохраняет основную информацию в малом количестве коэффициентов. Данное свойство сильнее всего проявляется на фотореалистичных изображениях.

Коэффициенты спектрального представления - это амплитуды пространственных частот изображения. В случае изображений с плавными переходами большая часть информации содержится в низкочастотном спектре.

Алгоритм сжатия, используемый в формате JPEG, построен на использовании дискретного косинусного преобразования Фурье. Схема сжатия в алгоритме представляет собой конвейер, где это преобразование - лишь одна из стадий, но одна из основных. Алгоритм содержит следующие основные операции:

1. Перевод в цветовое пространство YCbCr. Здесь Y - компонента яркости, Cb и Cr - компоненты цветности. Человеческий глаз более чувствителен к яркости, чем к цвету. Поэтому важнее сохранить большую точность при передаче Y, чем при передаче Cb и Cr.
2. Дискретное косинус-преобразование (ДКП). Изображение разбивается на блоки  $8 \times 8$ . К каждому блоку применяется дискретное косинус-преобразование (отдельно для компонент Y, Cb и Cr).
3. Сокращение высокочастотных составляющих в матрицах ДКП. Человеческий глаз практически не замечает изменения в высокочастотных составляющих, следовательно, коэффициенты, отвечающие за высокие частоты, можно хранить с меньшей точностью.



4. Зигзаг-упорядочивание матриц. Это особый проход матрицы для получения одномерной последовательности. Сначала идет элемент T00, затем T01, T10, T11 . . . Причем для типичных фотореалистических изображений сначала будут идти ненулевые коэффициенты, соответствующие низкочастотным компонентам, а затем - множество нулей (высокочастотные составляющие).

5. Сжатие сначала методом RLE, а затем методом Хаффмана.

Алгоритм восстановления изображения действует в обратном порядке. Степень сжатия от 5 до 100 и более раз. При этом визуальное качество для большинства фотореалистичных изображений остается на хорошем уровне при сжатии до 15 раз. Алгоритм и формат являются самыми распространенными для передачи и хранения полноцветных изображений.



**Вейвлет-преобразование** сигналов является обобщением классического преобразования Фурье. Термин "вейвлет" (wavelet) в переводе с английского означает "маленькая (короткая) волна". Вейвлеты - это обобщенное название семейств математических функций определенной формы, которые локальны во времени и по частоте, и в которых все функции получаются из одной базовой посредством ее сдвигов и растяжений по оси времени.

В алгоритмах сжатия с потерями, как правило, сохраняются все операции конвейера сжатия с заменой дискретного преобразования Фурье на дискретное вейвлет-преобразование. Вейвлет-преобразования имеют очень хорошую частотно-пространственную локализацию и по этому показателю превосходят традиционные преобразования Фурье. При этом становится возможно применять более сильное квантование, улучшая свойства последовательности для последующего сжатия. Алгоритмы сжатия изображений, основанные на этом преобразовании, при той же степени сжатия показывают лучшие результаты по сохранению качества изображения.

## Тема 18. РАСПОЗНАВАНИЕ ОБЪЕКТОВ ИЗОБРАЖЕНИЙ

Подробное рассмотрение теоретических аспектов данной темы не является задачей настоящего раздела. Более полную информацию по этим вопросам желающие могут получить в других источниках. Ниже даются практические рекомендации по реализации отдельных методов, непосредственно связанных с наиболее типичными задачами, которые решаются на основе использования алгоритмов, рассмотренных в предыдущей теме.

Распознавание объектов на изображениях оптических датчиков (фото-, аэро-, спутниковых и прочих снимках) – традиционная область обработки изображений. Вместе с тем методы распознавания объектов начали широко применяться в автоматических системах теленаблюдений, в рентгеновской и магнито-резонансной томографии, и прочих видах изначально цифровых изображений как для постобработки, так и для обработки в реальном времени.

### 18.1. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ИЗОБРАЖЕНИЙ

Операции распознавания на изображениях определенных объектов, как правило, предваряются обработкой изображений для создания условий, повышающих эффективность и качество выделения и распознавания искомым или изучаемых объектов. Методы предварительной обработки зависят от задач исследований, довольно разнообразны и могут включать, например, выделение наиболее информативных фрагментов, их увеличение, получение 3-мерных изображений, цветокартирование, реализация высокого пространственного разрешения, повышение контрастного разрешения, улучшение качества изображений и т.п. Рассмотрим среди них те, без которых, обычно, не обходится ни одна типовая задача.

#### **Коррекция яркости и контрастности изображений.**

Изображения, вводимые в компьютер, часто являются малоконтрастными. Слабый контраст, как правило, обусловлен широким диапазоном воспроизводимых яркостей, нередко сочетающийся с нелинейностью характеристики передачи уровней. Характер зависимости изменения яркости палитры пикселей от минимального значения до максимального также влияет на качество изображения. Оптимальной является линейная функция изменения интенсивности пикселей. При вогнутой характеристике изображение будет более темным, при выпуклой — более светлым. И в том, и в другом случае признаки объектов могут быть искажены и недостаточно хорошо идентифицируемы. Коррекция (линеаризация) яркости палитры существенно улучшает качество изображения.

Малая контрастность может быть обусловлена и тем, что вариации функции яркости пикселей на изображении намного меньше допустимого диапазона шкалы

яркостей. В этом случае контрастность изображения повышается путем "растягивания" реального динамического диапазона яркостей на всю шкалу при помощи линейного поэлементного преобразования.

Другой способ коррекции яркости палитры связан с инверсией входного изображения. Поскольку различать слабые сигналы на темном фоне достаточно сложно, то инверсная форма представления таких изображений имеет другую гистограмму яркостей, более приемлемую для наблюдения и визуальной идентификации.

Некоторые задачи обработки изображения связаны с преобразованием полутонового изображения (много градаций яркости) в бинарное (две градации). Преобразование осуществляется для того, чтобы сократить информационную избыточность изображения, оставить в нем только информацию, которая нужна для решения конкретной задачи. В бинарном изображении должны быть сохранены определенные детали (например, очертания изображенных объектов) и исключены несущественные особенности (фон).

Пороговая обработка полутонового изображения заключается в разделении всех элементов изображения на два класса  $A_1$  и  $A_2$  по признаку яркости с границей  $A_{гр}$ , и в выполнении соответствующей пороговой фильтрации с заменой пикселей изображения на установленную яркость классов. Выбор границы определяется видом гистограммы яркости исходного изображения. Для простейших изображений типа чертежей, машинописного текста и т.п., имеющих бимодальное распределение, граница устанавливается по минимуму между модами распределения. В общем случае изображение может быть многомодальным, и если устанавливается достаточно надежное соответствие между объектами и соответствующими модами их яркости, то пороговая фильтрация также может предусматривать несколько классов яркости пикселей.

Диапазон яркости изображения в компьютере может иметь отличия от диапазона яркостей исходного, например, в силу недостаточной экспозиции. Существует два возможных способа коррекции яркости. Согласно первому способу изображение линейно отображается в диапазоне яркостей исходного. Второй способ предусматривает ограничение яркости пикселей в обработанном изображении максимальным и минимальным пороговыми уровнями, и имеет более широкое применение. Присутствие в изображении самых светлых и самых темных тонов создает впечатление хорошей контрастности, однако излишняя контрастность приводит к тому, что максимальные градации влияют на средние тона, а большинство деталей изображения окрашены именно в средних тонах и излишняя контрастность может привести к потере этих деталей или затруднить их выделение.

**Гистограммы яркости.** Инструментом для оценки уровней интенсивности пикселей является гистограмма - графическое отображение количественной характеристики вероятностного распределения интенсивности (яркости) пикселей в выделенном участке изображения. Максимальному значению интенсивности пикселей присваивается уровень градации интенсивности 255 (белый цвет), самому темному - значение 0 (черный цвет). Интенсивности в диапазоне от 0 до 255 имеют линейную шкалу изменения, либо устанавливаемую в соответствии с принятой функцией изменения, например, усиливающей слабые сигналы (градации серого) и ослабляющей сильные сигналы (в области белого цвета), чем повышается пространственное и контрастное разрешение изображения или определенной зоны интереса.

Известен метод улучшения изображений, основанный на вычислении логарифма спектральных коэффициентов преобразования Фурье исходного изображения (вычисление кепстра). При обратном преобразовании кепстра в изображение происходит выравнивание гистограммы изображения за счет логарифмического преобразования спектра изображения.

Многие изображения характеризуются гистограммами с высокой концентрацией линий в определенных зонах распределения интенсивности. Часто гистограмма

распределения яркостей изображения имеет перекося в сторону малых уровней (яркость большинства элементов ниже средней). Одним из методов улучшения качества таких изображений является видоизменение их гистограммы. Выравнивание гистограммы может быть осуществлено на основе возведения в степень модуля спектральных коэффициентов Фурье-преобразования изображения, при этом знак и фаза коэффициентов сохраняется. Если обозначить показатель степени  $\alpha$ , то при  $\alpha < 1$  операция извлечения корня степени  $\alpha$  уменьшает большие спектральные коэффициенты и увеличивает малые. Такое перераспределение энергии в частотной плоскости изображения приводит к более эффективному использованию динамического диапазона интенсивностей пикселей изображения в пространственной области.

Выбор хорошей маски регулирования гистограммы интенсивности пикселей повышает контраст, тем самым улучшая контрастную разрешающую способность деталей. В программах обработки есть команды, позволяющие устанавливать цвета при цветном картировании изображений, имеющие плавные или, наоборот, резкие переходы отображаемых деталей в зоне интереса. В сочетании с обращением контраста, преобразующем негативное изображение в позитивное, данный способ позволяет также повысить контраст мелких и средних деталей изображения.

Существует достаточно большой арсенал математических моделей и алгоритмов, программная реализация которых позволяет значительно повысить контрастное разрешение изображений. Эти алгоритмы основаны на процессах линейной и нелинейной фильтрации изображений, преобразующей гистограмму интенсивности.

**Выравнивание освещенности изображений.** Часто некоторые участки на изображении бывают слишком темными, чтобы на них можно было что-то разглядеть. Если прибавить яркости ко всему изображению, то изначально светлые участки могут оказаться засвеченными. Чтобы улучшить вид изображения в таких случаях, применяется метод выравнивания освещенности.

Освещенность меняется в пространстве достаточно медленно и ее можно считать низкочастотным сигналом. Само же изображение можно считать в среднем более высокочастотным сигналом. Если бы в процессе фотографии эти сигналы складывались, то их можно было бы разделять с помощью обычных фильтров. Однако на реальной фотографии получается произведение той картины, которую мы хотим видеть, и карты освещенности. И поскольку эти сигналы не складываются, а перемножаются, то избавиться от неравномерностей освещенности простой фильтрацией не удастся.

Для решения таких задач применяется гомоморфная обработка. Идея обработки заключается в сведении нелинейной задачи к линейной. Например, можно свести задачу разделения перемноженных сигналов к задаче разделения сложенных сигналов. Для этого нужно взять логарифм от произведения изображений, который будет равен сумме логарифмов сомножителей. При этом задача разделения произведения сигналов сводится к задаче разделения суммы НЧ- и ВЧ- сигналов и решается с помощью ВЧ-фильтра, который удалит из суммы сигналов низкие частоты. Останется взять от полученного сигнала экспоненту, чтобы вернуться к исходному масштабу амплитуд.

ВЧ-фильтр можно реализовать следующим образом. Сначала к изображению применяется операция размытия (НЧ-фильтр), а потом из исходного изображения вычитается размытое. Наилучший радиус размытия зависит от конкретного изображения. Можно начать эксперименты с радиуса порядка десяти пикселей.

Обычно для размытия изображения применяется двумерный гауссовский фильтр, имеющий вид  $h(x, y) = A \exp(-(x^2+y^2) / 2\sigma^2)$ . Здесь  $A$  – нормирующая константа (сумма всех коэффициентов фильтра должна быть равна 1),  $\sigma$  – «ширина» фильтра, регулирующая степень размытия.

Непосредственное вычисление двумерной свертки с таким ядром требует больших вычислений даже при сравнительно небольшом размере ядра. Однако эквивалентного

эффекта можно достичь, отфильтровав одномерным гауссианом сначала строки изображения, а затем столбцы полученного изображения. Полученный от выравнивания освещенности эффект может оказаться слишком сильным (темные области станут по яркости такими же, как и светлые). Чтобы уменьшить эффект, можно просто смешать обработанное изображение с исходным в определенной пропорции.

**Улучшение пространственного разрешения.** Интерпретация изображений тесно связана с качеством представления мелких неискаженных деталей. При этом необходимо, чтобы с увеличением фрагментов не происходило ухудшения разрешающей способности изображения при выполнении математических операций 2D-интерполяции функции пространственного распределения интенсивности пикселей в строках и столбцах матрицы изображения. Важным фактором при идентификации объектов является также локализация и отображение зон одной и той же яркости или цветности, даже если эти области имеют размеры нескольких пикселей.

Четкость изображений в профессиональных программах корректируется, как правило, определением оптимальных значений яркости и контрастности путем выбора соответствующих опций:

- а) «величины» — степени влияния воздействия на резкость изображения;
- б) «радиуса» — толщины контура резкости;
- в) «порога дискриминации» — определения контуров объектов путем задания разности значений интенсивности соседних пикселей, достаточной для того, чтобы программа повысила контрастность между ними.

Некоторые программы содержат автоматическую установку оптимального соотношения контрастности и яркости, поддерживающих необходимую четкость изображения.

При обработке изображений важную роль играют алгоритмы ранговой фильтрации, позволяющие устранять «размытость» деталей (улучшить их фокусировку) за счет выбора двумерной  $n \times n$  маски пикселей, выполнения операции ранжирования значений интенсивности пикселей в пределах задаваемой маски и присвоения центральному пикселю значения, равного максимальному значению по ранжиру. К виду ранговой относится также процедура медианной фильтрации, устраняющая в изображении некоррелированные случайные сигналы и импульсные помехи без "размытия" резких перепадов яркости на границах объектов.

Информативное применение линейной фильтрации обусловлено ее возможностями коррекции искажений различного типа, возникающих вследствие несовершенства устройств, формирующих изображение. Посредством линейной фильтрации удается уменьшить влияние флуктуационных шумов и других дефектов на воспроизводимых изображениях в режиме повышения контраста малоконтрастных деталей при увеличении масштаба зон интереса.

Коррекция апертурных искажений изображений в случае отсутствия шумовой составляющей сигнала осуществляется путем инверсной фильтрации (деконволюции). Однако следует учитывать, фильтры деконволюции имеют коэффициент усиления дисперсии шумов, больший 1, и вместо улучшения изображения может увеличиться зашумленность.

Перспективными считаются нелинейные методы фильтрации на основе частотных масок, позволяющих уменьшить влияние низкочастотных компонент сигнала изображения и усилить влияние высокочастотных компонент, что повышает пространственное разрешение деталей, описываемых в спектре изображения более высокими пространственными частотами.

## **18.2. ОПРЕДЕЛЕНИЕ ГРАНИЦ ОБЪЕКТОВ НА ИЗОБРАЖЕНИИ**

С точки зрения распознавания и анализа объектов на изображении наиболее информативными являются не значения яркостей объектов, а характеристики их границ –

контуров. Другими словами, основная информация заключена не в яркости отдельных областей, а в их очертаниях. Задача выделения контуров состоит в построении изображения именно границ объектов и очертаний однородных областей.

Как правило, граница предмета на фотографии отображается перепадом яркости между двумя сравнительно однотонными областями. Но перепад яркости может быть вызван также текстурой предмета, тенями, бликами, перепадами освещенности, и т.п.

Будем называть контуром изображения совокупность его пикселей, в окрестности которых наблюдается скачкообразное изменение функции яркости. Так как при цифровой обработке изображение представлено как функция целочисленных аргументов, то контуры представляются линиями шириной, как минимум, в один пиксел. Если исходное изображение, кроме областей постоянной яркости, содержит участки с плавно меняющейся яркостью, то непрерывность контурных линий не гарантируется. С другой стороны, если на “кусочно-постоянном” изображении присутствует шум, то могут быть обнаружены “лишние” контуры в точках, которые не являются границами областей.

При разработке алгоритмов выделения контуров нужно учитывать указанные особенности поведения контурных линий. Специальная дополнительная обработка выделенных контуров позволяет устранять разрывы и подавлять ложные контурные линии.

Процедура построения бинарного изображения границ объектов обычно складывается из двух последовательных операций: выделения контуров и их пороговой обработки.

Исходное изображение подвергается линейной или нелинейной обработке, с реакцией на перепады яркости. В результате этой операции формируется изображение, функция яркости которого существенно отличается от нуля только в областях резких изменений яркости изображения. Пороговой обработкой из этого изображения формируется контурный объект. Выбор порога на втором этапе должен производиться из следующих соображений. При слишком высоком пороге могут появиться разрывы контуров, а слабые перепады яркости не будут обнаружены. При слишком низком пороге из-за шумов и неоднородности областей могут появиться ложные контуры.

**Поиск границ на основе градиента.** Одним из наиболее простых способов выделения границ является пространственное дифференцирование функции яркости. Для двумерной функции яркости  $A(x, y)$  перепады в направлениях  $x$  и  $y$  регистрируются частными производными  $\partial A(x, y)/\partial x$  и  $\partial A(x, y)/\partial y$ , которые пропорциональны скоростям изменения яркости в соответствующих направлениях.

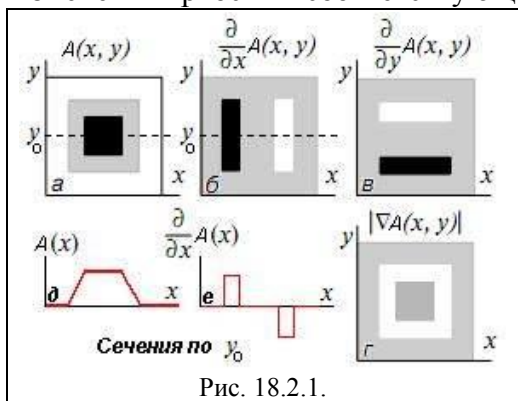


Рис. 18.2.1.

Выделение перепадов яркости иллюстрирует рис. 18.2.1. На нем можно видеть, что подчеркивание контуров, перпендикулярных к оси  $x$ , обеспечивает производная  $\partial A(x, y)/\partial x$  (рис. б), а подчеркивание контуров, перпендикулярных к оси  $y$ ,  $-\partial A(x, y)/\partial y$  (рис. в).

В практических задачах требуется выделять контуры, направление которых является произвольным. Для этих целей можно использовать модуль градиента функции яркости

$$|\nabla A(x, y)| = \sqrt{\left(\frac{\partial A(x, y)}{\partial x}\right)^2 + \left(\frac{\partial A(x, y)}{\partial y}\right)^2},$$

который пропорционален максимальной (по направлению) скорости изменения функции яркости в данной точке и не зависит от направления контура. Модуль градиента в отличие от частных производных принимает только неотрицательные значения, поэтому на получающемся изображении (рис. г) точки, соответствующие контурам, имеют

повышенный уровень яркости.

Для цифровых изображений аналогами частных производных и модуля градиента являются разностные функции.

Практический пример выделения границ на фотоизображении приведен на рис. 18.2.2. Исходное изображение (1) является однотонным. На изображении (2) представлен результат вычисления вектора градиента яркости  $\nabla A(x, y) = (\partial A/\partial x, \partial A/\partial y)$ . Как видно на рисунке, в точках большого перепада яркости градиент имеет большую длину. Отфильтровав пиксели с длиной градиента, большей определенного порога  $\alpha$ , мы получим изображение границ (3).

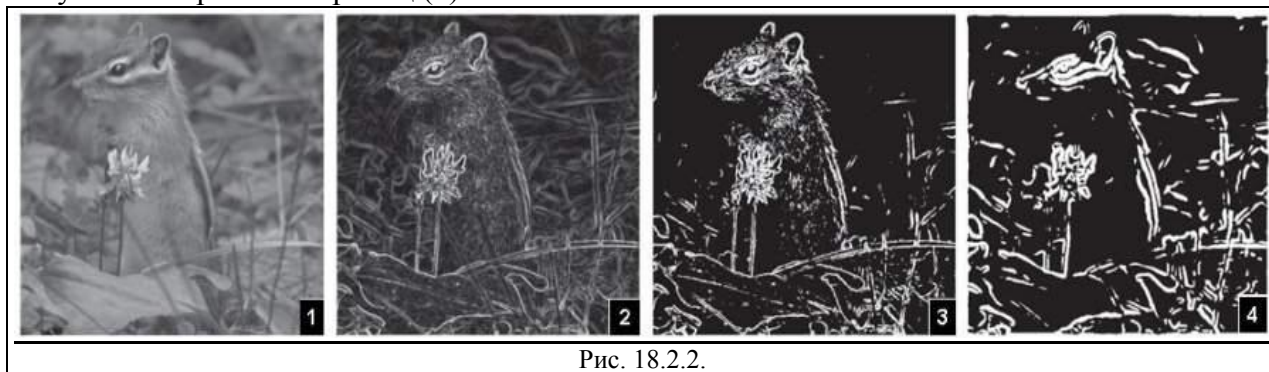


Рис. 18.2.2.

Недостаток алгоритма - пропуск границы с малыми перепадами яркости и включение в число границ деталей изображения с большими изменениями яркости (шкурка бурундука). При зашумлении изображения карту граничных точек будут загрязнять и просто шум, поскольку не учитывается, что граничные точки соответствуют не просто перепадам яркости, а перепадам яркости между относительно монотонными областями.

Для снижения влияния данного недостатка изображение сначала подвергают сглаживающей гауссовской фильтрации. При сглаживающей фильтрации мелкие несущественные детали размываются быстрее перепадов между областями. Результат операции можно видеть на изображении (4). Однако при этом четко выраженные границы расплываются в жирные линии.

Градиент яркости в каждой точке характеризуется длиной и направлением. Выше при поиске граничных точек использовалась только длина вектора. Направление градиента - это направление максимального возрастания функции, что позволяет использовать процедуру подавления немаксимумов. При этой процедуре для каждой точки рассматривается отрезок длиной в несколько пикселей, ориентированный по направлению градиента и с центром в рассматриваемом пикселе. Пиксель считается максимальным тогда и только тогда, когда длина градиента в нем максимальна среди всех длин градиентов пикселей отрезка. Граничными можно признать все максимальные пиксели с длинами градиента больше определенного порога. Градиент яркости в каждой точке перпендикулярен границе, поэтому после подавления немаксимумов жирных линий не остается. На каждом перпендикулярном сечении жирной линии останется один пиксель с максимальной длиной градиента.

Перпендикулярность градиента яркости к границе может быть использована для прослеживания границы, начиная с некоторого граничного пикселя. Такое прослеживание используется в гистерезисной фильтрации максимальных пикселей. Идея гистерезисной фильтрации заключается в том, что длинный устойчивый граничный контур, скорее всего, содержит в себе пиксели с особенно большим перепадом яркости, и, начиная с такого пикселя, контур можно проследить, переходя по граничным пикселям с меньшим перепадом яркости.



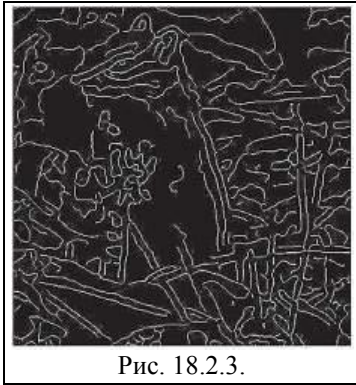


Рис. 18.2.3.

При проведении гистерезисной фильтрации вводят не одно, а два пороговых значения. Меньшее ( $\alpha$ ) соответствует минимальной длине градиента, при которой пиксель может быть признан граничным. Большее ( $\beta$ ), соответствует минимальной длине градиента, при которой пиксель может инициализировать контур. После того как контур инициализируется в максимальном пикселе  $P$  с длиной градиента, большей  $\beta$ , рассматриваются каждый соседний с ним максимальный пиксель  $Q$ . Если пиксель  $Q$  имеет длину градиента, большую  $\alpha$ , и угол между векторами  $PQ$  и  $\nabla(P)$  близок к  $90^\circ$ , то  $P$  добавляется к контуру, и процесс рекурсивно

переходит к  $Q$ . Его результат для исходного изображения на рис. 18.2.2 показан на рис. 18.2.3.

Таким образом, алгоритм нахождения границ на основе градиента заключается в последовательном применении следующих операций:

- гауссовская сглаживающая фильтрация;
- нахождение градиента яркости в каждом пикселе;
- нахождение максимальных пикселей;
- гистерезисная фильтрация максимальных пикселей.

Этот алгоритм носит названия алгоритма Кэнни и наиболее часто применяется для нахождения границ.

**Поиск границ на основе лапласиана.** Известно, что необходимым и достаточным условием экстремального значения первой производной функции в произвольной точке является равенство нулю второй производной в этой точке, причем вторая производная должна иметь разные знаки по разные стороны от точки.

В двумерном варианте аналогом второй производной является лапласиан - скалярный оператор  $\nabla^2(f) = (\partial^2 f / \partial x^2 + \partial^2 f / \partial y^2)$ .

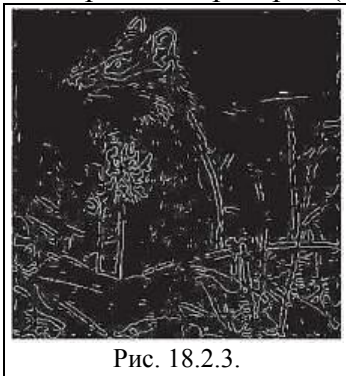


Рис. 18.2.3.

Нахождение границ на изображении с использованием лапласиана может производиться по аналогии с одномерным случаем: граничными признаются точки, в которых лапласиан равен нулю и вокруг которых он имеет разные знаки. Оценка лапласиана при помощи линейной фильтрации также предваряется гауссовской сглаживающей фильтрацией, чтобы снизить чувствительность алгоритма к шуму. Гауссовское сглаживание и поиск лапласиана можно осуществить одновременно, поэтому нахождение границ при помощи такого фильтра производится быстрее, чем при помощи алгоритма Кэнни. Фильтр применяется в системах, где имеет значение и

качество результата (обычно уступает алгоритму Кэнни), и быстродействие. Чтобы уменьшить чувствительность к несущественным деталям, из числа граничных точек также можно исключить те, длина градиента в которых меньше определенного порога (рис. 18.2.3).

### 18.3. ВЫДЕЛЕНИЕ ОБЪЕКТОВ НА ИЗОБРАЖЕНИИ

При работе с изображением часто возникает необходимость отделить одну, значимую для пользователя часть (объект), от всего остального (фон). Так, например, в медицинских изображениях выделяется определенный объект для измерений его параметров. Алгоритмы решения этой задачи являются интерактивными, т.к. только пользователь может определить, что является интересующим его объектом, а что относится к фону.

В принципе, поставленная задача может быть решена простейшим способом - пользователь просто обводит объект курсором мыши, но для сложных объектов это

потребуется значительных усилий и временных затрат.

**Алгоритм "Волшебная палочка"** (Magic wand) был одним из первых алгоритмов интерактивной сегментации. Пользователь последовательно указывает точки объекта, а алгоритм выделяет окрестные пиксели с похожим цветом и (по решению пользователя) добавляет выделенную область к объекту. Для оценки "похожести" пользователем задается матрица порогов чувствительности расстояния между цветами. Чем он больше, тем больше пикселей выделится на одном шаге. При этом в область "похожих" цветов выделяются только связные пиксели.

Алгоритм неплохо работает при выделении достаточно монотонных по цвету объектов. При сильных вариациях цвета аккуратно отделить объект от фона с помощью данного алгоритма невозможно. При слишком малом пороге чувствительности может не выделяться значительная часть объекта. Увеличение порога приводит к тому, что выделение "протекает" за пределы объекта. В случае пестрого объекта или размытой границы между фоном и объектом алгоритм практически беспомощен.

**Алгоритм "умные ножницы"** используется с 1996 года, завоевал популярность и был встроен в распространенный редактор фотоизображений Adobe Photoshop. При использовании алгоритма пользователь обводит границу между объектом и фоном, указывая точки на границе с некоторым промежутком, а "умные ножницы" проводят граничную линию между последовательно указанными точками.

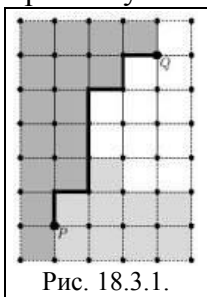


Рис. 18.3.1.

Представим себе растр изображения в виде графа (рис. 18.3.1) с ребрами, образованными сторонами пикселей. При указании пользователем двух последовательных точек P и Q алгоритм "ножниц" вычисляет минимальное расстояние между точками P и Q по ребрам графа, при этом условная геометрическая длина каждого ребра на этом пути имеет обратную зависимость от величины цветового перепада пикселей по его сторонам. Поскольку ребра, соответствующие резким цветовым перепадам, будут иметь меньшую условную длину, "умные ножницы" стремятся провести границу именно по таким ребрам.

"Умные ножницы" существенно ускоряют процесс выделения объекта. Однако и они работают не очень хорошо при наличии пестрого фона и/или пестрого объекта. В таких случаях требуется указывать большее количество граничных точек.

**Сегментация при помощи разрезов на графах.** Третий способ выделения объекта на фоне также основан на теории графов. Пользователь просто отмечает некоторое множество A пикселей, принадлежащих объекту, и некоторое множество B пикселей, принадлежащих фону. Поскольку эти пиксели не обязаны быть рядом с границей, такая разметка не требует от пользователя особых усилий. Результатом алгоритма служит сегментация, в которой все множество A относится к объекту, а множество B - к фону.

Если результат выделения с первого раза не удовлетворяет пользователя, он добавляет в исходные множества пиксели, доотмечая их на изображении. Например, если алгоритм ошибочно отнес кусок объекта к фону, пользователь отмечает часть пикселей этого куска как пиксели объекта (множество A). Результатом перезапуска алгоритма служит уточненная сегментация.

Рассмотрим, как работает алгоритм. Построим граф на растре следующим образом. Пиксельные вершины графа расположим в центре каждого пикселя, а под цветом вершины мы будем понимать цвет пикселя. Каждую вершину соединим с соседними вершинами и получим восемь ребер, которые соединяют центры соседних пикселей. Припишем каждому ребру вес:

$$(\lambda/L) \exp(-\sigma \rho(C_1, C_2)),$$

где L - геометрическая длина ребра, C<sub>1</sub> и C<sub>2</sub> - цвета вершин, соединяемых ребром, λ и σ - некоторые (положительные) параметры, ρ(..) - матрица чувствительности по перепаду цвета. Данный вес тем меньше, чем больше разница между цветами вершин.



Добавим в граф две терминальных вершины, называемые истоком и стоком, и соединим их ребрами с каждой вершиной графа. Ребрам, соединяющим исток с вершинами множества А, и ребрам, соединяющим сток с вершинами множества В, припишем бесконечный вес.

Рассмотрим распределение цветов вершин множества А (например, как гистограмму). Для всех пиксельных вершин не из множества А, припишем ребрам, соединяющим их с истоком, вес, пропорциональный согласованности их цвета с этим распределением цветов, при этом вес ребра будет тем больше, чем больше "похож" цвет вершины на цвета вершин множества А. Аналогичную процедуру сделаем для множества В и ребер, соединяющих пиксельные вершины со стоком.

Все ребра графа "разрежем" на два непересекающихся множества - истоковое и стоковое, и будем считать, что вершины, попавшие в истоковое множество, соответствуют пикселям объекта, а остальные, попавшие в стоковое множество соответствуют пикселям фона. Число возможных вариантов разрезов равно  $2^P$ , где P - число пикселей, так как каждую пиксельную вершину можно отнести либо в истоковое, либо в стоковое множество.



Рис. 18.3.2.

Весом разреза назовем сумму весов всех разрезанных ребер, за исключением ребер с бесконечным весом. Минимальным разрезом назовем разрез с минимальным весом, при этом истоковые пиксели этого разреза будут соответственно отнесены к пикселям объекта, а стоковые – к фону. Граница между объектом и фоном будет проведена по возможности между пикселями с сильно отличающимися цветами.

Идеального разделения, естественно, быть не может. Например, участок изображения может быть похож по цвету на фон (пиксели множества В), но окружен пикселями множества А и не отделен от них резкой границей. В таких случаях выбор

параметра  $\lambda$  в формуле веса ребер устанавливает баланс между последними двумя пунктами. При увеличении значения  $\lambda$ , увеличивается важность того, чтобы граница между фоном и объектом проходила между пикселями с разными цветами, а при уменьшении - увеличивается важность того, чтобы пиксели, похожие по цвету на пиксели множества А (или В), были отнесены к объекту (фону). Пример выделения объекта приведен на рис. 18.3.2.

#### 18.4. ВЫДЕЛЕНИЕ ПРИЗНАКОВ ОБЪЕКТОВ

Выделение признаков позволяет упростить реализацию распознавания или идентификации объектов. При выборе наиболее информативных признаков необходимо учитывать как свойства самих объектов, так и возможности разрешающей способности первичных формирователей сигнала изображения. Выделение признаков проведем на примере обработки монохромных (однослойных) изображений. В цветных изображениях рассмотренные алгоритмы можно применять к каждому цвету в отдельности.

При обработке предпочтительными являются следующие признаки объектов:

- площадь и периметр изображения объекта;
- размеры вписанных простейших геометрических фигур (окружностей, прямоугольников, треугольников и др.);
- число и взаимное расположение углов;
- моменты инерции изображений объектов.

Важной особенностью большинства геометрических признаков является инвариантность относительно разворота изображения объекта, а путем нормирования геометрических признаков друг относительно друга достигается инвариантность относительно масштаба изображения объекта.

**Определение площади и периметра.** Площадь изображения объекта вычисляется путём подсчёта числа элементов, относящихся к объекту:

$$A = \sum_x \sum_y S(x, y), \quad S(x, y) = \begin{cases} 1, & S(x, y) \in L \\ 0, & S(x, y) \notin L \end{cases}$$

где  $L$  - множество координат массива  $S(x, y)$ , принадлежащих выделенному объекту.

Периметр изображения объекта  $P$  вычисляется после выделения границ объекта путем суммирования множество контурных точек изображения объекта.

На основе выделенных признаков нормированный признак, инвариантный к масштабу изображения  $U = A/P^2$  или  $V = P/\sqrt{A}$ .

**Определение радиусов вписанных и описанных окружностей** (рис. 18.4.1) складывается из двух этапов.

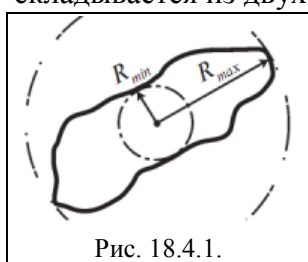


Рис. 18.4.1.

1. Определение координат геометрического центра изображения объекта:

$$X_{\text{ц}} = \sum_x \sum_y xS(x, y) / \sum_x \sum_y S(x, y), \quad Y_{\text{ц}} = \sum_x \sum_y yS(x, y) / \sum_x \sum_y S(x, y),$$

где  $x$  и  $y$  – номера строк и столбцов всех пикселей  $S(x, y)$ , входящих в объект.

2. Вычисление минимального и максимального расстояний от центра до границ изображения объекта.

$$r(x, y) = \sqrt{(x - X_{\text{ц}})^2 + (y - Y_{\text{ц}})^2}.$$

$$R_{\text{max}} = r(x, y_{\text{max}}); \quad R_{\text{min}} = r(x, y_{\text{min}}), \quad \text{где } x, y \in P \text{ (точки периметра).}$$

Нормированный признак  $R' = R_{\text{max}} / R_{\text{min}}$  инвариантен к масштабу изображения объекта.

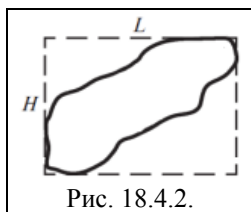


Рис. 18.4.2.

**Определение сторон описанного прямоугольника** (рис. 18.4.2) выполняется следующим образом. Определяются максимальные и минимальные значения абсцисс и ординат изображения объекта  $x_{\text{max}}$  и  $x_{\text{min}}$ ,  $y_{\text{max}}$  и  $y_{\text{min}}$ , после чего определяются высота и основание прямоугольника:

$$L = x_{\text{max}} - x_{\text{min}}, \quad H = y_{\text{max}} - y_{\text{min}}.$$

Данный признак не инвариантен к развороту изображения объекта.

**Определение числа и взаимного положения углов.** Классический способ определения угловых точек изображения объекта заключается в анализе небольшого фрагмента контура в окрестностях данной точки и в определении радиуса её кривизны. Если радиус окажется меньше установленного порога – это угловой элемент, в противном случае – нет.

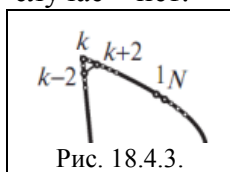


Рис. 18.4.3.

С практической точки зрения предпочтительным представляется более простой алгоритм. Он заключается в оценке расстояний между начальной и конечной точками фрагмента контура, например, между элементами контура с порядковыми номерами  $k - 2$  и  $k + 2$  на рис. 18.4.3.

Пусть  $x(k)$  и  $y(k)$  абсцисса и ордината контурных элементов соответственно. Тогда решающее правило может выглядеть следующим образом:

$$\{|x(k-2) - x(k+2)| + |y(k-2) - y(k+2)| \leq N\}.$$

Если условие выполняется, тогда данная точка контура принадлежит множеству угловых точек  $L$ . Здесь  $N$  – пороговое значение, выбираемое с учётом свойств объектов.

**Определение моментов инерции объекта.** Термин "моменты инерции изображения объекта" не имеет отношения к механике. Просто для вычисления указанного признака используются математические выражения, аналогичные вычислению моментов инерции материального тела, где вместо значений масс отдельных точек тела подставлены значения освещенностей в соответствующих точках его изображения. Моменты инерции являются информационными признаками для последующего распознавания образов.

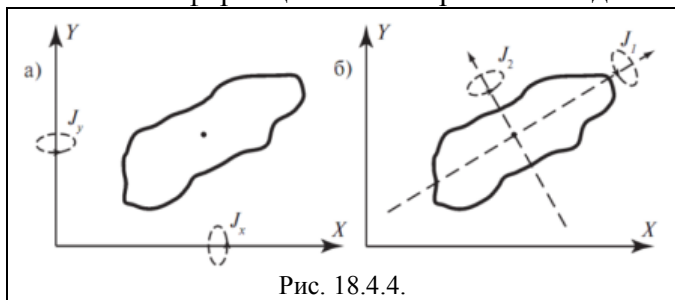


Рис. 18.4.4.

Обозначим главные искомые моменты инерции изображения объекта через  $J_1$  и  $J_2$ . Чтобы найти  $J_1$  и  $J_2$ , необходимо предварительно определить так называемые промежуточные моменты  $J_x$  и  $J_y$ , т.е. моменты инерции относительно вертикальной и горизонтальной осей системы

координат, а также смешанный момент  $J_{x,y}$  (рис. 18.4.4а).

Порядок вычислений:

1. Определяются координаты центра "тяжести" (энергетического центра) изображения объекта.

$$X_{цэ} = \sum_x \sum_y xA(x, y) / \sum_x \sum_y A(x, y), \quad Y_{цэ} = \sum_x \sum_y yA(x, y) / \sum_x \sum_y A(x, y),$$

2. Определяются промежуточные моменты  $J_x, J_y, J_{x,y}$ .

$$J_x = \sum_x \sum_y [(x - X_{цэ})^2 A(x, y)], \quad J_y = \sum_x \sum_y [(y - Y_{цэ})^2 A(x, y)],$$

$$J_{xy} = \sum_x \sum_y [(x - X_{цэ})(y - Y_{цэ}) A(x, y)].$$

3. Вычисляются главные моменты.

$$J_{1,2} = \frac{J_x + J_y}{2} \pm \sqrt{\frac{1}{4}(J_x - J_y)^2 + J_{x,y}^2}.$$

## 18.5. ОБНАРУЖЕНИЕ И РАСПОЗНАВАНИЕ ОБЪЕКТОВ ИЗОБРАЖЕНИЙ

**Обнаружение объектов.** Обнаружение (выявление) образов объектов в искажённом шумами и помехами изображении может быть определено в виде процедуры проверки определенного условия сравнения двух чисел - результата преобразования анализируемого изображения и определенного порогового значения:

$$L[A(x, y)] \geq \Pi[A(x, y)]. \quad (18.5.1)$$

Здесь  $L[.]$  – оператор преобразования исходного изображения,  $\Pi[.]$  – оператор формирования порогового значения. Решение о наличии объекта принимается в случае выполнения условия. Качество обнаружения характеризуется вероятностью выполнения условия при наличии объекта в анализируемом изображении.

Конкретный вид операторов  $L[.]$ ,  $\Pi[.]$  и качество обнаружения зависят от наличия априорных сведений об ожидаемых объектах, шумах, помехах и искажениях. Основой для определения оптимальных параметров операторов является теория статистических решений.

Так, например, при выделении точечных объектов форма изображения объекта однозначно определяется функцией пятна рассеяния объектива, которую можно считать известной, при этом задача сводится к классической процедуре обнаружения сигнала известной формы на фоне аддитивных нормальных шумов с нулевым средним значением. В этом случае в качестве операторов преобразования выступают корреляционные интегралы, вычисляемые с использованием заданного описания известного изображения объекта и типового шума соответственно. Идентификация заключается в сравнении изображения объекта с эталонами заданного класса. Решение об объекте выносится по наилучшему совпадению.

**Способ прямого сравнения объекта с эталонным изображением.** Пусть  $S(x, y)$  –

исходное изображение объекта,  $F(x, y)$  – эталонное изображение. Алгоритм прямого сравнения имеет вид:

$$T = \sum_x \sum_y (S(x, y) - F(x, y))^2 \leq D, \quad (18.5.2)$$

где  $D$  – порог различия.

Способ прост и может быть легко реализован. Однако, при наличии в реальных условиях дестабилизирующих факторов надёжность способа невелика. При большом значении порога  $D$  условию (18.5.2), могут удовлетворять различные объекты и могут возникнуть ошибки, связанные с неправильной идентификацией объекта (ошибки первого рода). При уменьшении  $D$  могут возникнуть ошибки типа пропуска объекта (ошибки второго рода). Регулируя величину  $D$ , можно лишь менять соотношение между вероятностями возникновения ошибок первого и второго рода в соответствии с заданным критерием оптимальности.

**Корреляционный метод** основан на вычислении взаимной корреляции между объектами и эталонами. Из множества  $k$  альтернативных вариантов выбирается тот объект (или эталон), при котором получается максимальная сумма взаимной корреляции:

$$K(k) = \sum_x \sum_y S_k(x, y) F_k(x, y). \quad (18.5.3)$$

При идентификации объектов удобно пользоваться коэффициентами корреляции, которые в первом приближении дают и оценку вероятности отнесения объекта к данному эталону:

$$R(k) = K(k)/K_{\max}(k), \quad K_{\max}(k) = \sum_x \sum_y F_k^2(x, y), \quad (18.5.3)$$

где  $K_{\max}(k)$  – значения автокорреляции эталонов.

Корреляционный метод более надёжен, но требует значительно большего объёма вычислений. Но при обработке бинарных изображений это не столь существенно, поскольку перемножение однобитовых чисел сводится к простой логической операции «И».

Рассмотренные выше методы требуют одинаковой ориентации изображений объекта и эталона, совмещения их по пространственным координатам и выдерживания одинаковых масштабов.

**Методы распознавания на основе системы признаков** также используют эталоны объектов, но в качестве элементов сравнения используются признаки объекта и эталона, что позволяет сократить объём эталонных данных и время обработки информации. Однако следует иметь в виду, что на практике выделение признаков объектов всегда осуществляется с некоторой погрешностью, а, следовательно, необходимо выявлять и учитывать характер и степень возможного рассеяния оценок используемых признаков для каждого из ожидаемых объектов, т.е. использовать гистограммы распределения значений признаков.

При большом числе возможных вариантов объектов рекомендуется многоступенчатый (иерархический) алгоритм. При этом на каждой ступени распознавания используется какой-либо из признаков объекта (площадь, периметр, радиусы вписанных и описанных окружностей, моменты инерции, число и расположение углов и т.д.). На нижних уровнях используются признаки, не требующие больших вычислительных затрат (например, площади и периметры объектов), а наиболее информативные (например, моменты инерции) – применяются на верхнем уровне, где число альтернатив минимально.

## Тема 19. ОСНОВЫ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ СИГНАЛОВ.

Вейвлетное преобразование сигналов является обобщением спектрального анализа, типичный представитель которого – классическое преобразование Фурье. Термин "вейвлет" (wavelet) в переводе с английского означает "маленькая (короткая) волна". Вейвлеты - это обобщенное название семейств математических функций определенной

формы, которые локальны во времени и по частоте, и в которых все функции получаются из одной базовой (порождающей) посредством ее сдвигов и растяжений по оси времени. Вейвлет-преобразования рассматривают анализируемые временные функции в терминах колебаний, локализованных по времени и частоте. Как правило, вейвлет-преобразования (WT) подразделяют на дискретное (DWT) и непрерывное (CWT).

DWT используется для преобразований и кодирования сигналов, CWT - для анализа сигналов. Вейвлет-преобразования в настоящее время принимаются на вооружение для огромного числа разнообразных применений, нередко заменяя обычное преобразование Фурье. Это наблюдается во многих областях, включая молекулярную динамику, квантовую механику, астрофизику, геофизику, оптику, компьютерную графику и обработку изображений, анализ ДНК, исследования белков, исследования климата, общую обработку сигналов и распознавание речи.

Вейвлетный анализ представляет собой особый тип линейного преобразования сигналов и физических данных. Базис собственных функций, по которому проводится вейвлетное разложение сигналов, обладает многими специфическими свойствами и возможностями. Вейвлетные функции базиса позволяют сконцентрировать внимание на тех или иных локальных особенностях анализируемых процессов, которые не могут быть выявлены с помощью традиционных преобразований Фурье и Лапласа. К таким процессам в геофизике относятся поля различных физических параметров природных сред. В первую очередь это касается полей температуры, давления, профилей сейсмических трасс и других физических величин.

Вейвлеты имеют вид коротких волновых пакетов с нулевым средним значением, локализованных по оси аргументов (независимых переменных), инвариантных к сдвигу и линейных к операции масштабирования (сжатия/растяжения). По локализации во временном и частотном представлении вейвлеты занимают промежуточное положение между гармоническими функциями, локализованными по частоте, и функцией Дирака, локализованной во времени.

Теория вейвлетов не является фундаментальной физической теорией, но она дает удобный и эффективный инструмент для решения многих практических задач. Основная область применения вейвлетных преобразований – анализ и обработка сигналов и функций, нестационарных во времени или неоднородных в пространстве, когда результаты анализа должны содержать не только частотную характеристику сигнала (распределение энергии сигнала по частотным составляющим), но и сведения о локальных координатах, на которых проявляют себя те или иные группы частотных составляющих или на которых происходят быстрые изменения частотных составляющих сигнала. По сравнению с разложением сигналов на ряды Фурье вейвлеты способны с гораздо более высокой точностью представлять локальные особенности сигналов, вплоть до разрывов 1-го рода (скачков). В отличие от преобразований Фурье, вейвлет-преобразование одномерных сигналов обеспечивает двумерную развертку, при этом частота и координата рассматриваются как независимые переменные, что дает возможность анализа сигналов сразу в двух пространствах.

Одна из главных и особенно плодотворных идей вейвлетного представления сигналов на различных уровнях декомпозиции (разложения) заключается в разделении функций приближения к сигналу на две группы: аппроксимирующую - грубую, с достаточно медленной временной динамикой изменений, и детализирующую - с локальной и быстрой динамикой изменений на фоне плавной динамики, с последующим их дроблением и детализацией на других уровнях декомпозиции сигналов. Это возможно как во временной, так и в частотной областях представления сигналов вейвлетами.

### **1.1. ИСТОКИ ВЕЙВЛЕТ - ПРЕОБРАЗОВАНИЯ**

**Историческая справка.** История спектрального анализа восходит к И. Бернулли, Эйлеру и Фурье, который впервые построил теорию разложения функций в

тригонометрические ряды. Однако это разложение долгое время применялось как математический прием и не связывалось с какими-либо физическими понятиями. Спектральные представления применялись и развивались сравнительно узким кругом физиков–теоретиков. Однако, начиная с 20-х годов прошлого века, в связи с бурным развитием радиотехники и акустики, спектральные разложения приобрели физический смысл и практическое применение. Основным средством анализа реальных физических процессов стал гармонический анализ, а математической основой анализа - преобразование Фурье. Преобразование Фурье разлагает произвольный процесс на элементарные гармонические колебания с различными частотами, а все необходимые свойства и формулы выражаются с помощью одной базисной функции  $\exp(j\omega t)$  или двух действительных функций  $\sin(\omega t)$  и  $\cos(\omega t)$ . Гармонические колебания имеют широкое распространение в природе, и поэтому смысл преобразования Фурье интуитивно понятен независимо от математической аналитики.

Преобразование Фурье обладает рядом замечательных свойств. Областью определения преобразования является пространство  $L^2$  интегрируемых с квадратом функций, и многие физические процессы в природе можно считать функциями, принадлежащими этому пространству. Для применения преобразования разработаны эффективные вычислительные процедуры типа быстрого преобразования Фурье (БПФ). Эти процедуры входят в состав всех пакетов прикладных математических программ и реализованы аппаратно в процессорах обработки сигналов.

Было также установлено, что функции можно разложить не только по синусам и косинусам, но и по другим ортогональным базисным системам, например, полиномам Лежандра и Чебышева, функциям Лагерра и Эрмита. Однако практическое применение они получили только в последние десятилетия XX века благодаря развитию вычислительной техники и методов синтеза цифровых линейных систем обработки данных. Непосредственно для целей спектрального анализа подобные ортогональные функции не нашли широкого применения из-за трудностей интерпретации получаемых результатов. По тем же причинам не получили развития в спектральном анализе функции типа "прямоугольной волны" Уолша, Радемахера, и пр.

Теоретические исследования базисных систем привели к созданию теории обобщенного спектрального анализа, которая позволила оценить пределы практического применения спектрального анализа Фурье, создала методы и критерии синтеза ортогональных базисных систем. Иллюстрацией этому является активно развивающаяся с начала 80-х годов прошлого столетия теория базисных функций типа вейвлет. Благодаря прозрачности физической интерпретации результатов анализа, сходной с "частотным" подходом в преобразовании Фурье, ортогональный базис вейвлетов стал популярным и эффективным средством анализа сигналов и изображений в акустике, сейсмике, медицине и других областях науки и техники.

Вейвлет-анализ является разновидностью спектрального анализа, в котором роль простых колебаний играют функции особого рода, называемые вейвлетами. Базисная функция вейвлет – это некоторое "короткое" колебание, но не только. Понятие частоты спектрального анализа здесь заменено масштабом, и, чтобы перекрыть "короткими волнами" всю временную ось, введен сдвиг функций во времени. Базис вейвлетов – это функции типа  $\psi((t-b)/a)$ , где  $b$  - сдвиг,  $a$  – масштаб. Функция  $\psi(t)$  должна иметь нулевую площадь и, еще лучше, равными нулю первый, второй и прочие моменты. Фурье-преобразование таких функций равно нулю при  $\omega=0$  и имеет вид полосового фильтра. При различных значениях масштабного параметра 'a' это будет набор полосовых фильтров. Семейства вейвлетов во временной или частотной области используются для представления сигналов и функций в виде суперпозиций вейвлетов на разных масштабных уровнях декомпозиции (разложения) сигналов.



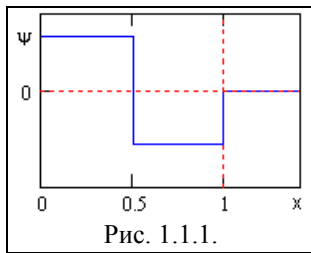


Рис. 1.1.1.

Первое упоминание о подобных функциях (которые вейвлетами не назывались) появилось в работах Хаара (Haar) еще в начале прошлого века. Вейвлет Хаара - это короткое прямоугольное колебание на интервале  $[0, 1]$ , показанное на рис. 1.1.1. Однако он интересен больше теоретически, так как не является непрерывно дифференцируемой функцией и имеет длинные "хвосты" в частотной области. В 30-е годы физик Paul Levy, исследуя броуновское движение, обнаружил, что базис Хаара лучше, чем базис Фурье, подходит для изучения деталей броуновского движения.

Сам термин "вейвлет", как понятие, ввели в своей статье J. Morlet и A. Grossman, опубликованной в 1984 г. Они занимались исследованиями сейсмических сигналов с помощью базиса, который и назвали вейвлетом. Весомый вклад в теорию вейвлетов внесли Гуппилауд, Гроссман и Морлет, сформулировавшие основы CWT, Ингрид Добеши, разработавшая ортогональные вейвлеты (1988), Натали Делпрат, создавшая время-частотную интерпретацию CWT (1991), и многие другие. Математическая формализация в работах Mallat и Meuer привела к созданию теоретических основ вейвлет-анализа, названного мультиразрешающим (кратномасштабным) анализом.

В настоящее время специальные пакеты расширений по вейвлетам присутствуют в основных системах компьютерной математики (Matlab, Mathematica, Mathcad, и др.), а вейвлет-преобразования и вейвлетный анализ используются во многих областях науки и техники для самых различных задач. Многие исследователи называют вейвлет-анализ "математическим микроскопом" для точного изучения внутреннего состава и структур неоднородных сигналов и функций.

Не следует рассматривать вейвлет-методы обработки и анализа сигналов в качестве новой универсальной технологии решения любых задач. Возможности вейвлетов еще не раскрыты полностью, однако это не означает, что их развитие приведет к полной замене традиционных средств обработки и анализа информации, хорошо отработанных и проверенных временем. Вейвлеты позволяют расширить инструментальную базу информационных технологий обработки данных.

**Преобразование Фурье (ПФ).** В основе спектрального анализа сигналов лежит интегральное преобразование и ряды Фурье. Напомним некоторые математические определения.

В пространстве функций, заданных на конечном интервале  $(0, T)$ , норма, как числовая характеристика произвольной функции  $s(t)$ , вычисляется как корень квадратный из скалярного произведения функции. Для комплексных функций, квадрат нормы (энергия сигнала) соответствует выражению:

$$\|s(t)\|^2 = \langle s(t), s(t) \rangle = \int_0^T s(t) s^*(t) dt, \quad (1.1.1)$$

где  $s^*(t)$  – функция, комплексно сопряженная с  $s(t)$ .

Если норма функции имеет конечное значение (интеграл сходится), то говорят, что функция принадлежит пространству функций  $L^2[R]$ ,  $R=[0, T]$ , интегрируемых с квадратом (пространство Гильберта), и имеет конечную энергию. В пространстве Гильберта на основе совокупности ортогональных функций с нулевым скалярным произведением

$$\langle v(t), w(t) \rangle = \int_0^T v(t) w^*(t) dt = 0$$

может быть создана система ортонормированных "осей" (базис пространства), при этом любой сигнал, принадлежащий этому пространству, может быть представлен в виде весовой суммы проекций сигнала на эти "оси" – базисных векторов. Значения проекций определяются скалярными произведениями сигнала с соответствующими функциями базисных "осей".

Базис пространства может быть образован любой ортогональной системой

функций. Наибольшее применение в спектральном анализе получила система комплексных экспоненциальных функций. Проекция сигнала на данный базис определяются выражением:

$$S_n = (1/T) \int_0^T s(t) \exp(-jn\Delta\omega t) dt, \quad n \in (-\infty, \infty), \quad (1.1.2)$$

где  $\Delta\omega=2\pi/T$  – частотный аргумент векторов. При известных выражениях базисных функций сигнал  $s(t)$  однозначно определяется совокупностью коэффициентов  $S_n$  и может быть абсолютно точно восстановлен (реконструирован) по этим коэффициентам:

$$s(t) = \sum_{n=-\infty}^{\infty} S_n \exp(jn\Delta\omega t). \quad (1.1.3)$$

Уравнения (1.1.2) и (1.1.3) называют прямым и обратным преобразованием Фурье сигнала  $s(t)$ . Любая функция гильбертова пространства может быть представлена в виде комплексного ряда Фурье (1.1.3), который называют спектром сигнала или его Фурье-образом.

Ряд Фурье ограничивается определенным количеством членов  $N$ , что означает аппроксимацию с определенной погрешностью бесконечномерного сигнала  $N$  – мерной системой базисных функций спектра сигнала. Ряд Фурье равномерно сходится к  $s(t)$ :

$$\lim_{N \rightarrow \infty} \left\| s(t) - \sum_{n=-N}^N S_n \exp(jn\Delta\omega t) \right\| = 0. \quad (1.1.4)$$

Таким образом, ряд Фурье - это разложение сигнала  $s(t)$  по базису пространства  $L^2(0,T)$  ортонормированных гармонических функций  $\exp(jn\Delta\omega t)$  с изменением частоты, кратным частоте первой гармоники  $\omega_1=\Delta\omega$ . Отсюда следует, что ортонормированный базис пространства  $L^2(0,T)$  построен из одной функции  $\exp(j\Delta\omega t) = \cos(\Delta\omega t) + j \cdot \sin(\Delta\omega t)$  с помощью масштабного преобразования независимой переменной.

Для коэффициентов ряда Фурье справедливо равенство Парсеваля сохранения энергии сигнала в различных представлениях:

$$(1/T) \int_0^T |s(t)|^2 dt = \sum_{n=-\infty}^{\infty} |S_n|^2. \quad (1.1.5)$$

Разложение в ряд Фурье произвольной функции  $y(t)$  корректно, если функция  $y(t)$  принадлежит этому же пространству  $L^2(0,T)$ , т.е. квадратично интегрируема с конечной энергией:

$$\int_0^T |y(t)|^2 dt < \infty, \quad t \in (0,T), \quad (1.1.6)$$

при этом она может быть периодически расширена и определена на всей временной оси пространства  $R(-\infty, \infty)$  так, что

$$y(t) = y(t-T), \quad t \in R,$$

при условии сохранения конечности энергии в пространстве  $R(-\infty, \infty)$ .

С позиций анализа произвольных сигналов и функций в частотной области и точного восстановления после преобразований можно отметить ряд недостатков разложения сигналов в ряды Фурье, которые привели к появлению оконного преобразования Фурье и стимулировали развитие вейвлетного преобразования. Основные из них:

- Ограниченная информативность анализа нестационарных сигналов и практически полное отсутствие возможностей анализа их особенностей (сингулярностей), т.к. в частотной области происходит «размазывание» особенностей сигналов (разрывов, ступенек, пиков и т.п.) по всему частотному диапазону спектра.
- Гармонические базисные функции разложения не способны отображать перепады сигналов с бесконечной крутизной типа прямоугольных импульсов, т.к. для этого требуется бесконечно большое число членов ряда. При ограничении числа членов ряда



Фурье в окрестностях скачков и разрывов при восстановлении сигнала возникают осцилляции (явление Гиббса).

- Преобразование Фурье отображает глобальные сведения о частотах исследуемого сигнала и не дает представления о локальных свойствах сигнала при быстрых временных изменениях его спектрального состава. Так, например, преобразование Фурье не различает стационарный сигнал с суммой двух синусоид от нестационарного сигнала с двумя последовательно следующими синусоидами с теми же частотами, т.к. спектральные коэффициенты (1.1.2) вычисляются интегрированием по всему интервалу задания сигнала. Преобразование Фурье не имеет возможности анализировать частотные характеристики сигнала в произвольные моменты времени.

**Оконное преобразование Фурье.** Частичным выходом из этой ситуации является оконное преобразование Фурье с движущейся по сигналу оконной функцией, имеющей компактный носитель. Временной интервал сигнала разделяется на подинтервалы и преобразование выполняется последовательно для каждого подинтервала в отдельности. Тем самым осуществляется переход к частотно-временному (частотно-координатному) представлению сигналов, при этом в пределах каждого подинтервала сигнал "считается" стационарным. Результатом оконного преобразования является семейство спектров, которым отображается изменение спектра сигнала по интервалам сдвига окна преобразования. Это позволяет выделять на координатной оси и анализировать особенности нестационарных сигналов. Размер носителя оконной функции  $w(t)$  обычно устанавливается соизмеримым с интервалом стационарности сигнала. По существу, таким преобразованием один нелокализованный базис разбивается на определенное количество базисов, локализованных в пределах функции  $w(t)$ , что позволяет представлять результат преобразования в виде функции двух переменных - частоты и временного положения окна.

Оконное преобразование выполняется в соответствии с выражением:

$$S(\omega, b_k) = \int_{-\infty}^{\infty} s(t) w^*(t-b_k) \exp(-j\omega t) dt. \quad (1.1.7)$$

Функция  $w^*(t-b)$  представляет собой функцию окна сдвига преобразования по координате  $t$ , где параметром  $b$  задаются фиксированные значения сдвига. При сдвиге окон с равномерным шагом значения  $b_k$  принимаются равными  $k\Delta b$ . В качестве окна преобразования может использоваться как простейшее прямоугольное окно, так и специальные весовые окна (Бартлетта, Гаусса, и пр.), обеспечивающие малые искажения спектра при вырезке оконных отрезков сигналов (нейтрализация явления Гиббса).

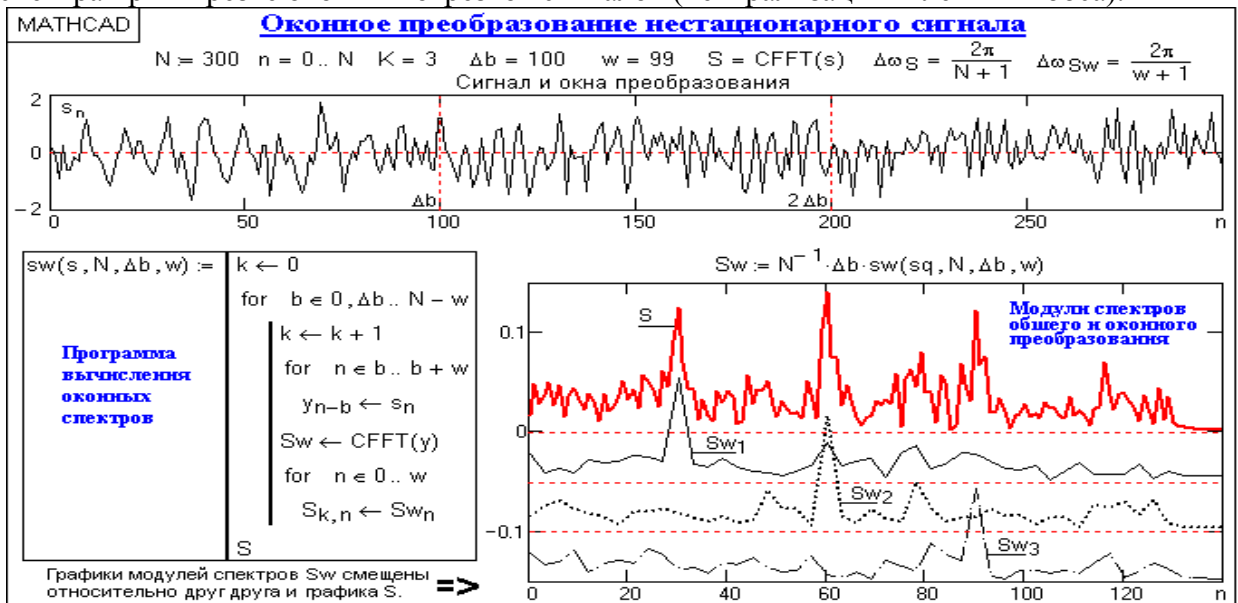


Рис. 1.1.2.

Пример оконного преобразования для нестационарного сигнала на большом уровне шума приведен на рис. 1.1.2. По спектру сигнала можно судить о наличии в его составе гармонических колебаний на трех частотах, определять соотношение между амплитудами этих колебаний и конкретизировать локальность колебаний по интервалу сигнала.

Координатная разрешающая способность оконного преобразования определяется шириной оконной функции и обратно пропорциональна частотной разрешающей способности. При ширине оконной функции, равной  $b$ , частотная разрешающая способность определяется значением  $\Delta\omega = 2\pi/b$ . При требуемой величине частотного разрешения  $\Delta\omega$  соответственно ширина оконной функции должна быть равна  $b = 2\pi/\Delta\omega$ . Для оконного преобразования Фурье эти ограничения являются принципиальными. Так, для рис. 1.1.2 при размере массива данных  $N = 300$  и ширине оконной функции  $\Delta b = 100$  частотная разрешающая способность результатов преобразования уменьшается в  $N/\Delta b = 3$  раза по сравнению с исходными данными, и графики  $Sw(n\Delta\omega_{sw})$  по координате  $n$  для наглядного сопоставления с графиком  $S(n\Delta\omega_s)$  построены с шагом по частоте  $\Delta\omega_{sw} = 3\Delta\omega_s$ , т.е. по точкам  $n = 0, 3, 6, \dots, N$ .

**Частотно-временное оконное преобразование** применяется для анализа нестационарных сигналов, если их частотный состав изменяется во времени. Функция оконного преобразования (23.1.1) может быть переведена в вариант с независимыми переменными и по времени, и по частоте:

$$S(t, \omega) = \int_{\tau} s(t-\tau) w(\tau) \exp(-j\omega\tau) dt. \quad (23.1.2)$$

Координатная разрешающая способность оконного преобразования определяется шириной оконной функции и, в силу принципа неопределенности Гейзенберга, обратно пропорциональна частотной разрешающей способности. Хорошая разрешающая способность по времени подразумевает небольшое окно времени, которому соответствует плохая частотная разрешающая способность и наоборот. Оптимальным считается ОПФ с гауссовым окном, которое получило название преобразование Габора (Gabor). Пример преобразования приведен на рис. 23.1.2 (в дискретном варианте вычислений).

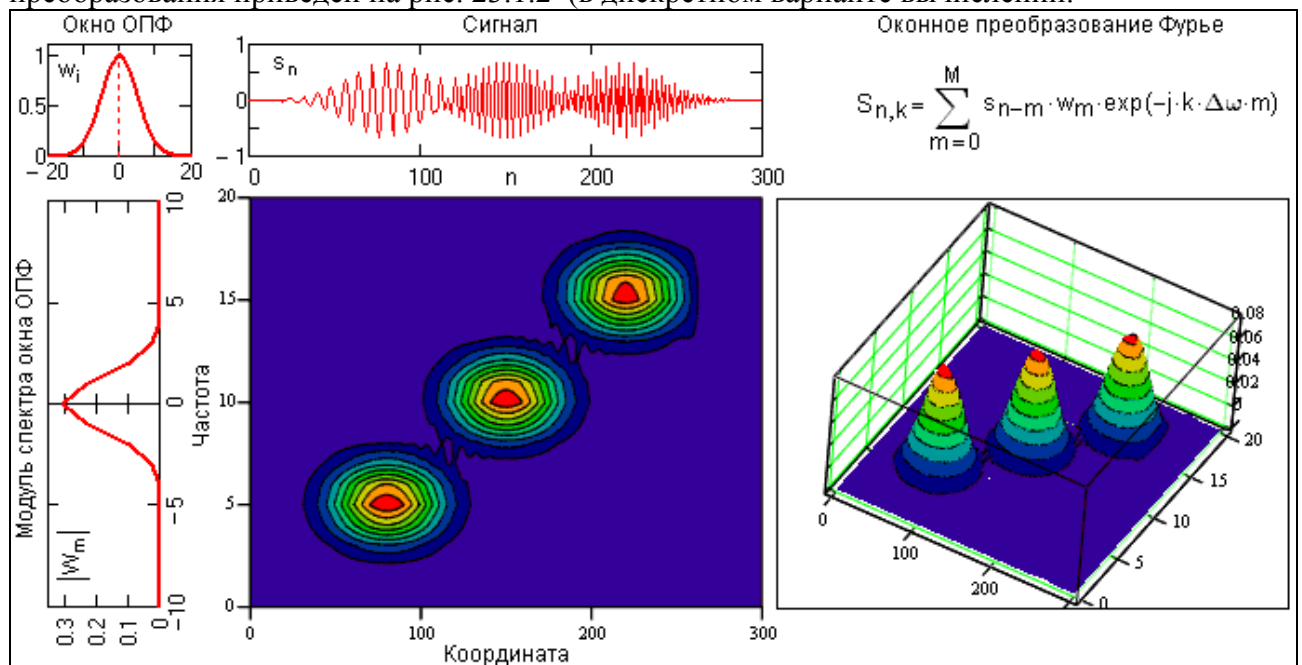


Рис. 23.1.2.

На рис. 23.1.3 приведен пример вычисления и представления (модуль правой части главного диапазона спектра) результатов спектрограммы при дискретном задании зашумленного входного сигнала  $sq(n)$ . Сигнал представляет собой сумму трех последовательных радиоимпульсов с разными частотами без пауз, с отношением

сигнал/шум, близким к 1. Оконная функция  $w_i$  задана в одностороннем варианте с эффективной шириной окна  $b \cong 34$  и полным размером  $M = 50$ . Установленный для результатов шаг по частоте  $\Delta\omega = 0.1$  несколько выше фактической разрешающей способности  $2\pi/M = 0.126$ . Для обеспечения работы оконной функции по всему интервалу сигнала задавались начальные и конечные условия вычислений (продление на  $M$  точек обоих концов сигнала нулевыми значениями).

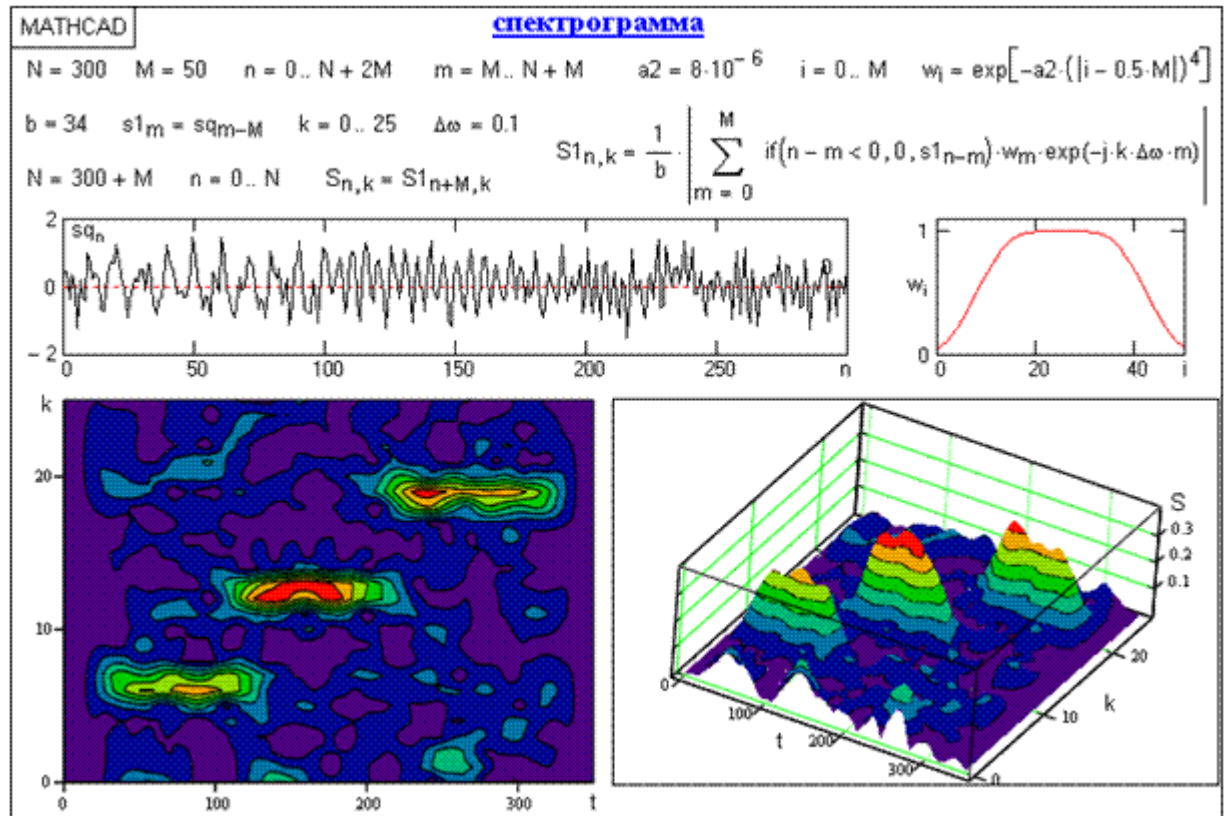
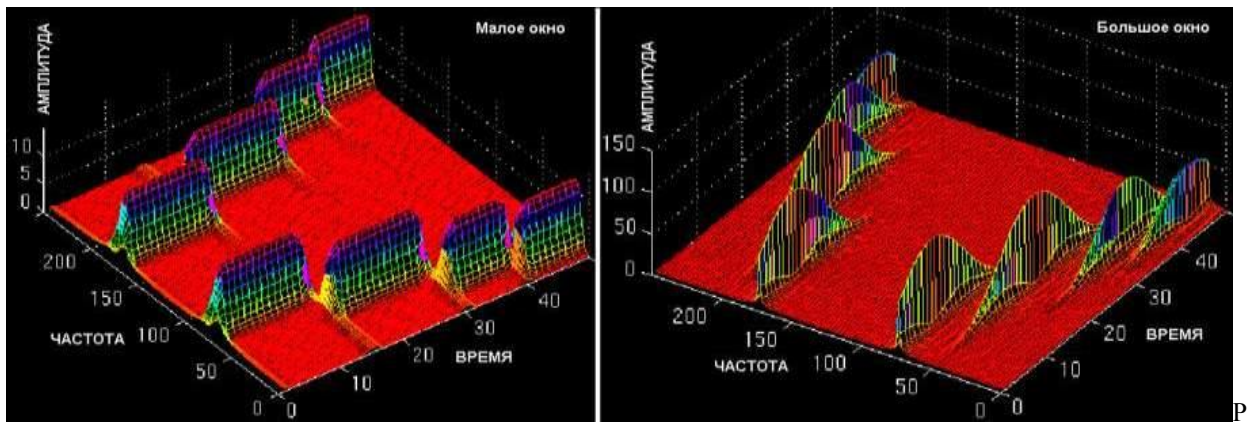


Рис. 23.1.3.

Как видно из приведенных примеров, оконное преобразование позволяет выделить информативные особенности сигнала и по времени, и по частоте. Разрешающая способность локализации по координатам и по частоте определяется принципом неопределенности Гейзенберга. В силу этого принципа невозможно получить произвольно точное частотно-временное представление сигнала. На рис. 23.1.4 приведен пример частотно-временного оконного преобразования сигнала, состоящего из 4-х непересекающихся интервалов, в каждом из которых сумма двух гармоник разной частоты. В качестве окна применена гауссова функция разной ширины. Узкое окно обеспечивает лучшее временное разрешение и четкую фиксацию границ интервалов, но широкие пики частот в пределах интервалов. Широкое окно напротив – четко отмечает частоты интервалов, но с перекрытием границ временных интервалов. При решении практических задач приходится выбирать окно для анализа всего сигнала, тогда как разные его участки могут требовать применения разных окон.



ис. 23.1.4.

На рис. 23.1.3 приведен пример частотно-временного оконного преобразования сигнала, состоящего из 4-х непересекающихся интервалов, в каждом из которых сумма двух гармоник разной частоты. В качестве окна применена гауссова функция разной ширины. Узкое окно обеспечивает лучшее временное разрешение и четкую фиксацию границ интервалов, но широкие пики частот в пределах интервалов. Широкое окно напротив – четко отмечает частоты интервалов, но с перекрытием границ временных интервалов. При решении практических задач приходится выбирать окно для анализа всего сигнала, тогда как разные его участки могут требовать применения разных окон. Если сигнал состоит из далеко отстоящих друг от друга частотных компонент, то можно пожертвовать спектральным разрешением в пользу временного, и наоборот.

**Функции оконного спектрального анализа** в Mathcad находятся в пакете Signal Processing. Они позволяют разбивать сигнал на поддиапазоны (с перекрытием или без перекрытия) и выполнять следующие операции:

- $cspectrum(x,n,r[,w])$  – расчет кросс-спектра сигнала  $x$ ;
- $pspectrum(x,n,r[,w])$  – расчет распределения спектральной мощности сигнала;
- $coherence(x,y,n,r[,w])$  – расчет когерентности сигналов  $x$  и  $y$ ;
- $snr(x,y,n,r[,w])$  – расчет отношения сигнал/шум для векторов  $x$  и  $y$ .

Здесь:  $x$  и  $y$  – вещественные или комплексные массивы данных (векторы),  $n$  – число поддиапазонов разбиения входного сигнала  $x$  (от 1 до  $N$  – размера массива),  $k$  – фактор перекрытия поддиапазонов (от 0 до 1),  $w$  - код окна (1- прямоугольное, 2- трапеция, 3- треугольное, 4- окно Хеннинга, 5- окно Хемминга, 6- окно Блекмана).

**Принцип вейвлет-преобразования.** Гармонические базисные функции преобразования Фурье предельно локализованы в частотной области (до импульсных функций Дирака при  $T \rightarrow \infty$ ) и не локализованы во временной (определены во всем временном интервале от  $-\infty$  до  $\infty$ ).

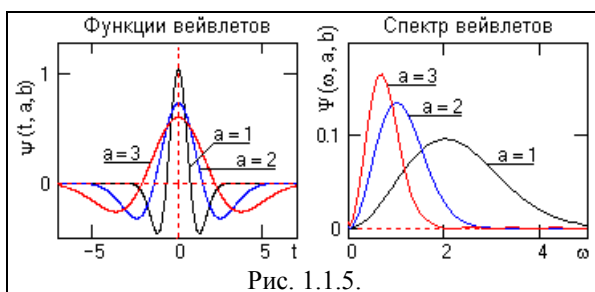


Рис. 1.1.5.

Их противоположностью являются импульсные базисные функции типа импульсов Кронекера, которые предельно локализованы во временной области и "размыты" по всему частотному диапазону. Вейвлеты по локализации в этих двух представлениях можно рассматривать как функции, занимающие промежуточное положение между гармоническими и импульсными функциями. Они должны быть локализованными как во временной, так и в частотной области представления. Однако при проектировании таких функций мы неминуемо столкнемся с принципом неопределенности, связывающим эффективные значения длительности функций и ширины их спектра. Чем точнее мы будем осуществлять локализацию временного положения функции, тем шире будет становиться

ее спектр, и наоборот, что наглядно видно на рис. 1.1.5.

Отличительной особенностью вейвлет-анализа является то, что в нем можно использовать семейства функций, реализующих различные варианты соотношения неопределенности. Соответственно, исследователь имеет возможность гибкого выбора между ними и применения тех вейвлетных функций, которые наиболее эффективно решают поставленные задачи.

Вейвлетный базис пространства  $L^2(\mathbb{R})$ ,  $\mathbb{R}(-\infty, \infty)$ , целесообразно конструировать из финитных функций, принадлежащих этому же пространству, которые должны стремиться к нулю на бесконечности. Чем быстрее эти функции стремятся к нулю, тем удобнее использовать их в качестве базиса преобразования при анализе реальных сигналов. Допустим, что такой функцией является  $\psi$  - функция  $\psi(t)$ , равная нулю за пределами некоторого конечного интервала и имеющая нулевое среднее значение по интервалу задания. Последнее необходимо для задания локализации спектра вейвлета в частотной области. На основе этой функции сконструируем базис в пространстве  $L^2(\mathbb{R})$  с помощью масштабных преобразований независимой переменной.

Функция изменения частотной независимой переменной в спектральном представлении сигналов отображается во временном представлении растяжением/сжатием сигнала. Для вейвлетного базиса это можно выполнить функцией типа  $\psi(t) \Rightarrow \psi(a^m t)$ ,  $a = \text{const}$ ,  $m = 0, 1, \dots, M$ , т.е. путем линейной операции растяжения/сжатия, обеспечивающей самоподобие функции на разных масштабах представления. Однако локальность функции  $\psi(t)$  на временной оси требует дополнительной независимой переменной последовательных сдвигов функции  $\psi(t)$  вдоль оси, типа  $\psi(t) \Rightarrow \psi(t+k)$ , для перекрытия всей числовой оси пространства  $\mathbb{R}(-\infty, \infty)$ . С учетом обеих условий одновременно структура базисной функции может быть принята следующей:

$$\psi(t) \Rightarrow \psi(a^{m/2} \psi(a^m t + k)). \quad (1.1.10)$$

Для упрощения дальнейших выкладок значения переменных  $m$  и  $k$  примем целочисленными. При приведении функции (1.1.10) к единичной норме, получаем:

$$\psi_{mk}(t) = a^{m/2} \psi(a^m t + k). \quad (1.1.11)$$

Если для семейства функций  $\psi_{mk}(t)$  выполняется условие ортогональности:

$$\langle \psi_{nk}(t), \psi_{lm}(t) \rangle = \int_{-\infty}^{\infty} \psi_{nk}(t) \cdot \psi_{lm}^*(t) dt = \delta_{nl} \cdot \delta_{km}, \quad (1.1.12)$$

то семейство  $\psi_{mk}(t)$  можно использовать в качестве ортонормированного базиса пространства  $L^2(\mathbb{R})$ . Произвольную функцию этого пространства можно разложить в ряд по базису  $\psi_{mk}(t)$ :

$$s(t) = \sum_{m,k=-\infty}^{\infty} S_{mk} \psi_{mk}(t), \quad (1.1.13)$$

где коэффициенты  $S_{mk}$  – проекции сигнала на новый ортогональный базис функций, как и в преобразовании Фурье, определяются скалярным произведением

$$S_{mk} = \langle s(t), \psi_{mk}(t) \rangle = \int_{-\infty}^{\infty} s(t) \psi_{mk}^*(t) dt, \quad (1.1.14)$$

при этом ряд равномерно сходиться:

$$\lim_{M,K \rightarrow \infty} \left\| s(t) - \sum_{m=-M}^M \sum_{k=-K}^K S_{mk} \psi_{mk}(t) \right\| = 0.$$

При выполнении этих условий базисная функция преобразования  $\psi(t)$  называется ортогональным вейвлетом.

Простейшим примером ортогональной системы функций такого типа являются функции Хаара. Базисная функция Хаара определяется соотношением



$$\psi(t) = \begin{cases} 1, & 0 < t < 1/2 \\ -1, & 1/2 < t < 1 \\ 0, & t < 0, t > 1. \end{cases} \quad (1.1.15)$$

Легко проверить, что при  $a = 2$ ,  $m = 0, 1, 2, \dots$ ,  $k = 0, 1, 2, \dots$  две любые функции, полученные с помощью этого базисного вейвлета путем масштабных преобразований и переносов, имеют единичную норму и ортогональны. На рис. 1.1.6 приведены примеры функций для первых трех значений  $m$  и  $b$  при различных их комбинациях, где ортогональность функций видна наглядно.

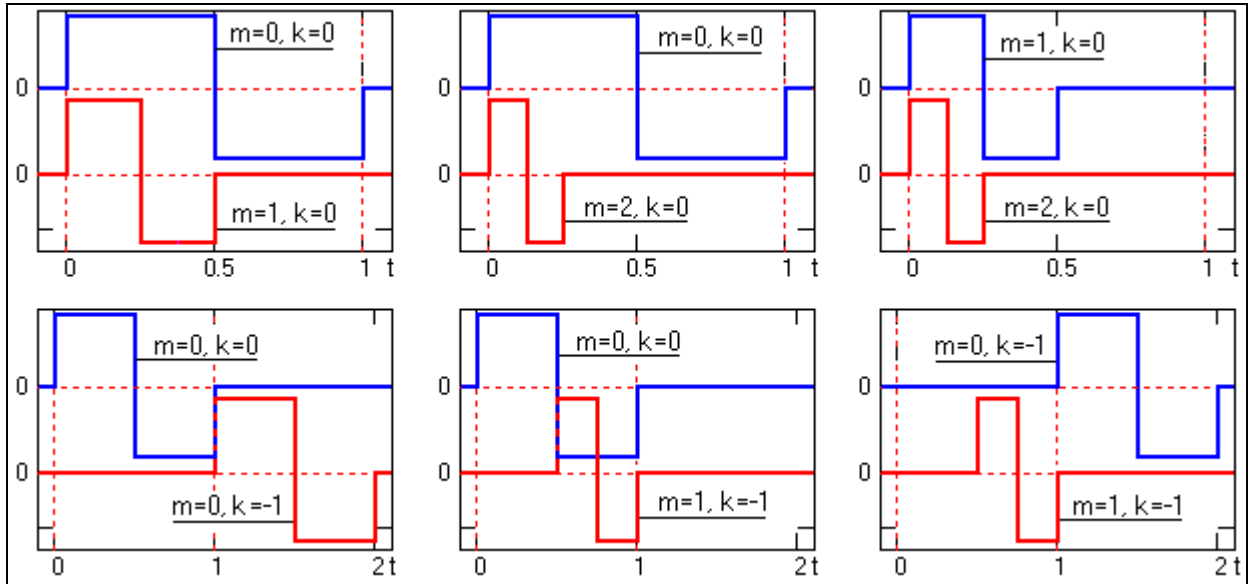


Рис. 1.1.6. Функции Хаара.

**Вейвлетный спектр**, в отличие от преобразования Фурье, является двумерным и определяет двумерную поверхность в пространстве переменных  $m$  и  $k$ . При графическом представлении параметр растяжения/сжатия спектра  $m$  откладывается по оси абсцисс, параметр локализации  $k$  по оси ординат – оси независимой переменной сигнала. Математику процесса вейвлетного разложения сигнала в упрощенной форме рассмотрим на примере разложения сигнала  $s(t)$  вейвлетом Хаара с тремя последовательными по масштабу  $m$  вейвлетными функциями с параметром  $a=2$ , при этом сам сигнал  $s(t)$  образуем суммированием этих же вейвлетных функций с одинаковой амплитудой с разным сдвигом от нуля, как это показано на рис. 1.1.7.

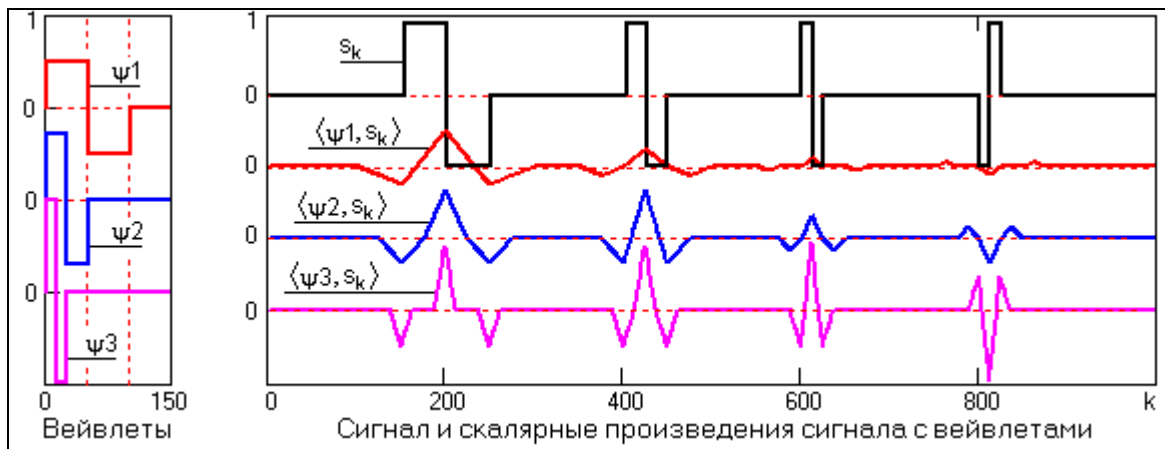


Рис. 1.1.7. Скалярные произведения сигнала с вейвлетами.

Для начального значения масштабного коэффициента сжатия  $m$  определяется функция вейвлета ( $\psi_1(t)$  на рис. 1.1.7), и вычисляется скалярное произведение сигнала с вейвлетом

$\langle \psi(t), s(t+k) \rangle$  с аргументом по сдвигу  $k$ . Для наглядности результаты вычисления скалярных произведений на рис. 1.1.7 построены по центрам вейвлетных функций (т.е. по аргументу  $k$  от нуля со сдвигом на половину длины вейвлетной функции). Как и следовало ожидать, максимальные значения скалярного произведения отмечаются там, где локализована эта же вейвлетная функция.

После построения первой масштабной строки разложения, меняется масштаб вейвлетной функции ( $\psi_2$  на рис. 1.1.7) и выполняется вычисление второй масштабной строки спектра, и т.д.

Как видно на рис. 1.1.7, чем точнее локальная особенность сигнала совпадает с соответствующей функцией вейвлета, тем эффективнее выделение этой особенности на соответствующей масштабной строке вейвлетного спектра. Можно видеть, что для сильно сжатого вейвлета Хаара характерной хорошо выделяемой локальной особенностью является скачок сигнала, причем выделяется не только скачок функции, но и направление скачка.

На рис. 1.1.8 приведен пример графического отображения вейвлетной поверхности реального физического процесса [4]. Вид поверхности определяет изменения во времени спектральных компонент различного масштаба и называется частотно-временным спектром. Поверхность изображается на рисунках, как правило, в виде изолиний или условными цветами. Для расширения диапазона масштабов может применяться логарифмическая шкала.

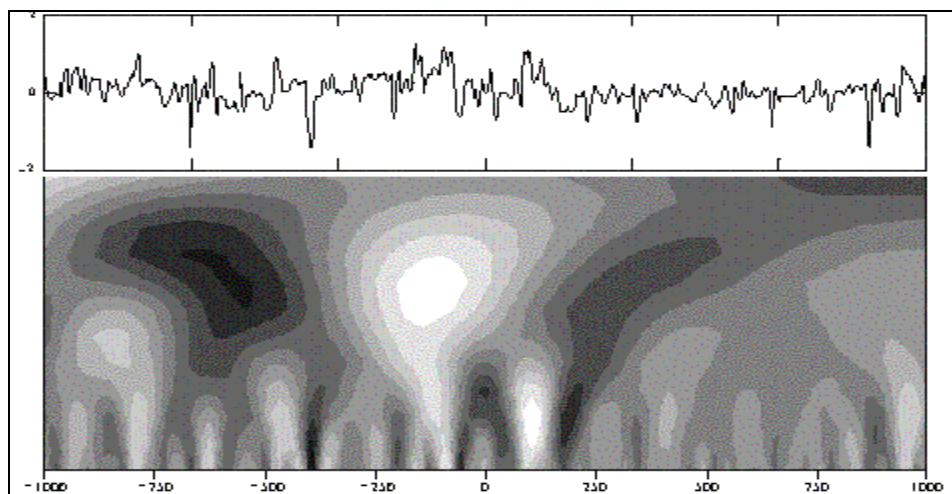


Рис. 1.1.8. Пример вейвлетного преобразования.

## 1.2. ОСНОВЫ ВЕЙВЛЕТ - ПРЕОБРАЗОВАНИЯ

В основе вейвлет-преобразований, в общем случае, лежит использование двух непрерывных, взаимозависимых и интегрируемых по независимой переменной функций:

- Вейвлет-функции  $\psi(t)$ , как psi-функции времени с нулевым значением интеграла и частотным фурье-образом  $\Psi(\omega)$ . Этой функцией, которую обычно и называют вейвлетом, выделяются локальные особенности сигнала. В качестве вейвлетов обычно выбирают функции, хорошо локализованные и во временной, и в частотной области. Пример временного и частотного образа функции приведен на рис. 1.2.1.
- Масштабирующей функции  $\phi(t)$ , как временной скейлинг-функции  $\phi$  с единичным значением интеграла, которой выполняется грубое приближение (аппроксимация) сигнала.

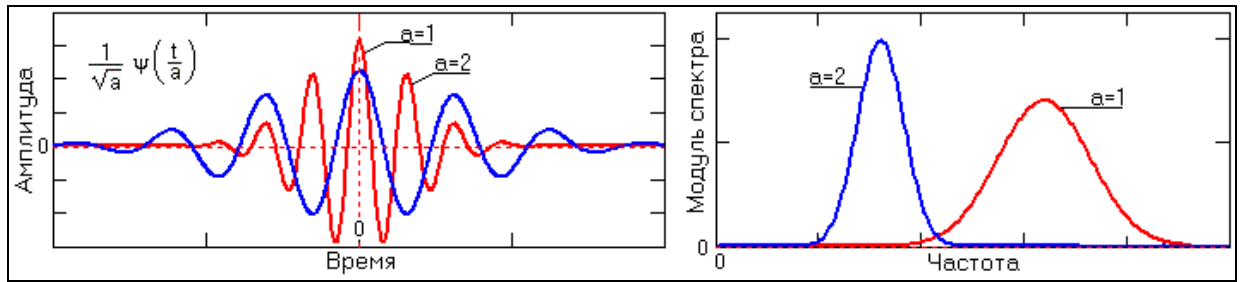


Рис. 1.2.1. Вейвлетные функции в двух масштабах.

Phi-функции присущи не всем, а, как правило, только ортогональным вейвлетам. Они необходимы для преобразования нецентрированных и достаточно протяженных сигналов при отдельном анализе низкочастотных и высокочастотных составляющих. Роль и использование phi-функции рассмотрим несколько позже.

**Непрерывное вейвлет-преобразование** (НВП, CWT- Continuous Wavelet Transform). Допустим, что мы имеем функции  $s(t)$  с конечной энергией в пространстве  $L^2(\mathbb{R})$ , определенные по всей действительной оси  $\mathbb{R}(-\infty, \infty)$ . Для финитных сигналов с конечной энергией средние значения сигналов должны стремиться к нулю на  $\pm\infty$ .

Непрерывным вейвлет-преобразованием (или вейвлетным образом) функции  $s(t) \in L^2(\mathbb{R})$  называют функцию двух переменных:

$$C(a,b) = \langle s(t), \psi(a,b,t) \rangle = \int_{-\infty}^{\infty} s(t) \psi(a,b,t) dt, \quad a, b \in \mathbb{R}, \quad a \neq 0. \quad (1.2.1)$$

где вейвлеты  $\psi(a,b,t) \equiv \psi_{ab}(t)$  – масштабированные и сдвинутые копии порождающего вейвлета  $\psi(t) \in L^2(\mathbb{R})$ , совокупность которых создает базис пространства  $L^2(\mathbb{R})$ .

Порождающими функциями могут быть самые различные функции с компактным носителем - ограниченные по времени и местоположению на временной оси, и имеющие спектральный образ, локализованный на частотной оси. Базис пространства  $L^2(\mathbb{R})$  целесообразно конструировать из одной порождающей функции, норма которой должна быть равна 1. Для перекрытия функцией вейвлета всей временной оси пространства используется операция сдвига (смещения по временной оси):  $\psi(b,t) = \psi(t-b)$ , где значение  $b$  для НВП является величиной непрерывной. Для перекрытия всего частотного диапазона пространства  $L^2(\mathbb{R})$  используется операция временного масштабирования вейвлета с непрерывным изменением независимой переменной:  $\psi(a,t) = |a|^{-1/2} \psi(t/a)$ . На рис. 1.2.1. видно, что если временной образ вейвлета будет расширяться (изменением значения параметра 'a'), то его "средняя частота" будет понижаться, а частотный образ (частотная локализация) перемещаться на более низкие частоты. Таким образом, путем сдвига по независимой переменной  $(t-b)$  вейвлет имеет возможность перемещаться по всей числовой оси произвольного сигнала, а путем изменения масштабной переменной 'a' (в фиксированной точке  $(t-b)$  оси) "просматривать" частотный спектр сигнала по определенному интервалу окрестностей этой точки.

С использованием этих операций вейвлетный базис функционального пространства образуется путем масштабных преобразований и сдвигов порождающего вейвлета  $\psi(t)$ :

$$\psi(a,b,t) = |a|^{-1/2} \psi[(t-b)/a], \quad a, b \in \mathbb{R}, \quad a \neq 0, \quad \psi(t) \in L^2(\mathbb{R}). \quad (1.2.2)$$

Нетрудно убедиться, что нормы вейвлетов  $\psi(a,b,t)$  равны норме  $\psi(t)$ , что обеспечивает нормировочный множитель  $|a|^{-1/2}$ . При нормировке к 1 порождающего вейвлета  $\psi(t)$  все семейство вейвлетов также будет нормированным. Если при этом выполняется требование ортогональности функций, то функции  $\psi(a,b,t)$  образуют ортонормированный базис пространства  $L^2(\mathbb{R})$ .

**Понятие масштаба ВП** имеет аналогию с масштабом географических карт. Большие значения масштаба соответствуют глобальному представлению сигнала, а низкие значения масштаба позволяют различить детали. В терминах частоты низкие частоты соответствуют глобальной информации о сигнале, а высокие частоты - детальной



информации и особенностям, которые имеют малую протяженность, т.е. масштаб вейвлета, как единица шкалы частотно-временного представления сигналов, обратен частоте. Масштабирование, как математическая операция, расширяет или сжимает сигнал. Большие значения масштабов соответствуют расширениям сигнала, а малые значения - сжатым версиям. В определении вейвлета коэффициент масштаба  $a$  стоит в знаменателе. Соответственно,  $a > 1$  расширяет сигнал,  $a < 1$  сжимает его.

**Процедура преобразования** стартует с масштаба  $a=1$  и продолжается при увеличивающихся значениях  $a$ , т.е. анализ начинается с высоких частот и проводится в сторону низких частот. Первое значение 'a' соответствует наиболее сжатому вейвлету. При увеличении значения 'a' вейвлет расширяется. Вейвлет помещается в начало сигнала ( $t=0$ ), перемножается с сигналом, интегрируется на интервале своего задания и нормализуется на  $1/\sqrt{a}$ . Результат вычисления  $C(a,b)$  помещается в точку ( $a=1, b=0$ ) масштабно-временного спектра преобразования. Сдвиг  $b$  может рассматриваться как время с момента  $t=0$ , при этом координатная ось  $b$  повторяет временную ось сигнала. Для полного включения в обработку всех точек входного сигнала требуется задание начальных и конечных условий преобразования (определенных значений входного сигнала при  $t < 0$  и  $t > t_{\max}$  на полуширину окна вейвлета). При одностороннем задании вейвлетов результат относится, как правило, к временному положению средней точки окна вейвлета.

Затем вейвлет масштаба  $a=1$  сдвигается вправо на значение  $b$  и процедура повторяется. Получаем значение, соответствующее  $t=b$  в строке  $a=1$  на частотно-временном плане. Процедура повторяется до тех пор, пока вейвлет не достигнет конца сигнала. Таким образом получаем строку точек на масштабно-временном плане для масштаба  $a=1$ .

Для вычисления следующей масштабной строки значение  $a$  увеличивается на некоторое значение. При НВП в аналитической форме  $\Delta b \rightarrow 0$  и  $\Delta a \rightarrow 0$ . При выполнении преобразования в компьютере выполняется увеличение обоих параметров с определенным шагом. Тем самым осуществляется дискретизация масштабно-временной плоскости.

Начальное значение масштабного коэффициента может быть и меньше 1. Для детализации самых высоких частот сигнала минимальных размер окна вейвлета не должен превышать периода самой высокочастотной гармоники. Если в сигнале присутствуют спектральные компоненты, соответствующие текущему значению  $a$ , то интеграл произведения вейвлета с сигналом в интервале, где эта спектральная компонента присутствует, дает относительно большое значение. В противном случае - произведение мало или равно нулю, т.к. среднее значение вейвлетной функции равно нулю. С увеличением масштаба (ширины окна) вейвлета преобразование выделяет все более низкие частоты.

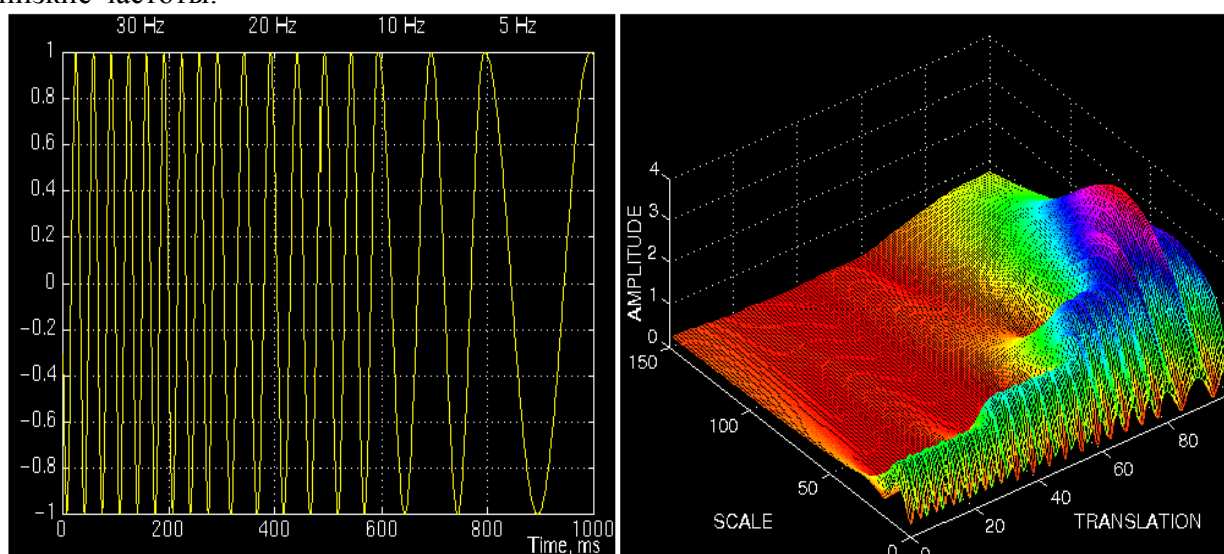


Рис. 1.2.2.

На рис. 1.2.2 приведен пример модельного сигнала и спектра его непрерывного вейвлет-преобразования.

Значения параметров 'a' и 'b' в (1.2.2) являются непрерывными, и множество базисных функций является избыточным. Сигналу, определенному на  $\mathbb{R}$ , соответствует вейвлетный спектр  $\mathbb{R} \times \mathbb{R}$ . Отсюда следует, что вейвлетный спектр НПВ имеет огромную избыточность.

**Обратное преобразование.** Так как форма базисных функций  $\psi(a,b,t)$  зафиксирована, то вся информация о сигнале в (1.2.1) переносится на значения функции  $C(a,b)$ . Точность обратного интегрального вейвлет-преобразования зависит от выбора базисного вейвлета и способа построения базиса, т.е. от значений базисных параметров  $a, b$ . Строго теоретически вейвлет может считаться базисной функцией  $L^2(\mathbb{R})$  только в случае его ортонормированности. Для практических целей непрерывного преобразования часто бывает вполне достаточно устойчивость и "приблизительность" ортогональности системы разложения функций. Под устойчивостью понимается достаточно точная реконструкция произвольных сигналов. Для ортонормированных вейвлетов обратное вейвлет-преобразование записывается с помощью того же базиса, что и прямое:

$$s(t) = (1/C_\psi) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1/a^2) C(a,b) \psi(a,b,t) da db. \quad (1.2.3)$$

где  $C_\psi$  - нормализующий коэффициент:

$$C_\psi = \int_{-\infty}^{\infty} (|\Psi(\omega)|^2 / \omega) d\omega < \infty. \quad (1.2.4)$$

Условие конечности  $C_\psi$  ограничивает класс функций, которые можно использовать в качестве вейвлетов. В частности, при  $\omega=0$ , для обеспечения сходимости интеграла (1.2.4) в нуле, значение  $\Psi(\omega)$  должно быть равно нулю. Это обеспечивает условие компактности фурье-образа вейвлета с локализацией вокруг некоторой частоты  $\omega_0$  – средней частоты вейвлетной функции. Следовательно, функция  $\psi(t)$  должна иметь нулевое среднее значение по области его определения (интеграл функции по аргументу должен быть нулевым):

$$\int_{-\infty}^{\infty} \psi(t) dt = 0.$$

Однако это означает, что не для всех сигналов возможна их точная реконструкция вейвлетом  $\psi(t)$ , т.к. при нулевом первом моменте вейвлета коэффициент передачи постоянной составляющей сигнала в преобразовании (1.2.3) равен нулю. Условия точной реконструкции сигналов будут рассмотрены при описании кратномасштабного анализа. Кроме того, даже при выполнении условия (1.2.4) далеко не все типы вейвлетов могут гарантировать реконструкцию сигналов, как таковую. Однако и такие вейвлеты могут быть полезны для анализа особенностей сигналов, как дополнительного метода к другим методам анализа и обработки данных. В общем случае, при отсутствии строгой ортогональности вейвлетной функции (1.2.1), для обратного преобразования применяется выражение:

$$s(t) = (1/C_\psi) \int_{\mathbb{R}} (1/a^2) C(a,b) \psi^\#(a,b,t) da db, \quad (1.2.3')$$

где индексом  $\psi^\#(a,b,t)$  обозначен ортогональный "двойник" базиса  $\psi(a,b,t)$ , о котором будет изложено ниже.

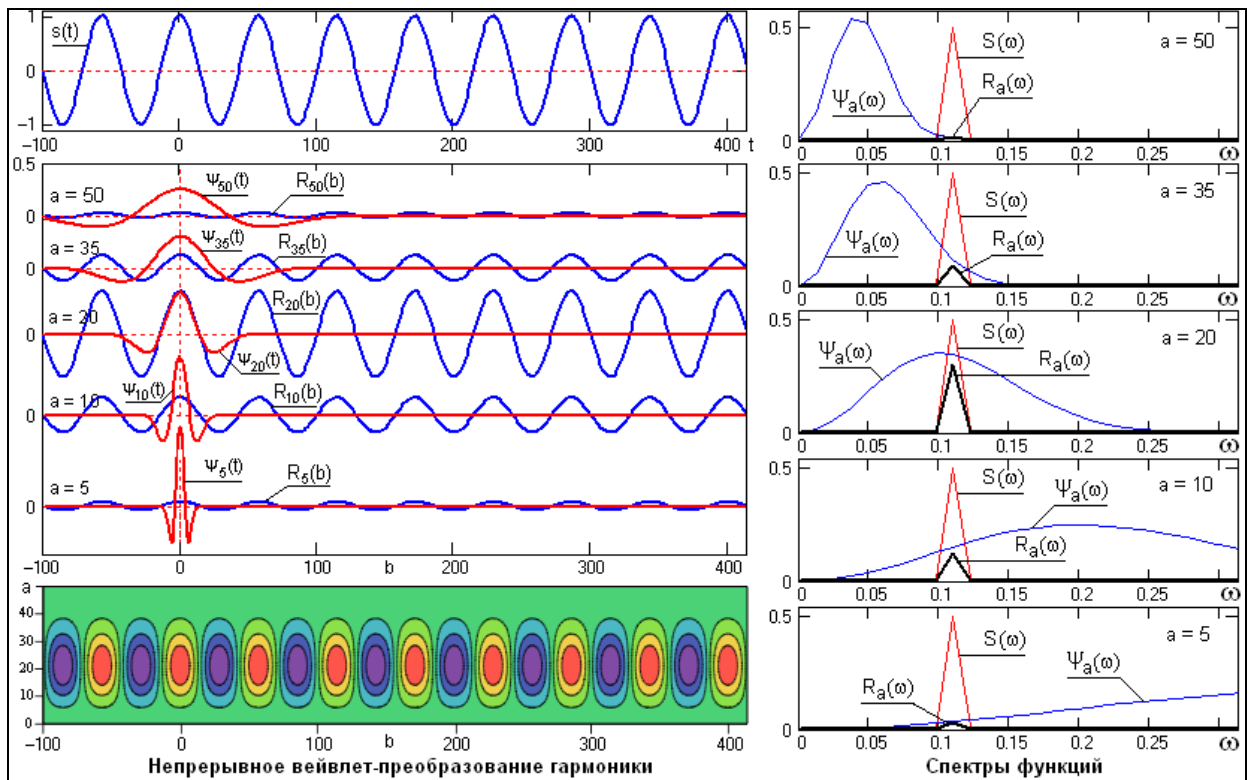


Рис. 1.2.3.

Таким образом, непрерывное вейвлет-преобразование представляет собой разложение сигнала по всем возможным сдвигам и сжатиям/растяжениям некоторой локализованной финитной функции - вейвлета. При этом переменная 'a' определяет масштаб вейвлета и эквивалентна частоте в преобразованиях Фурье, а переменная 'b' – сдвиг вейвлета по сигналу от начальной точки в области его определения, шкала которого повторяет временную шкалу анализируемого сигнала. Вейвлетный анализ является частотно-пространственным анализом сигналов.

В качестве примера рассмотрим вейвлет-преобразование чистого гармонического сигнала  $s(t)$ , приведенного на рис. 1.2.3. На этом же рисунке ниже приведены вейвлеты  $\psi_a(t)$  симметричного типа разных масштабов.

Скалярное произведение (1.2.1) "просмотра" сигнала вейвлетом определенного масштаба 'a' может быть записано в следующей форме:

$$C_a(b) = \langle s(t), \psi_a(t+b) \rangle = \int_{-\infty}^{\infty} s(t) \psi_a(t+b) dt. \quad (1.2.5)$$

Но выражение (1.2.5) эквивалентно взаимной корреляционной функции  $R_a(b)$  сигналов  $s(t)$  и  $\psi_a(t)$ . Если сигнал  $s(t)$  представляет собой гармонику, а второй сигнал симметричен, задан на компактном носителе и имеет нулевое среднее значение, то, как известно, форма взаимной корреляционной функции таких сигналов также является центрированным гармоническим сигналом. В частотной области скалярное произведение двух функций отображается произведением Фурье-образов этих функций, которые приведены на рисунке в правом столбце спектров. Масштабы спектров  $\Psi_a(\omega)$  и  $R_a(\omega)$  для наглядности сопоставления нормированы к спектру  $s(t)$ . Максимальная амплитуда гармоники  $R_a(b)$  будет наблюдаться при совпадении средней частоты локализации вейвлета  $\psi_a(t)$  определенного масштаба 'a' в частотной области с частотой сигнала  $s(t)$ , что и можно видеть на рис. 1.2.3 для функции  $R_a(b)$  при масштабе вейвлета  $a=20$ . Результирующий вейвлетный спектр непрерывного вейвлет-преобразования гармоники приведен на левом нижнем графике и показывает точное положение на временной оси 'b' максимумов и минимумов гармонического сигнала.

**Дискретное вейвлет-преобразование.** В принципе, при обработке данных на ПК может выполняться дискретизированная версия непрерывного вейвлет-преобразования с заданием дискретных значений параметров  $(a, b)$  вейвлетов с произвольным шагом  $\Delta a$  и  $\Delta b$ . В результате получается избыточное количество коэффициентов, намного превосходящее число отсчетов исходного сигнала, которое не требуется для реконструкции сигналов.

Дискретное вейвлет-преобразование (ДВП) обеспечивает достаточно информации, как для анализа сигнала, так и для его синтеза, являясь вместе с тем экономным по числу операций и по требуемой памяти. ДВП оперирует с дискретными значениями параметров  $a$  и  $b$ , которые задаются, как правило, в виде степенных функций:

$$a = a_0^{-m}, \quad b = k \cdot a_0^{-m}, \quad a_0 > 1, \quad m, k \in I,$$

где  $I$  – пространство целых чисел  $\{-\infty, \infty\}$ ,  $m$  – параметр масштаба,  $k$  – параметр сдвига. Базис пространства  $L^2(\mathbb{R})$  в дискретном представлении:

$$\psi_{mk}(t) = |a_0|^{m/2} \psi(a_0^m t - k), \quad m, k \in I, \quad \psi(t) \in L^2(\mathbb{R}). \quad (1.2.6)$$

Вейвлет-коэффициенты прямого преобразования:

$$C_{mk} = \int_{-\infty}^{\infty} s(t) \psi_{mk}(t) dt. \quad (1.2.7)$$

Значение 'a' может быть произвольным, но обычно принимается равным 2, при этом преобразование называется *диадным вейвлет-преобразованием*. Для диадного преобразования разработан быстрый алгоритм вычислений, аналогичный быстрому преобразованию Фурье, что предопределило его широкое использование при анализе массивов цифровых данных.

Обратное дискретное преобразование для непрерывных сигналов при нормированном ортогональном вейвлетном базисе пространства:

$$s(t) = \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_{mk} \psi_{mk}(t). \quad (1.2.8)$$

Число использованных вейвлетов по масштабному коэффициенту  $m$  задает уровень *декомпозиции* сигнала, при этом за нулевой уровень ( $m = 0$ ) обычно принимается уровень максимального временного разрешения сигнала, т.е. сам сигнал, а последующие уровни ( $m < 0$ ) образуют ниспадающее *вейвлет-дерево*. В программном обеспечении вычислений для исключения использования отрицательной нумерации по  $m$  знак 'минус' обычно переносится непосредственно в (1.2.6), т.е. используется следующее представление базисных функций:

$$\psi_{mk}(t) = |a_0|^{-m/2} \psi(a_0^{-m} t - k), \quad m, k \in I, \quad \psi(t) \in L^2(\mathbb{R}). \quad (1.2.6')$$

Устойчивость дискретного базиса определяется следующим образом.

Функция  $\psi(t) \in L^2(\mathbb{R})$  называется R-функцией, если базис на ее основе по (1.2.6) является базисом Рисса (Riesz). Для базиса Рисса существуют значения  $A$  и  $B$ ,  $0 < A \leq B < \infty$ , для которых выполняется соотношение

$$A \|C_{mk}\|^2 \leq \left\| \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_{mk} \psi_{mk}(t) \right\|^2 \leq B \|C_{mk}\|^2,$$

если энергия ряда  $C_{mk}$  конечна. При этом для любой R-функции существует базис  $\psi^{\#}_{mk}(t)$ , который ортогонален базису  $\psi_{mk}(t)$ . Его называют ортогональным "двойником" базиса  $\psi_{mk}(t)$ , таким, что

$$\langle \psi_{mk}(t), \psi^{\#}_{nl}(t) \rangle = \delta_{mn} \cdot \delta_{kl}.$$

Если  $A = B = 1$  и  $a_0 = 2$ , то семейство базисных функций  $\{\psi_{mk}(t)\}$  является ортонормированным базисом и возможно полное восстановление исходного сигнала, при этом  $\psi_{mk}(t) \equiv \psi^{\#}_{mk}(t)$  и для реконструкции сигналов используется формула (1.2.8). Если  $\psi(t)$  не ортогональный вейвлет, но имеет "двойника", то на базе "двойника" вычисляется семейство  $\psi^{\#}_{mk}(t)$ , которое и используется при обратном преобразовании вместо  $\psi_{mk}(t)$ ,

при этом точное восстановление исходного сигнала не гарантировано, но оно будет близко к нему в среднеквадратическом смысле.

Как и для непрерывного вейвлет-преобразования, обратное дискретное преобразование (1.2.8) не может выполнить восстановление нецентрированных сигналов в силу нулевого первого момента вейвлетных функций и, соответственно, центрирования значения вейвлет-коэффициентов  $C_{mk}$  при прямом вейвлет-преобразовании. Поэтому при обработке числовых массивов данных дискретные вейвлеты используются, как правило, в паре со связанными с ними дискретными скейлинг-функциями. Скейлинг-функции имеют с вейвлетами общую область задания и определенное соотношение между значениями, но первый момент скейлинг-функций по области определения равен 1. Если вейвлеты рассматривать, как аналоги полосовых фильтров сигнала, в основном, высокочастотных при выделении локальных особенностей в сигнале, то скейлинг-функции вейвлетов представляет собой аналоги низкочастотных фильтров, которыми из сигнала выделяются в отдельный массив составляющие, не прошедшие вейвлетную фильтрацию. Так, например, порождающая скейлинг-функция вейвлета Хаара (1.1.15) задается следующим выражением:

$$\varphi(t) = \begin{cases} 1, & 0 < t < 1 \\ 0, & t < 0, t > 1. \end{cases}$$

При обозначении скейлинг-функций индексом  $\varphi_{mk}(t)$  аналитика скейлинг-функций повторяет выражения (1.2.6-1.2.7) и образует дополнительный базис пространства  $L^2(\mathbb{R})$ . Сумма вейвлет-коэффициентов и скейлинг-коэффициентов разложения сигналов соответственно дает возможность выполнять точную реконструкцию сигналов, при этом вместо (1.2.8) используется следующее выражение обратного вейвлет-преобразования:

$$s(t) = \sum_{k=-\infty}^{\infty} C_{ak} \varphi_k(t) + \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_{d_{mk}} \psi_{mk}(t), \quad (1.2.9)$$

где  $C_{ak}$  – скейлинг-коэффициенты, которые обычно называют коэффициентами аппроксимации сигнала,  $C_{d_{mk}}$  – вейвлет-коэффициенты или коэффициенты детализации. Более подробно использование скейлинг-функций будет рассмотрено в теме вейвлетного кратномасштабного анализа.

**Частотно-временная локализация вейвлет-анализа.** Реальные сигналы, как правило, конечны и принадлежат пространству  $L^2(\mathbb{R})$ . Частотный спектр сигналов обратно пропорционален их длительности. Соответственно, достаточно точный низкочастотный анализ сигнала должен производиться на больших интервалах его задания, а высокочастотный – на малых. Если частотный состав сигнала претерпевает существенные изменения на интервале его задания, то преобразование Фурье дает только усредненные данные частотного состава сигнала с постоянным частотным разрешением. Определенная частотно-временная локализация анализа создается применением оконного преобразования Фурье, что дает семейства частотных спектров, локализованных во времени, но в пределах постоянной ширины окна оконной функции, а, следовательно, также с постоянным значением и частотного, и временного разрешения. В отличие от оконного преобразования Фурье, вейвлет-преобразование, при аналогичных дискретных значениях сдвигов  $b$ , дает семейства спектров масштабных коэффициентов  $a$  сжатия-растяжения

$$C(a,b) = \int_{-\infty}^{\infty} s(t) |a|^{-1/2} \psi_0[(t-b)/a] dt.$$

Если считать, что каждый вейвлет имеет определенную "ширину" своего временного окна, которому соответствует определенная "средняя" частота спектрального образа вейвлета, обратная его масштабному коэффициенту  $a$ , то семейства масштабных коэффициентов вейвлет-преобразования можно считать аналогичными семействам частотных спектров оконного преобразования Фурье, но с одним принципиальным

отличием. Масштабные коэффициенты изменяют "ширину" вейвлетов и, соответственно, "среднюю" частоту их фурье-образов, а, следовательно, каждой частоте соответствует своя длительность временного окна анализа, и наоборот. Так малые значения параметра  $a$ , характеризующие быстрые составляющие в сигналах, соответствуют высоким частотам, а большие значения – низким частотам. За счёт изменения масштаба вейвлеты способны выявлять различия на разных частотах, а за счёт сдвига (параметр  $b$ ) проанализировать свойства сигнала в разных точках на всём исследуемом временном интервале. Многомерное временное окно вейвлет-преобразования адаптировано для оптимального выявления и низкочастотных, и высокочастотных характеристики сигналов.

Для произвольной оконной функции  $z(t) \in L^2(\mathbb{R})$  ее центр и радиус определяются формулами:

$$t_0 = \frac{1}{\|z(t)\|^2} \int_{-\infty}^{\infty} t |z(t)|^2 dt,$$

$$\Delta_z = \frac{1}{\|z(t)\|} \sqrt{\int_{-\infty}^{\infty} (t - t_0)^2 |z(t)|^2 dt}$$

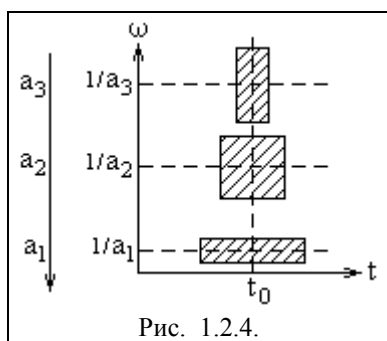


Рис. 1.2.4.

Если по этим функциям определить центры и радиусы вейвлетов и их фурье-образов, то временная локализация происходит с центрами окон  $b+at_0$  шириной  $\text{win}_t = 2a\Delta_{\psi(t)}$ , а частотная – с центрами  $\omega_0/a$ , и с шириной окна  $\text{win}_\omega = 2\Delta_{\psi(\omega)}/a$ . При этом значение отношения центральной частоты к ширине окна не зависит от местоположения центральной частоты. Частотно-временное окно  $\text{win}_t \cdot \text{win}_\omega = 4\Delta_{\psi(t)}\Delta_{\psi(\omega)}$ , сужается при высокой центральной частоте, и расширяется при низкой.

Схематическое изображение частотно-временных окон преобразования приведено на рис. 1.2.4. Таким образом, на высоких частотах лучше разрешение по времени, а на низких - по частоте. Для высокочастотной компоненты сигнала мы можем точнее указать ее временную позицию, а для низкочастотной - ее значение частоты.

Изменение частотно-временного окна вейвлета определяет угол влияния значений функции в произвольных точках  $t_i$  на значения коэффициентов  $C(a,b)$ . И наоборот, угол влияния из точки  $C(a_i,b_i)$  на ось  $t$  определяет интервал значений функции, которые принимают участие в вычислении данного коэффициента  $C(a_i,b_i)$  – область достоверности. Схематически это показано на рис. 1.2.5.

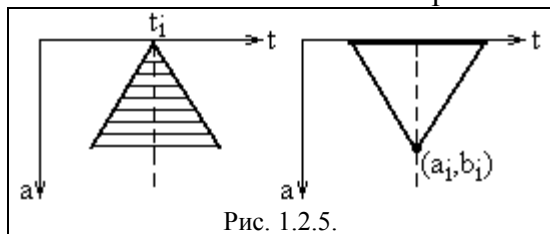


Рис. 1.2.5.

По углу влияния наглядно видно, что высокочастотная (мелкомасштабная) информация вычисляется на основе малых интервалов сигналов, а низкочастотная – на основе больших. Поскольку анализируемые сигналы всегда конечны, то при вычислении коэффициентов на границах задания сигнала область достоверности выходит за пределы сигнала, и для уменьшения погрешности вычислений сигнал дополняется заданием начальных и конечных условий.

**Образное представление преобразования.** Представим себе длинный и узкий стеклянный ларь, произвольно заполненный шарами трех разных диаметров: 5, 10 и 15 см. Взглянем на ларь сбоку, и линию высоты насыпки будем считать значением сигнала в зависимости от расстояния от одного из торцов ларя (условно – нулевого).

Возьмем первый "вейвлет" – идеальное дифференциальное сито с диаметром отверстий  $d=5$  см, через которое проходят только пятисантиметровые шары (аналог значения  $a_0$ ). Передвигаясь вдоль ларя, "просеем" через это сито шары в ларе, не перемешивая их по расстоянию от нулевого торца ларя и размещая отсеиваемые шары в

таким же ларе, сохраняя расстояние от начала ларя. Сменим масштаб "вейвлета" и повторим эту операцию ситом с диаметром отверстий 10, а затем 15 см. Если все три ларя расположить рядом, мы получим двумерную "поверхность" насыпки отсеянных шаров, которая наглядно покажет распределение шаров в ларе и по размерам, и по их концентрации в различных участках ларя.

Данная модель разложения является довольно грубой, но интуитивно понятно, что обратная сборка шаров в ларь с сохранением их местоположения с определенной точностью восстановит высоту насыпки. Замените шары короткими фрагментами электронных сигналов произвольной, но одной и той формы в пределах диаметра шаров, например такими, как  $\psi(t)$  на рис.1.2.1, сложите все значения сигналов по текущим значениям  $t$ , и Вы получите сложный суммарный сигнал. Используя прямое вейвлет-преобразование с вейвлетами этих же составляющих, Вы можете разложить суммарный сигнал (и любой другой произвольный сигнал) на составляющие в масштабнo-временной плоскости. Замените масштабную ось ширины вейвлетов на обратную ей частотную ось, и Вы представите результаты в частотно-временной плоскости. Заметим только, что точность, представительность и информативность результатов анализа во многом будут зависеть как от формы и особенностей анализируемого сигнала, так и от формы выбранных вами вейвлетов и параметров масштабирования и сдвига. Это определяется тем, что дифференциальное сито в примере с шарами – идеальная операция разделения, в то время как при вейвлет-преобразовании "идентификация" составляющих выполняется по скалярному произведению сигнала и функции вейвлета. Скалярное произведение в принципе не может давать однозначного ответа типа "да-нет", а только "наносит" на масштабнo-временную плоскость определенные значения величины скалярного произведения. С одной стороны, выбор типа вейвлета вносит определенную субъективность исследователя в методику исследования сигналов, но, с другой стороны, дает исследователю новые возможности и свободу в поиске наиболее эффективных и оптимальных методов обработки сигналов и извлечения из них необходимой информации.

#### **Достоинства и недостатки вейвлетных преобразований.**

- Вейвлетные преобразования обладают всеми достоинствами преобразований Фурье.
- Вейвлетные базисы могут быть хорошо локализованными как по частоте, так и по времени. При выделении в сигналах хорошо локализованных разномасштабных процессов можно рассматривать только те масштабные уровни разложения, которые представляют интерес.
- Вейвлетные базисы, в отличие от преобразования Фурье, имеют много разнообразных базовых функций, свойства которых ориентированы на решение различных задач. Базисные вейвлеты могут реализоваться функциями различной гладкости.
- Недостатком вейвлетных преобразований является их относительная сложность.

**Практическое использование** вейвлет-преобразований связано, в основном, с дискретными вейвлетами как в силу повсеместного использования цифровых методов обработки данных, так и в силу ряда различий дискретного и непрерывного вейвлет-преобразований.

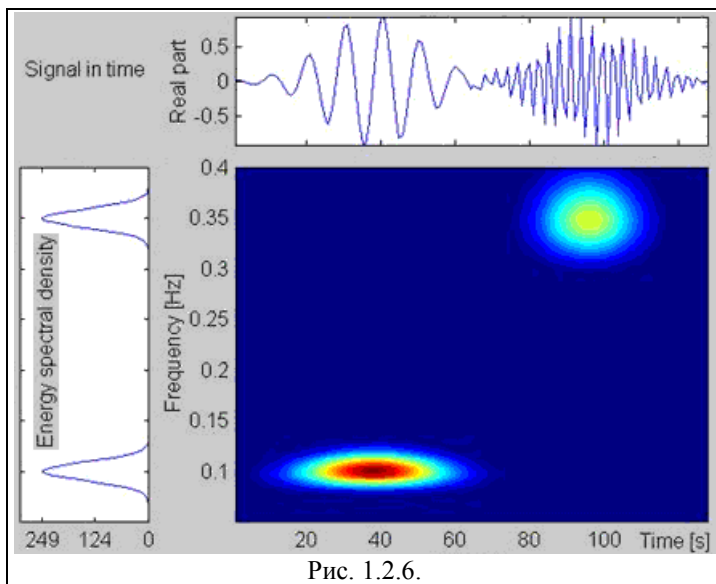


Рис. 1.2.6.

Непрерывные вейвлеты дают несколько более наглядное представление результатов анализа в виде поверхностей вейвлет-коэффициентов по непрерывным переменным. На рис. 1.2.6 анализируемый сигнал состоит из двух модулированных гауссианов. Преобразование вейвлетом Морлета четко показывает их пространственную и частотную локализацию, в то время как спектр Фурье дает только частотную локализацию. Однако базисы на основе непрерывных вейвлетов, как правило, не являются строго ортонормированными, поскольку элементы базиса бесконечно дифференцируемы и экспоненциально

спадают на бесконечности. У дискретных вейвлетов эти проблемы легко снимаются, что обеспечивает более точную реконструкцию сигналов.

Выбор конкретного вида и типа вейвлетов во многом зависит от анализируемых сигналов и задач анализа, при этом немалую роль играет интуиция и опыт исследователя. Для получения оптимальных алгоритмов преобразования разработаны определенные критерии, но их еще нельзя считать окончательными, т.к. они являются внутренними по отношению к самим алгоритмам преобразования и, как правило, не учитывают внешних критериев, связанных с сигналами и целями их преобразований. Отсюда следует, что при практическом использовании вейвлетов необходимо уделять достаточное внимание проверке их работоспособности и эффективности для поставленных целей по сравнению с известными методами обработки и анализа.

## Тема 20. СВОЙСТВА ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

Аналитика вейвлетных преобразований сигналов определяются математической базой разложения сигналов, которая аналогична преобразованиям Фурье. Основной отличительной особенностью вейвлет-преобразований является новый базис разложения сигналов - вейвлетные функции. Свойства вейвлетов принципиально важны как для самой возможности разложения сигналов по единичным вейвлетным функциям, так и для целенаправленных действий над вейвлетными спектрами сигналов, в том числе с последующей реконструкцией сигналов по обработанным вейвлетным спектрам.

Вейвлеты могут быть ортогональными, полуортогональными, биортогональными. Вейвлетные функции могут быть симметричными, асимметричными и несимметричными, с компактной областью определения и не имеющие таковой, а также иметь различную степень гладкости. Некоторые функции имеют аналитическое выражение, другие – быстрый алгоритм вычисления вейвлет-преобразования. Для практики желательно было бы иметь ортогональные симметричные и асимметричные вейвлеты, но таких идеальных вейвлетов не существует. Наибольшее применение находят биортогональные вейвлеты.

### 2.1. БАЗИСНЫЕ ФУНКЦИИ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

Базисными функциями вейвлет-преобразований могут быть самые различные функции с компактным носителем - модулированные импульсами синусоиды, функции со скачками уровня и т.п. Они обеспечивает хорошее отображение и анализ сигналов с локальными особенностями, в том числе со скачками, разрывами и перепадами значений с большой крутизной.

Следует различать вейвлеты по целевым задачам вейвлетных преобразований с



позиций декомпозиции – реконструкции сигналов. Было бы желательно иметь такое вейвлет-преобразование сигналов, которое обеспечивало бы полную информационную эквивалентность вейвлетного спектра сигналов временному (динамическому, координатному) представлению, и, соответственно, однозначность как декомпозиции сигналов, так и их реконструкции из вейвлетных спектров. Однако это возможно только при использовании ортогональных и биортогональных вейвлетов. Этим вейвлетам и будет уделено основное внимание. Для качественного анализа сигналов и локальных особенностей в сигналах может применяться более обширная номенклатура вейвлетных функций, которые хотя и не обеспечивают реконструкцию сигналов, но позволяют оценить информационное содержание сигналов и динамику изменения этой информации.

**Определение вейвлета.** К вейвлетам относятся локализованные функции, которые конструируются из одного материнского вейвлета  $\psi(t)$  (или по любой другой независимой переменной) путем операций сдвига по аргументу ( $b$ ) и масштабного изменения ( $a$ ):

$$\psi_{ab}(t) = (1/\sqrt{|a|}) \psi((t-b)/a), \quad (a, b) \in \mathbb{R}, \quad \psi(t) \in L^2(\mathbb{R}).$$

где множитель  $(1/\sqrt{|a|})$  обеспечивает независимость нормы функций от масштабного числа 'a'.

Непрерывное вейвлет-преобразование сигнала  $s(t) \in L^2(\mathbb{R})$ , которое применяется для качественного частотно-временного анализа, по смыслу соответствует преобразованию Фурье с заменой гармонического базиса  $\exp(-j\omega t)$  на вейвлетный  $\psi((t-b)/a)$ :

$$C(a, b) = \langle s(t), \psi_{ab}(t) \rangle = (1/\sqrt{|a|}) \int_{-\infty}^{\infty} s(t) \psi((t-b)/a) dt, \quad (a, b) \in \mathbb{R}, \quad a \neq 0.$$

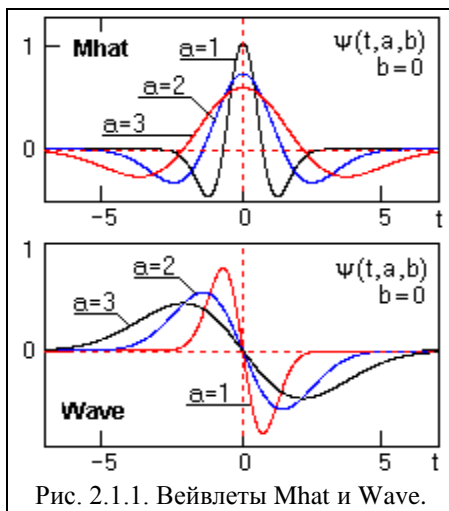


Рис. 2.1.1. Вейвлеты Mhat и Wave.

Вейвлетный масштабно-временной спектр  $C(a,b)$  в отличие от фурье-спектра является функцией двух аргументов: масштаба вейвлета 'a' (в единицах, обратных частоте), и временного смещения вейвлета по сигналу 'b' (в единицах времени), при этом параметры 'a' и 'b' могут принимать любые значения в пределах областей их определения.

На рис. 2.1.1 приведены примеры простейших неортогональных вейвлетов четного (Mhat) и нечетного (Wave) типов.

Для количественных методов анализа (декомпозиция сигналов с возможностью последующей линейной реконструкции сигналов из обработанных вейвлет-спектров) в качестве вейвлетных базисов можно

использовать любые локализованные функции  $\psi(t)$ , если для них существуют функции-двойники  $\psi^\#(t)$ , такие, что семейства  $\{\psi_{ab}(t)\}$  и  $\{\psi^\#_{ab}(t)\}$  могут образовывать парные базисы функционального пространства  $L^2(\mathbb{R})$ . Вейвлеты, определенные таким образом, позволяют представить любую произвольную функцию в пространстве  $L^2(\mathbb{R})$  в виде ряда:

$$s(t) = \sum_{a,b} C(a,b) \psi^\#_{ab}(t), \quad (a, b) \in I,$$

где коэффициенты  $C(a,b)$  – проекции сигнала на вейвлетный базис пространства, которые определяются скалярным произведением

$$C(a,b) = \langle s(t), \psi_{ab}(t) \rangle = \int_{-\infty}^{\infty} s(t) \psi_{ab}(t) dt.$$

Если вейвлет  $\psi(t)$  обладает свойством ортогональности, то  $\psi^\#(t) \equiv \psi(t)$  и вейвлетный базис ортогонален. Вейвлет может быть неортогональным, однако если он имеет двойника, и пара  $(\psi(t), \psi^\#(t))$  дает возможность сформировать семейства  $\{\psi_{mk}(t)\}$  и  $\{\psi^\#_{zp}(t)\}$ , удовлетворяющие условию биортогональности на целых числах I:

$$\langle \psi_{mk}(t), \psi_{zp}^{\#}(t) \rangle = \delta_{mz} \cdot \delta_{kp}, \quad m, k, z, p \in I,$$

то возможно разложение сигналов на вейвлетные ряды с обратной формулой реконструкции.

**Свойства вейвлета.**

- **Локализация.** Вейвлет должен быть непрерывным, интегрируемым, иметь компактный носитель и быть локализованным как во времени (в пространстве), так и по частоте. Если вейвлет в пространстве сужается, то его "средняя" частота повышается, спектр вейвлета перемещается в область более высоких частот и расширяется. Этот процесс должен быть линейным – сужение вейвлета вдвое должно повышать его "среднюю" частоту и ширину спектра также вдвое.

Вейвлетную функцию можно считать хорошо локализованной при выполнении условий:

$$\psi(t) \leq C/(1+|t|)^{1+\varepsilon}, \quad \Psi(f) \leq C/(1+|f|)^{1+\varepsilon}, \quad C = \text{const}, \quad \text{при } \varepsilon > 0.$$

- **Нулевое среднее значение,** т.е. выполнение условия для нулевого момента:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0,$$

что обеспечивает выделение локальных особенностей сигналов в пределах вейвлетного носителя на уровне региональных изменений и тренда, нулевое усиление постоянной составляющей сигналов, нулевое значение частотного спектра вейвлета при  $\omega=0$ , и локализацию спектра вейвлета в виде полосового фильтра с центром на определенной (доминирующей) частоте  $\omega_0$ . Для анализа мелкомасштабных флуктуаций и особенностей высокого порядка, как правило, требуются и нулевые значения определенного количества последующих моментов:

$$\int_{-\infty}^{\infty} t^m \psi(t) dt = 0.$$

Такие вейвлеты называются вейвлетами  $m$ -го порядка.

- **Ограниченность.** Необходимое и достаточное условие:

$$\|\psi(t)\|^2 = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty$$

Оценка ограниченности и локализации может выполняться с использованием выражений:

$$|\psi(t)| < 1/(1+|t|^n), \quad \text{или} \quad |\Psi(\omega)| < 1/(1+|\omega_0|^n),$$

где  $\omega_0$  – средняя частота вейвлета. Число  $n$  должно быть как можно больше.

- **Автомодельность базиса** или самоподобие. Форма всех базисных вейвлетов  $\psi_{ab}(t)$  должна быть подобна материнскому вейвлету  $\psi(t)$ , т.е. должна оставаться одной и той же при сдвигах и масштабировании (растяжении/сжатии), иметь одно и то же число осцилляций.

**Отображение преобразования.** Результатом вейвлет-преобразования одномерного числового ряда (сигнала) является двумерный массив значений коэффициентов  $C(a,b)$ . Распределение этих значений в пространстве  $(a,b)$  - временной масштаб, временная локализация, дает информацию об изменении во времени относительного вклада в сигнале вейвлетных компонент разного масштаба и называется спектром коэффициентов вейвлет-преобразования, масштабнo-временным (частотно-временным) спектром или просто вейвлет-спектром (wavelet spectrum).

Спектр  $C(a,b)$  одномерного сигнала представляет собой поверхность в трехмерном пространстве. Способы визуализации спектра могут быть самыми различными. Наиболее распространенный способ – проекция на плоскость  $ab$  с изолиниями (изоуровнями), что позволяет проследить изменения коэффициентов на разных масштабах во времени, а также выявить картину локальных экстремумов этих поверхностей ("холмов" и "впадин"), так называемый "скелет" (skeleton) структуры анализируемого процесса. При широком диапазоне масштабов применяются логарифмические координаты  $(\log a, b)$ . Пример

вейвлетного спектра простейшего сигнала при его разложении вейвлетом  $M_{hat}$  приведен на рис. 2.1.2.

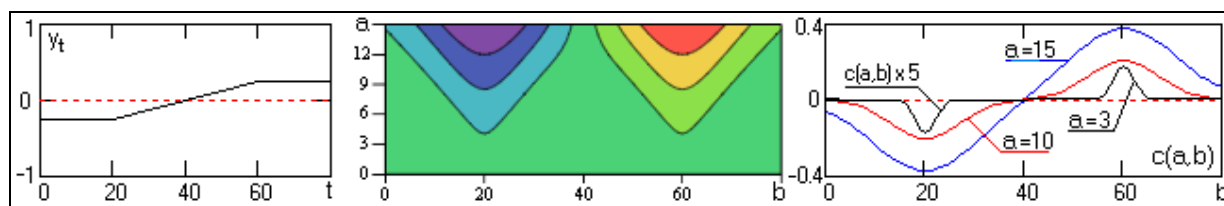


Рис. 2.1.2. Сигнал, вейвлетный  $M_{hat}$  - спектр и масштабные сечения спектра.

По вертикальным сечениям (сечениям сдвига  $b$ ) вейвлет-спектр отражает компонентный состав сигнала (из данного комплекта вейвлетов) в каждый текущий момент. По смыслу преобразования, как скалярного произведения сигнала с вейвлетом, ясно, что значения коэффициентов в каждой текущей временной точке по масштабным сечениям тем больше, чем сильнее корреляция между вейвлетом данного масштаба и поведением сигнала в окрестностях этой точки. Соответственно, сечения по параметру 'a' демонстрируют изменения в сигнале компоненты данного масштаба 'a' со временем.

Вейвлетные составляющие сигнала в сечениях его спектра не имеют ничего общего с синусоидами, и представлены, как правило, сигналами достаточно сложной и не всегда понятной формы, что может затруднять их наглядное представление и понимание.

**Вейвлетные функции.** Выбор анализирующего вейвлета определяется тем, какую информацию необходимо извлечь из сигнала. С учетом характерных особенностей различных вейвлетов во временном и в частотном пространстве, можно выявлять в анализируемых сигналах те или иные свойства и особенности, которые незаметны на графиках сигналов, особенно в присутствии шумов. При этом задача реконструкции сигнала может и не ставится, что расширяет семейство используемых регулярных вейвлетных функций, в том числе неортогональных. Более того, вейвлет может конструироваться непосредственно под ту локальную особенность в сигнале, которая подлежит выделению или обнаружению, если ее форма априорно известна.

При анализе сигналов вейвлетами четного типа (симметричными или близкими к симметричным) гармоническим сигналам обычно соответствуют яркие горизонтальные полосы вейвлетных пиков и впадин на доминирующих частотах вейвлетов, совпадающих с частотой гармоник сигналов. Нарушения гладкости сигналов фиксируются вертикальными полосами, пики в сигналах выделяются максимумами, а впадины – минимумами вейвлетных коэффициентов. Напротив, вейвлеты нечетного типа более резко реагируют на скачки и быстрые изменения в сигналах, отмечая их максимумами или минимумами в зависимости от знака дифференциалов. Чем резче выражены особенности сигналов, тем сильнее они выделяются на спектрограммах.

Для конструирования таких вейвлетов часто используются производные функции Гаусса, которые имеют наилучшую локализацию как во временной, так и в частотной областях. В общей форме уравнение базового вейвлета:

$$\psi_n(x) = (-1)^{n+1} d^n[\exp(-x^2/2)]/dx^n, \quad n \geq 1, \quad (2.1.1)$$

Уравнения базовых вейвлетов для первых четырех производных:

$$\begin{aligned} \psi_1(x) &= -x \exp(-x^2/2), \quad \psi_2(x) = (1-x^2) \exp(-x^2/2), \\ \psi_3(x) &= (3x-x^3) \exp(-x^2/2), \quad \psi_4(x) = (-4x^4+6x^2-3) \exp(-x^2/2), \end{aligned}$$

Уравнения нормированных базисов для временных сигналов:

$$\psi(t, a, b) = (K_n/\sqrt{2}) \psi_n(x), \quad x=(t-b)/a, \quad K_1=1.062, \quad K_2=0.867, \quad K_3 =0.548, \quad K_4=0.293.$$

Для сужения базовой формы вейвлетов применяется также упрощенная форма:

$$\psi_n(x) = (-1)^{n+1} d^n[\exp(-x^2)]/dx^n, \quad n \geq 1, \quad (2.1.1')$$

**WAVE-вейвлет** вычисляется по первой производной ( $n=1$ ) и приведен на рис. 2.1.3 во временной и частотной области для трех значений масштабных коэффициентов 'a'. Форма

вейвлета относится к нечетным функциям и, соответственно, спектр вейвлета является мнимым. Уравнение вейвлета по (2.1.1') с единичной нормой:

$$\psi(t, a, b) = \frac{-1.786}{\sqrt{2}} \cdot \frac{t-b}{a} \cdot \exp\left[-\left(\frac{t-b}{a}\right)^2\right]. \quad (2.1.2)$$

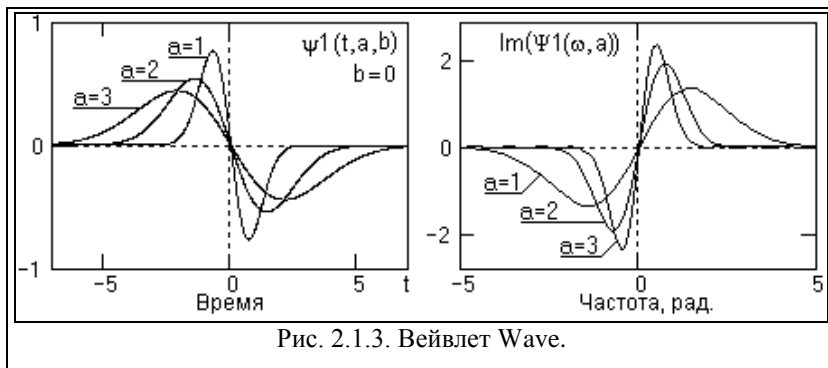


Рис. 2.1.3. Вейвлет Wave.

На рис. 2.1.4 приведен пример применения вейвлета для анализа двух однопериодических сигналов, один из которых осложнен шумами с мощностью на уровне мощности самого сигнала. Как следует из рисунка, контурная масштабно-

временная картина вейвлетных коэффициентов, а равно и ее сечения на больших значениях масштабных коэффициентов 'a' (малых доминирующих частотах вейвлетов) очень точно и уверенно фиксирует положение вершины информационного сигнала сменой знака коэффициентов C(a,b).

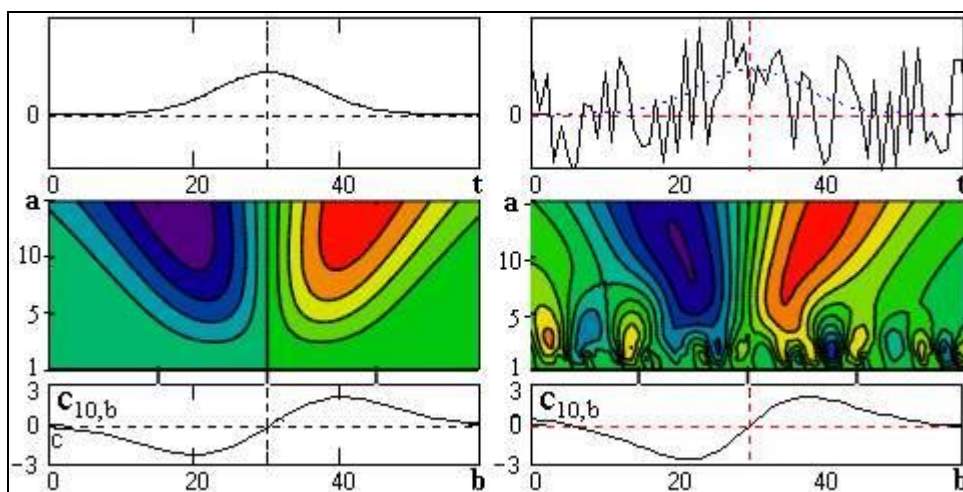


Рис. 2.1.4.

**МНАТ-вейвлет** (Mexican hat – мексиканская шляпа) вычисляется по второй производной (n=2) и приведен на рис. 2.1.5. Вейвлет симметричен, спектр вейвлета представлен только действительной частью и хорошо локализован по частоте, нулевой и первый моменты вейвлета равны нулю. Применяется для анализа сложных сигналов. Уравнение вейвлета по (2.1.1'):

$$\psi(t, a, b) = \frac{1.031}{\sqrt{2}} \cdot \left\{ \exp\left[-\left(\frac{t-b}{a}\right)^2\right] - 2\left(\frac{t-b}{a}\right)^2 \cdot \exp\left[-\left(\frac{t-b}{a}\right)^2\right] \right\}. \quad (2.1.3)$$

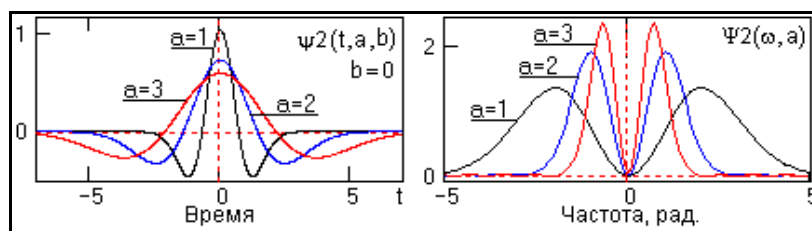


Рис. 2.1.5. Вейвлет МНАТ.

На рис. 2.1.6 приведен пример использования вейвлета для анализа сложного сигнала y(t). Модель сигнала образована суммой сигналов разной структуры. Сигналы y1-y2 представляют собой функции Гаусса разного масштабного уровня, сигнал y3 -

прямоугольный импульс, сигнал  $y_4$  задан в виде тренда с постоянным значением дифференциала. На контурном графике вейвлет-коэффициентов можно видеть выделение всех трех основных структур сигнала при полном исключении тренда. Особенно четко выделяются границы скачков прямоугольной структуры. Справа на рисунке приведена полная трехмерная картина вейвлет-преобразования.

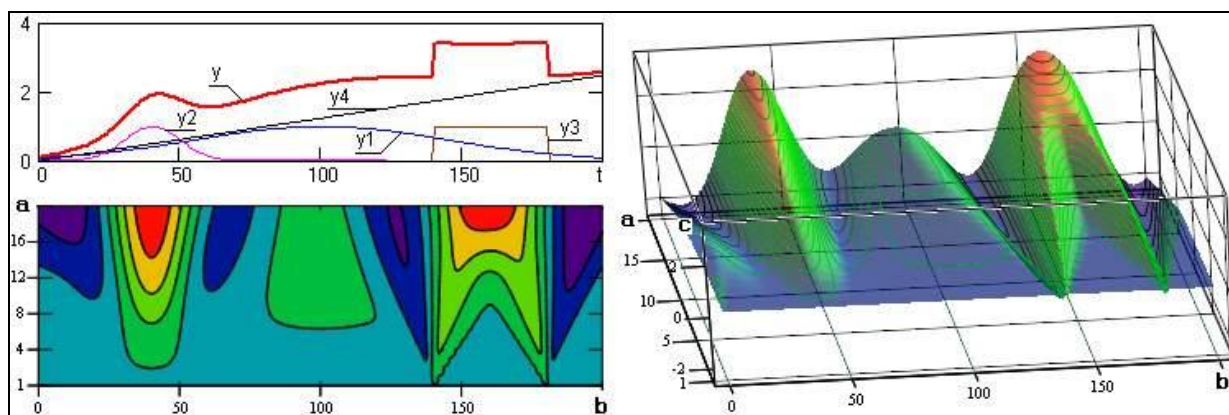


Рис. 2.1.6.

Вейвлет широко используется в двумерном варианте для анализа изотропных полей. На его основе возможно также построение двумерного неизотропного базиса с хорошей угловой избирательностью при добавлении к сдвигам и масштабированию вейвлета его вращения.

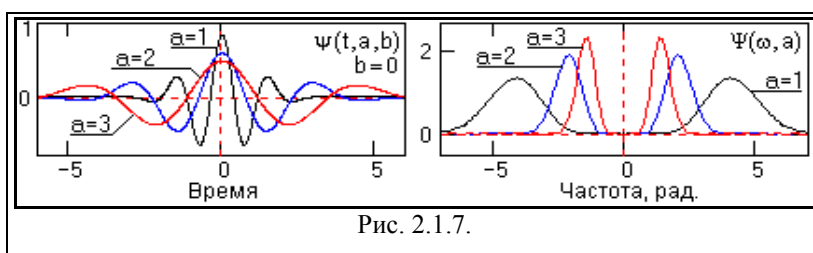


Рис. 2.1.7.

При повышении номера производной функции (2.1.1) временная область определения вейвлета несколько увеличивается при существенном повышении доминирующей частоты вейвлета и степени его

локализации в частотной области. Вейвлеты  $n$ -го порядка позволяют анализировать более тонкие высокочастотные структуры сигналов, подавляя низкочастотные компоненты. Пример вейвлета по восьмой производной приведен на рис. 2.1.7.

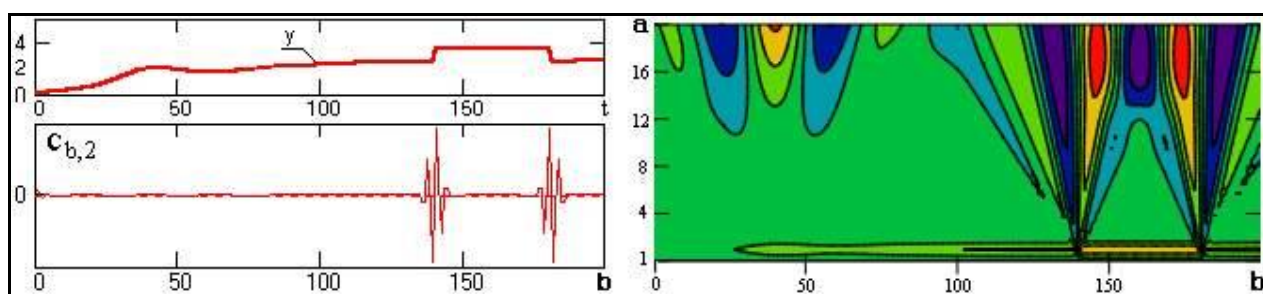


Рис. 2.1.8.

Практическое следствие повышения степени локализации вейвлетов в частотной области наглядно видно на рис. 2.1.8 на примере преобразования той же функции, что и на рис. 2.1.6. Сравнение рисунков показывает существенное повышение чувствительности вейвлета к высокочастотным составляющим сигнала на малых масштабных коэффициентах.

## 2.2. СВОЙСТВА ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

Результаты вейвлет-преобразования, как скалярного произведения вейвлета и сигнальной функции, содержат комбинированную информацию об анализируемом сигнале и самом вейвлете. Получение объективной информации о сигнале базируется на



свойствах вейвлет-преобразования, общих для вейвлетов всех типов. Рассмотрим основные из этих свойств. Для обозначения операции вейвлет-преобразования произвольных функций  $s(t)$  будем применять индекс  $TW[s(t)]$ .

**Линейность.**

$$TW[\alpha \cdot s_1(t) + \beta \cdot s_2(t)] = \alpha \cdot TW[s_1(t)] + \beta \cdot TW[s_2(t)]. \quad (2.2.1)$$

Для векторных функций из этого следует, что  $TW$  векторной функции есть вектор с компонентами  $TW$  каждой из компонент анализируемого вектора в отдельности.

**Инвариантность относительно сдвига.** Сдвиг сигнала во времени на  $t_0$  приводит к сдвигу вейвлет-спектра также на  $t_0$ :

$$TW[s(t-t_0)] = C(a, b-t_0). \quad (2.2.2)$$

**Инвариантность относительно масштабирования.** Растяжение (сжатие) сигнала приводит к сжатию (растяжению) вейвлет-спектра сигнала:

$$TW[s(t/a_0)] = (1/a_0) \cdot C(a/a_0, b/a_0). \quad (2.2.3)$$

**Дифференцирование.**

$$d^n \{ TW[s(t)] \} / dt^n = TW[d^n(s(t)) / dt^n]. \quad (2.2.4)$$

$$TW[d^n(s(t)) / dt^n] = (-1)^n \int_{-\infty}^{\infty} s(t) [d^n(\psi(t)) / dt^n] dt. \quad (2.2.5)$$

Отсюда следует, что безразлично, дифференцировать ли функцию или анализирующий вейвлет. Если анализирующий вейвлет задан формулой, то это может быть очень полезным для анализа сигналов. Проанализировать особенности высокого порядка или мелкомасштабные вариации сигнала  $s(t)$  с игнорированием крупномасштабных полиномиальных составляющих (тренда и регионального фона) можно дифференцированием нужного числа раз либо вейвлета, либо самого сигнала. Это свойство особенно полезно, когда сигнал задан дискретным рядом.

**Аналог теоремы Парсеваля** для ортогональных и биортогональных вейвлетов.

$$\int_{\mathbb{R}} s_1(t) \cdot s_2^*(t) dt = C_{\psi}^{-1} \int_{\mathbb{R}} a^{-2} C(a,b) C^*(a,b) da db. \quad (2.2.6)$$

Отсюда следует, что энергия сигнала может вычисляться через коэффициенты вейвлет-преобразования.

Определения и свойства одномерного непрерывного вейвлет-преобразования обобщаются на многомерный и на дискретный случаи.

### 2.3. ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ ПРОСТЫХ СИГНАЛОВ.

Вейвлет-преобразование, выполняемое при анализе сигналов для выявления в них каких-либо особенностей и места их локализации без обратной реконструкции, допускает применение любых типов вейвлетов, как ортогональных, так и неортогональных. Чаще всего для этих целей используются симметричные вейвлеты. Ниже приводятся результаты применения вейвлета  $M$  для анализа сигналов простых форм. Вычисления выполнены с вейвлетом (2.1.3) по формуле:

$$c(a,b) = \sum_{t=b-3a}^{b+3a} s(t) \psi(t,a,b), \quad (2.3.1)$$

где суммирование выполняется в растворе угла влияния (по области достоверности) с шагом  $\Delta t = \Delta b = \Delta a = 1$ . Так как при непрерывном разложении скейлинг-функция не используется, отсчет значений 'a' начинается с 1, а ряд коэффициентов  $c(0,b)$  оставляется нулевым и определяет нулевой фон контурных графиков спектра.

Импульсы Кронекера (положительный и отрицательный), вейвлет-спектр импульсов и сечения спектра на трех значениях параметра 'a' приведены на рис. 2.3.1. Цветовая гамма спектра здесь и в дальнейшем соответствует естественному цветоряду от красного (большие значения) к фиолетовому (малые значения коэффициентов).

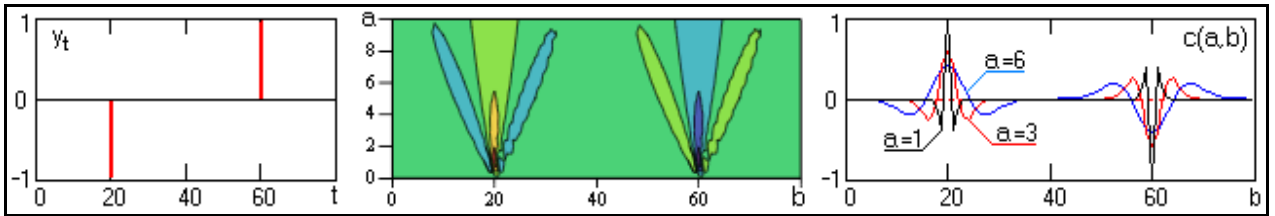


Рис. 2.3.1. Преобразование импульсов Кронекера.

На сечениях спектра видно, что свертка единичных импульсов с разномасштабными вейвлетами повторяет форму вейвлетов, как это и положено при операции свертки. Соответственно, линии максимальных экстремумов на сечениях ("хребты" и "долины", в зависимости от полярности) определяют временное положение импульсов, а боковые экстремумы противоположной полярности образуют характерные лепестки в конусе угла влияния, который хорошо выражен.

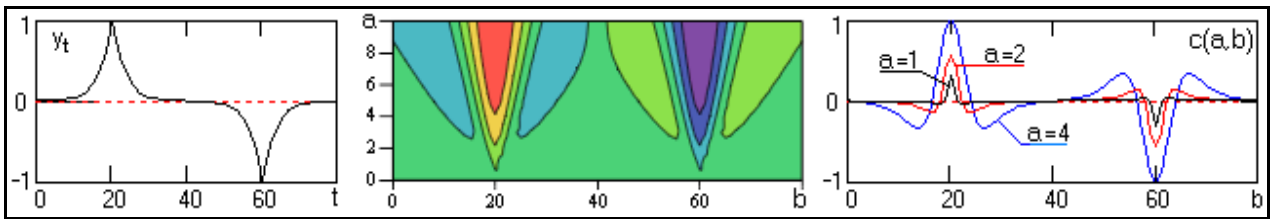


Рис. 2.3.2. Преобразование функций Лапласа.

Аналогичный характер спектра сохраняется и для любых локальных неоднородностей на сигналах в форме пиков (рис. 2.3.2) со смещением максимумов (минимумов) коэффициентов  $c(a,b)$  со значений  $a=1$  в область больших значений 'a' (в зависимости от эффективной ширины пиков).

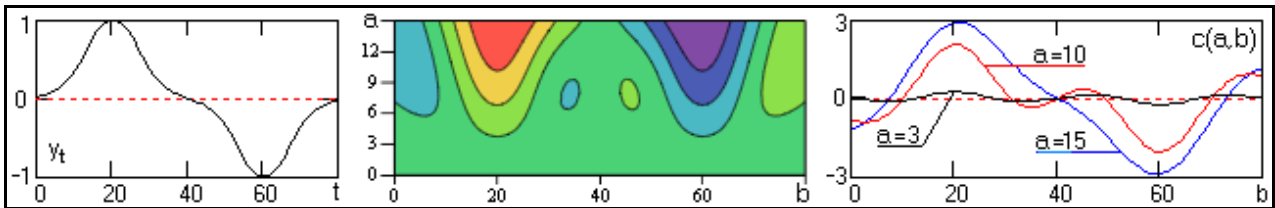


Рис. 2.3.3. Преобразование функций Гаусса.

На рис. 2.3.3 приведен спектр функций Гаусса. При сглаживании вершин пиковых неоднородностей форма цветковых конусов также сглаживается, но "хребтовые" ("долинные") линии достаточно точно фиксируют на временной оси положение центров локальных неоднородностей.

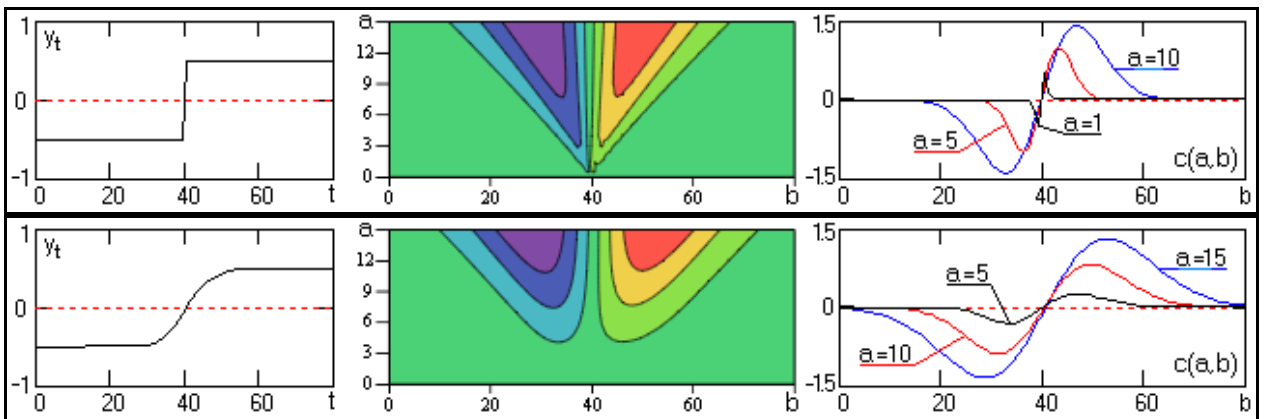


Рис. 2.3.4. Преобразование перепада постоянного значения функций.

На рис. 2.3.4 приведены спектры двух разных по крутизне перепадов постоянных значений функции. Центры перепадов фиксируются по переходу через нуль значений коэффициентов  $c(a,b)$ , а крутизна перепадов отражается, в основном, на значениях

функции  $c(a,b)$  при малых значениях параметра 'a'.

При изломах функций спектрограммы уверенно фиксируют место изломов максимумами (минимумами) значений коэффициентов  $c(a,b)$ , как это показано на рис. 2.3.5. При наложении на такие функции шумов точное определение места изломов по масштабным сечениям на малых значениях параметра 'a' становится невозможным, однако на больших значениях параметра 'a' такая возможность сохраняется, естественно, с уменьшением точности локализации.

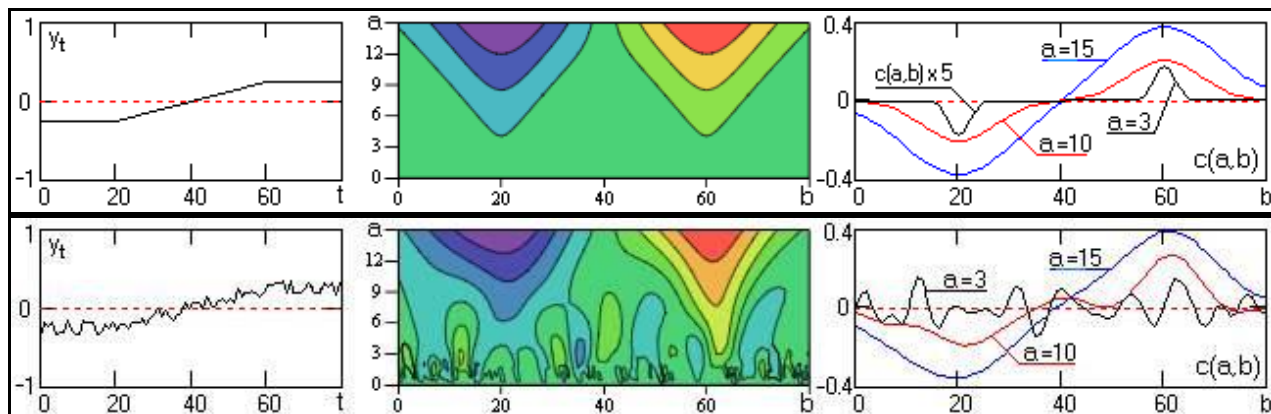


Рис. 2.3.5. Преобразование изломов функций.

Аналогичный характер имеет влияние шумов и на другие локальные сигналы, приведенные на рис. 2.3.1-2.3.4, и если спектральные особенности сигналов достаточно глубоки по диапазону значений параметра 'a', то остается возможность идентификации этих локальных сигналов и их места на временной оси.

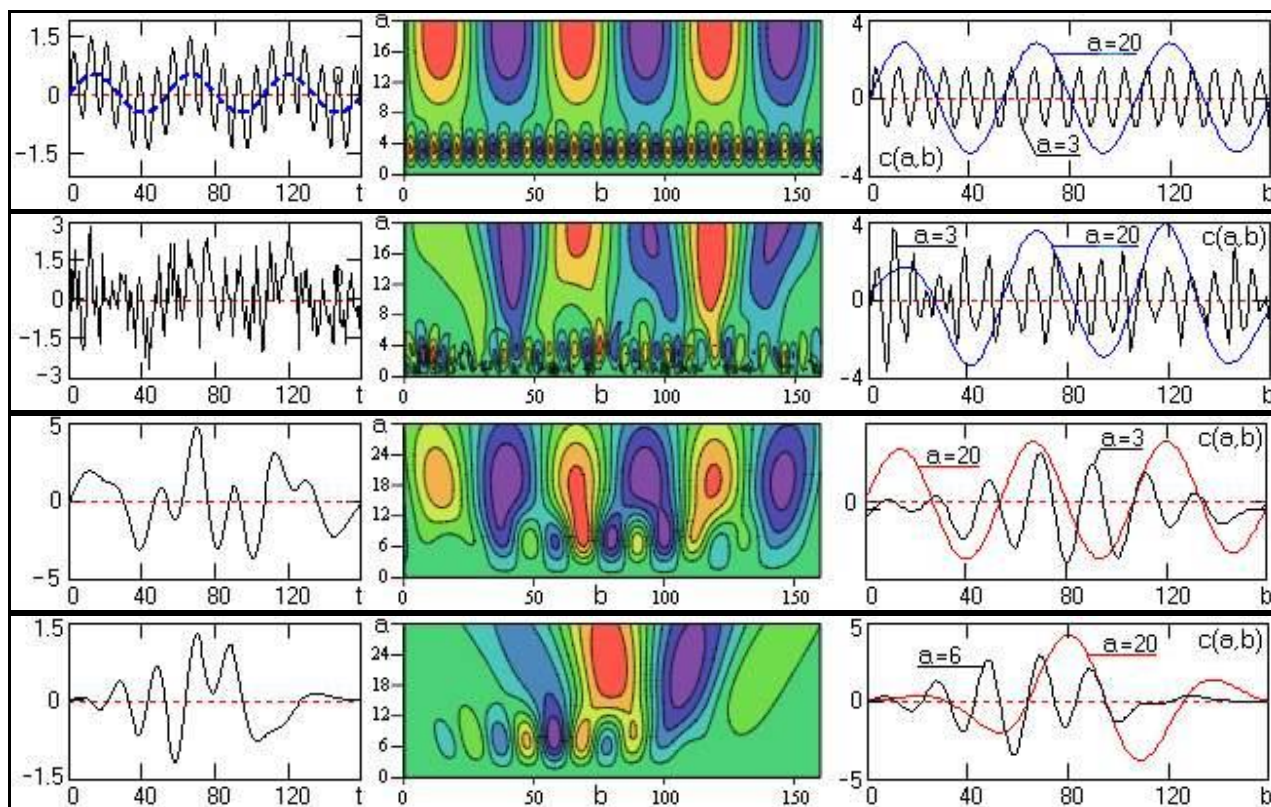


Рис. 2.3.6. Преобразование гармонических функций.

Разделение гармонических функций на масштабной оси спектров, в том числе при наложении сильных шумовых процессов, приведено в примерах на рис. 2.3.6. Приведенный пример имеет чисто иллюстративный характер, так как для выделения гармонических процессов с постоянной частотой во времени целесообразно использовать



спектральный анализ и частотные полосовые фильтры. Тем не менее, для локальных сигналов, типа модулированных гармоник, вейвлет-спектры достаточно хорошо показывают место их локализации на временной оси.

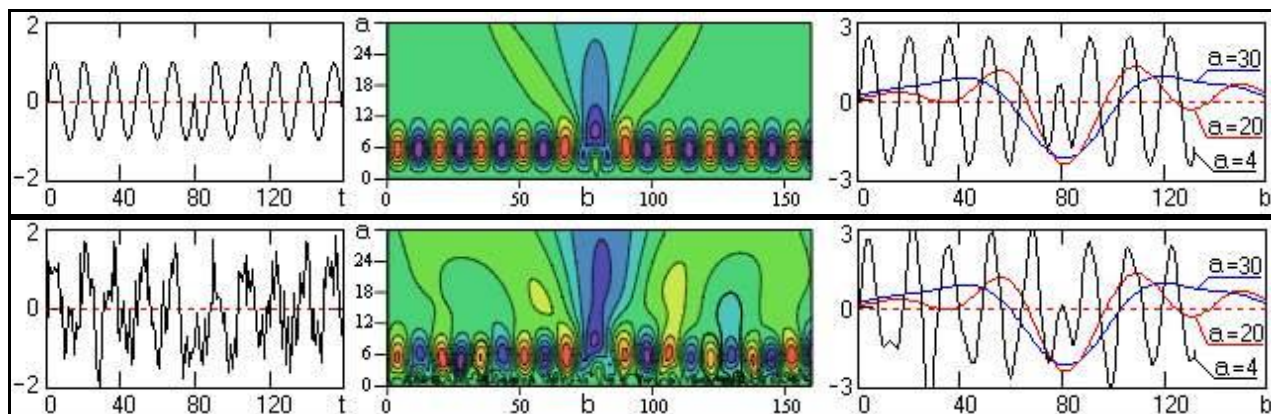


Рис. 2.3.7. Изменение фазы гармонического сигнала.

На рис. 2.3.7 приведен пример еще одной характерной особенности гармонического сигнала – изменение его фазы на  $180^\circ$ , которое хорошо фиксируется на всех масштабах вейвлета, а, следовательно, достаточно легко определяется даже в присутствии сильных шумовых сигналов.

При наложении синусоидальных сигналов на тренд вейвлет-преобразование на больших масштабах позволяет достаточно уверенно выделять характерные особенности тренда. Пример выделения изломов тренда приведен на рис. 2.3.8.

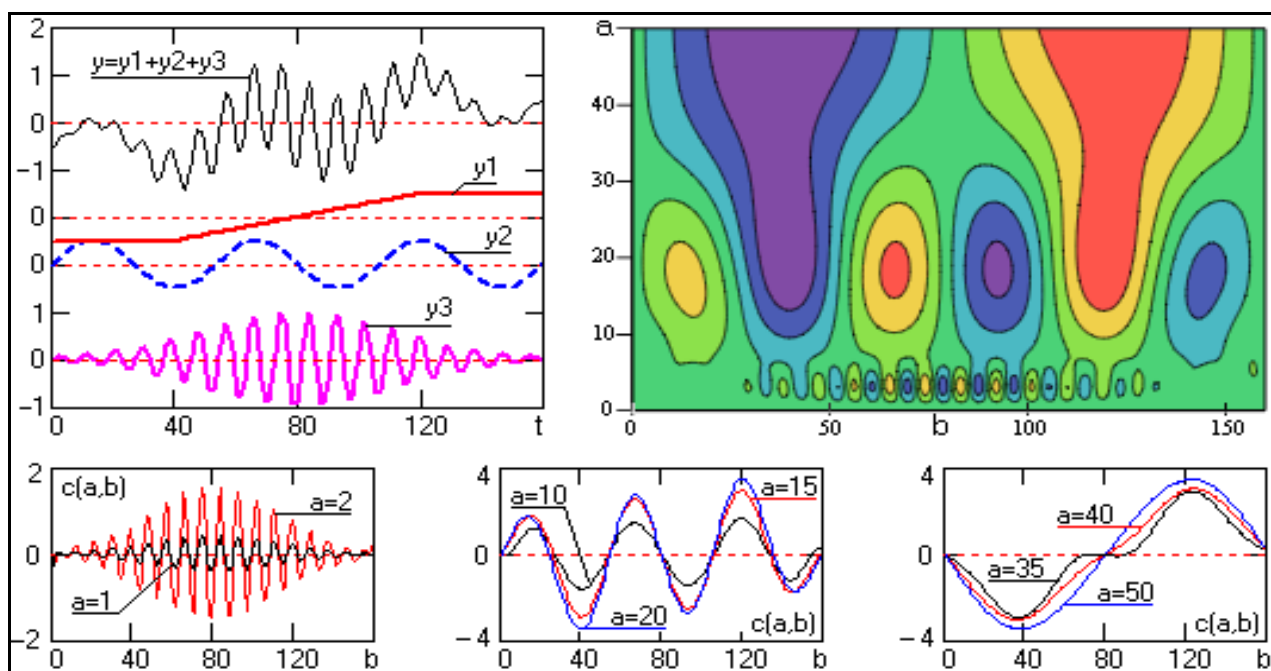


Рис. 2.3.8. Преобразование суммы трех сигналов.

Форма вейвлета (четность или нечетность), доминирующая частота и степень ее локализации существенно влияют на вейвлет-спектры анализируемых сигналов и на возможности выделения его локальных особенностей. На нижеследующих рисунках приведены сравнительные спектры простых сигналов при использовании вейвлетов Wave (нечетный, рис. 2.1.3), Mhat (четный, рис. 2.1.5) и вейвлета по 8-й производной Гаусса (рис. 2.3.9-2.3.16), который также является четным, и имеет в 4 раза более высокую доминирующую частоту, чем вейвлет Mhat.

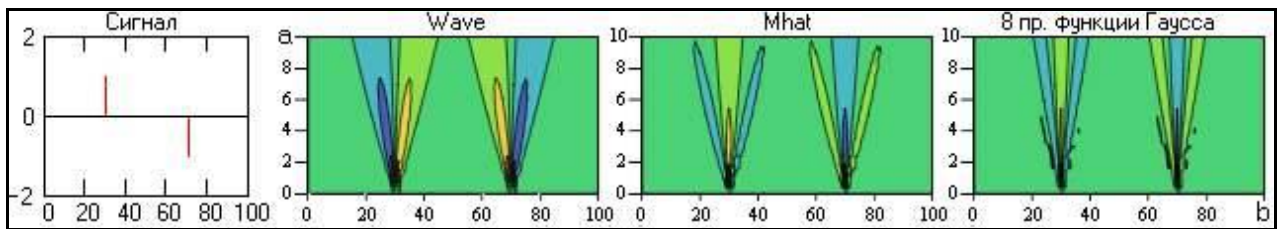


Рис. 2.3.9. Импульсы Кронекера.

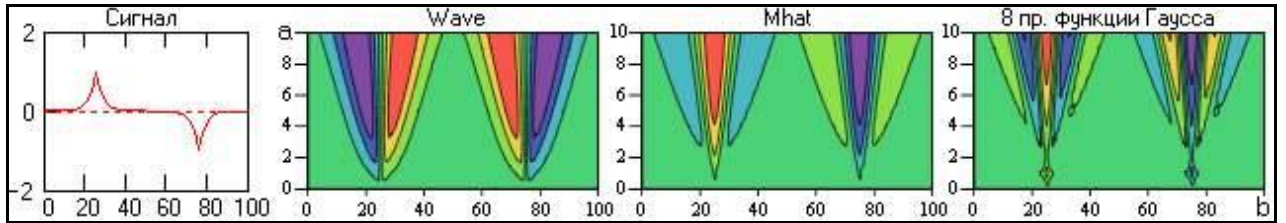


Рис. 2.3.10. Пики Лапласа.

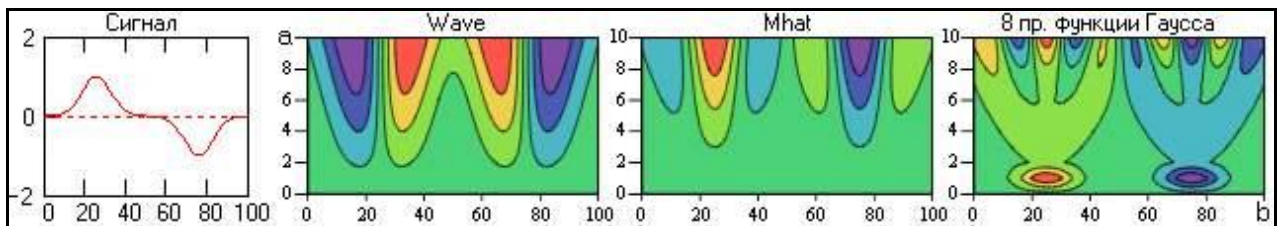


Рис. 2.3.11. Функции Гаусса.

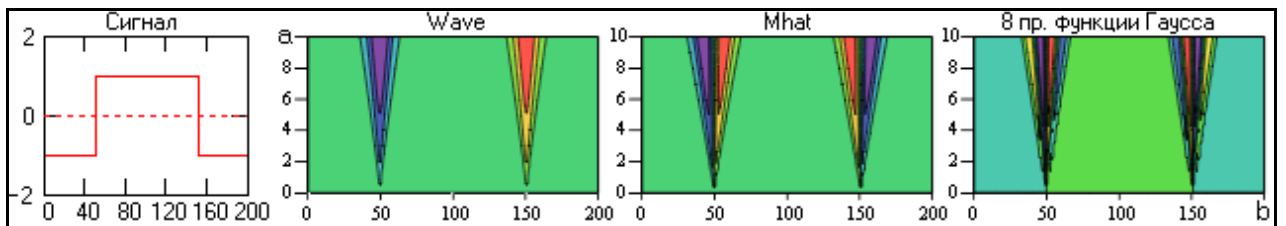


Рис. 2.3.12. Крутые скачки.

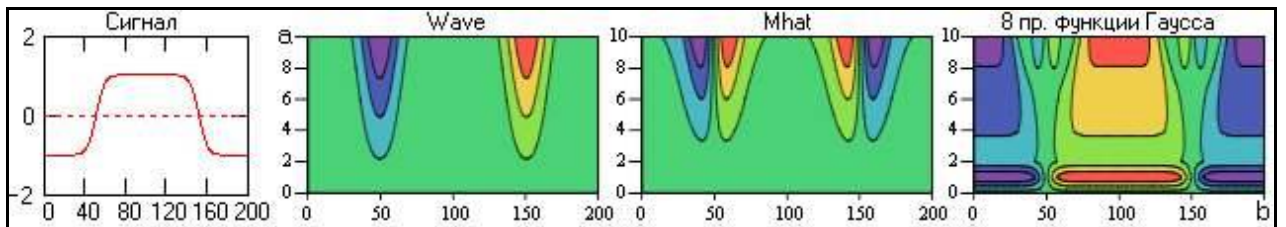


Рис. 2.3.13. Сглаженные скачки.

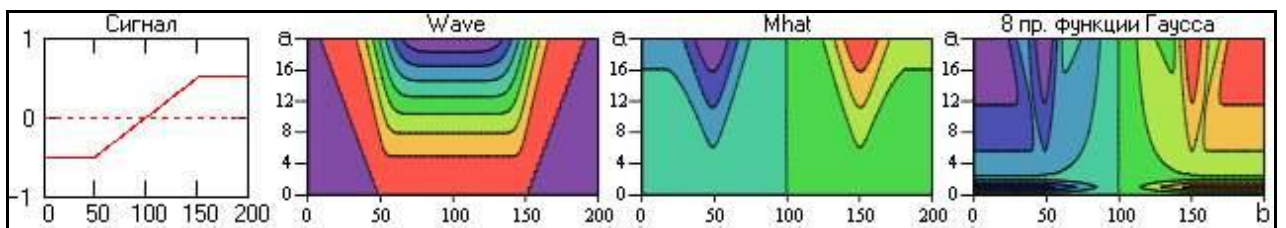


Рис. 2.3.14. Изломы функций

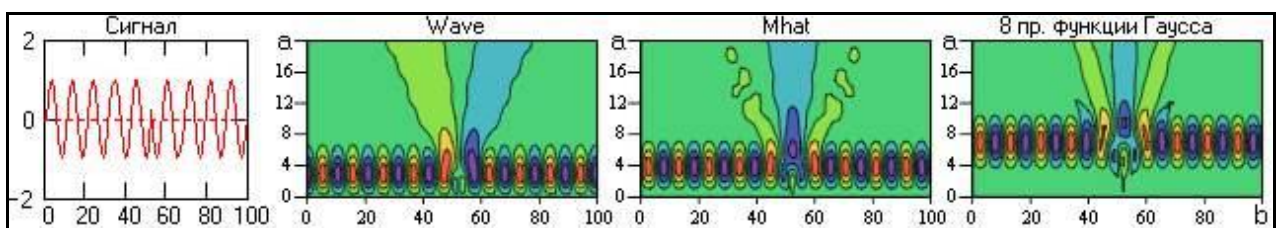


Рис. 2.3.15. Фазовые скачки гармоник.

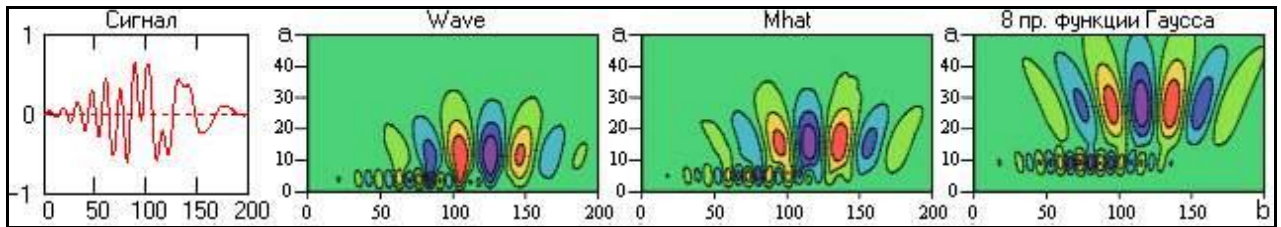


Рис. 2.3.16. Сумма двух модулированных синусоид.

Заметим, что при анализе произвольных сигналов использование разнотипных вейвлетов позволяет повысить достоверность выделения локальных особенностей сигналов.

## Тема 21. ВЕЙВЛЕТНЫЙ КРАТНОМАСШТАБНЫЙ АНАЛИЗ

В практике передачи информации часто требуется представить сигнал в виде совокупности его последовательных приближений. Например, при просмотре и передаче изображений с выборкой из некоторой базы данных можно сначала передать грубую его версию, а затем (при необходимости) последовательно ее уточнять. При сжатии изображений часто без потери качества можно убирать из изображения незначимые мелкомасштабные детали.

Произвольный информационный сигнал обычно рассматривается в виде суммы разнотипных составляющих: региональной функции тренда, циклических компонент с определенным периодом повторения, локальных особенностей (аномалий) разного порядка и флуктуаций (шумов) вокруг всех вышеперечисленных составляющих сигнала. Инструментом разделения (декомпозиции) сигналов на такие составляющие, анализа их порядка и реконструкции сигналов из определенных составляющих (или с исключением определенных составляющих, например шумов или малозначимых деталей) является кратномасштабный (многомасштабный) анализ (КМА).

КМА позволяет получить хорошее разрешение по времени (плохое по частоте) на высоких частотах и хорошее разрешение по частоте (плохое по времени) на низких частотах. Этот подход становится эффективным, если сигнал имеет короткие высокочастотные компоненты и протяженные низкочастотные компоненты. Именно такие сигналы и встречаются чаще всего.

Идея кратномасштабного анализа заключается в том, что разложение сигнала производится по ортогональному базису, образованному сдвигами и кратномасштабными копиями вейвлетной функции. Свертка сигнала с вейвлетами позволяет выделить характерные особенности сигнала в области локализации этих вейвлетов, причем, чем больший масштаб имеет вейвлет, тем более широкая область сигнала будет оказывать влияние на результат свертки.

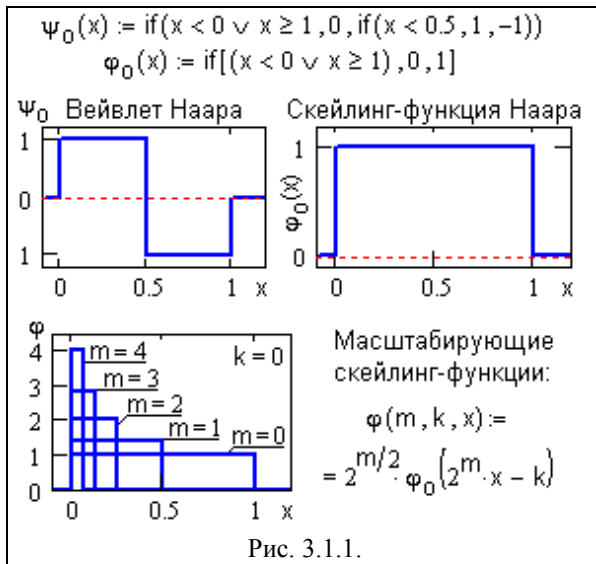
Понятие кратномасштабного анализа (Multiresolution analyses) является фундаментальным в теории вейвлетов. Для кратномасштабного анализа разработан быстрый каскадный алгоритм вычислений, подобный быстрому преобразованию Фурье.

### 3.1. ПРИНЦИП КРАТНОМАСШТАБНОГО АНАЛИЗА

**Дискретные ортогональные преобразования.** Непрерывное вейвлет-преобразование, равно как и его дискретный аналог с произвольным шагом по масштабу и сдвигу, обладает сильной избыточностью. Интуитивно понятно, что если какая-либо информация заключена в  $N$  отсчетах сигнала, то при любых преобразованиях сигнала для отображения этой информации без потерь в новом базисном пространстве должно быть необходимо и достаточно то же самое количество отсчетов  $N$ . С учетом принципа неопределенности Гейзенберга это означает, что для точного восстановления сигнала достаточно знать его вейвлет-преобразование на некоторой решетке частотно-временной области, густой в области высоких частот сигнала, и редкой в области низких частот. Идея КМА заключается в том, чтобы масштабировать вейвлет в некоторое постоянное число



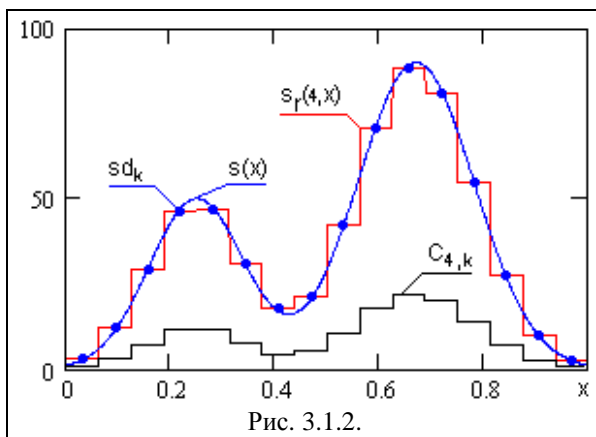
раз (например, 2), и при скольжении по сигналу сдвигать его во времени с шагом, равным интервалу носителя масштабированного вейвлета. Если обозначить количество масштабных строк индексом  $m$ , и принять  $N=2^m$ , то при  $N=32$  решетка вейвлетного спектра будет иметь всего  $m=5$  масштабных строк с количеством отсчетов в первой строке 16, во второй 8, в третьей 4, в четвертой 2, и в пятой 1, с общим количеством отсчетов 32, как и в исходном сигнале. При этом все сдвиги одного масштаба будут попарно ортогональны (нет перекрытия сдвигов), равно как и вейвлеты разных масштабов в силу их нулевого первого момента.



**Вейлет Хаара.** Простейшие методы КМА, без всякой теоретической базы, использовались при обработке числовых данных уже достаточно давно. Рассмотрим один из таких методов на практическом примере анализа гистограмм, который обычно выполняется функцией Хаара (Haar), в дальнейшем получившей название вейвлета Хаара (рис. 3.1.1).

Допустим, что мы анализируем определенную зависимость  $s(x)$  на интервале  $0 \leq x \leq 1$ , показанную на рис. 3.1.2. Функция нецентрированная, и для использования вейвлет-преобразования с последующим восстановлением исходного сигнала требует применения как вейвлета, так и его скейлинг-

функции. На основе базовых функций вейвлета и скейлинг-функции Хаара, приведенных на рис. 3.1.1, записываем масштабированные функции:



$$\varphi_{m,k}(x) = 2^{m/2} \varphi(2^m x - k), \quad (3.1.1)$$

$$\Psi_{m,k}(x) = 2^{m/2} \psi(2^m x - k). \quad (3.1.2)$$

Эти функции образуют нормированные взаимно ортогональные базисы пространства вейвлетных коэффициентов, на которые может быть разложен анализируемый сигнал. Ортогональность базисных функций является обязательным условием КМА, обеспечивающим возможность обратной реконструкции сигнала.

Для коротких и достаточно гладких кривых нет смысла устанавливать много уровней декомпозиции сигнала. Примем максимальное значение  $m$ , равным 4, при этом  $N=1/2^m=16$  с интервалом дискретизации данных, соответственно,  $\Delta x=1/N$ . В принципе, можно применять и задание исходного значения  $\Delta x$  с последующим определением количества уровней разложения.

При сдвиговой ортогональности прямоугольных базисных функций прямое преобразование (проекции сигнала на базис (3.1.1)) для непрерывных сигналов выполняется по формуле:

$$C_{m,k} = \int_{k\Delta x}^{(k+1)\Delta x} s(x) \varphi(2^m x - k) dx. \quad (3.1.3)$$

Значения коэффициентов при  $m=4$ :

$C_{m,k} =$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	0	0.774	2.996	7.384	11.596	11.65	7.733	4.471	5.355	10.615	17.721	22.048	20.246	13.711	6.848	2.522	0.684

Восстановление сигнала с четвертого уровня декомпозиции соответственно выполняется по формуле реконструкции:

$$s_r(m,x) = \sum_{k=0}^{N-1} C_{m,k} \varphi_{m,k}(x), \quad m=4, N=16. \quad (3.1.4)$$

$$s_r\left(k\Delta x + \frac{\Delta x}{2}\right) =$$

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	0	3.095	11.984	29.536	46.385	46.599	30.933	17.884	21.418	42.462	70.883	88.193	80.983	54.846	27.391	10.086	2.738

Восстановление исходной непрерывной функции сигнала  $s(x)$  скейлинг-функцией Хаара невозможно в силу того, что значение скейлинг-функции – константа шириной  $\Delta x$ , на которую умножается соответствующее значение  $C_{4,k}$  и распространяется на весь интервал  $k\Delta x - (k+1)\Delta x$  (кривая  $s_r(x)$  на рис. 3.1.2). Если выполнить перевод сигнала  $s(t)$  во временной ряд  $sd_k, k=0 \dots N-1$ , с осреднением по интервалам  $\Delta x$ , или с использованием (в общем случае произвольного вейвлета) его скейлинг-функции:

$$sd_k = 2^{m/2} \int_{k\Delta x}^{(k+1)\Delta x} s(x) \varphi_{m,k}(x) dx, \quad (3.1.5)$$

$$sd =$$

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	0	3.095	11.984	29.536	46.385	46.599	30.933	17.884	21.418	42.462	70.883	88.193	80.983	54.846	27.391	10.086	2.738

то нетрудно убедиться, что  $sd_k = s_r(k\Delta x + \Delta x/2)$  (числовые отсчеты  $sd_k$  на рис. 3.1.2 отнесены к середине интервалов  $\Delta x$ ).

В принципе, гистограмма  $sd_k$  может представлять собой непосредственные исходные дискретные данные (результаты измерений и т.п.). Сравнением выражений (3.1.5) и (3.1.3) нетрудно убедиться, что нулевой уровень разложения ( $m=m_{\max}$ ) может быть получен непосредственно из дискретных данных:

$$C_{m,k} = sd_k / 2^{m/2}. \quad (3.1.3')$$

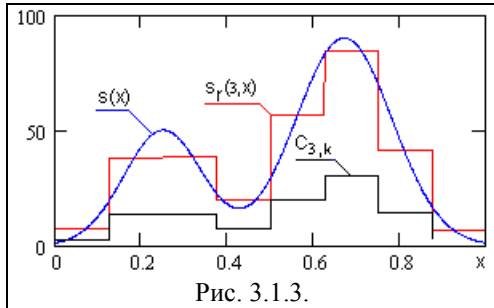


Рис. 3.1.3.

На следующем уровне разложения функции, при  $m=3$ , скейлинг-функция (3.1.1) расширяется по  $x$  в 2 раза (в нашем примере до  $1/8$ ), т.е. производится усреднение отсчетов по двум соседним интервалам исходной гистограммы. Количество коэффициентов соответственно в 2 раза уменьшается. Расчет коэффициентов  $C_{3,k}$  может выполняться непосредственно по (3.1.3), реконструкция  $s_{r3}(x)$  – по (3.1.4), при  $m=3, N=8$ . Тем самым аппроксимация

исходного сигнала выполняется на более "грубом" уровне декомпозиции, на основании чего скейлинг-функции вейвлетов называют аппроксимирующими или масштабными функциями, а сами коэффициенты, выделенные скейлинг-функциями – аппроксимирующими.

Но при известных значениях коэффициентов  $C_{4,k}$  предшествующего уровня следующий уровень может выполняться непосредственно по ним с учетом изменения нормировочного множителя в формуле скейлинг-функции (3.1.1). В общей форме:

$$C_{m-1,k} = (1/\sqrt{2}) (C_{m,2k} + C_{m,2k+1}). \quad (3.1.6)$$

$$C_{3,k} = \{2.665, 13.421, 13.706, 6.948, 20.037, 29.906, 14.538, 2.267\}.$$

Кроме аппроксимирующих коэффициентов  $C_{m-1,k}$  из предшествующей гистограммы аппроксимации  $C_{m,k}$  могут быть выделены также коэффициенты изменения сигнала в пределах нового интервала усреднения, т.е. коэффициенты разности значений первой и второй половины интервала:

$$D_{m-1,k} = (1/\sqrt{2}) (C_{m,2k} - C_{m,2k+1}), \quad (3.1.7)$$

которые называют детализирующими коэффициентами.

$$D_{3,k} = \{-1.571, -2.979, 2.769, -0.625, -5.024, 1.275, 4.853, 1.299\}.$$

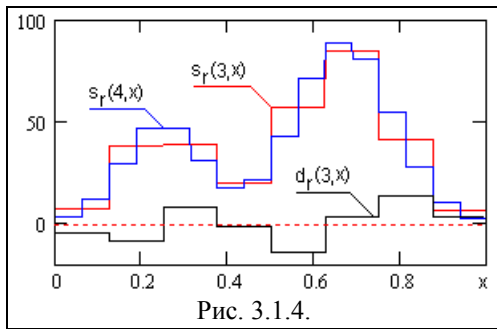


Рис. 3.1.4.

На рис. 3.1.4 показан график  $d_r(3,x)$  детализирующих коэффициентов ( $m=3$ ), приведенный к масштабу исходного сигнала по формуле (3.1.4) при  $m=3$  и  $N=2^m=8$ , по которому нетрудно понять их физическую сущность. Так как значения сигнала в интервале разложения  $2\Delta t$  по  $m=3$  представляют собой среднее значение сигналов в двух интервалах  $\Delta t$  разложения по  $m=4$ , которые они перекрывают, а детализирующий коэффициент

(с учетом приведения к масштабу исходного сигнала) равен половине разности сигналов этих двух интервалов, то его значение есть не что иное, как флюктуация сигнала по  $m=4$  относительно его аппроксимации по  $m=3$ . Если детализирующий коэффициент отрицателен, то эта флюктуация отрицательна относительно аппроксимированного значения в первой половине его интервала и положительна во второй, и наоборот. Т.е. соответствующие коэффициенты аппроксимации  $C_{m-1,k}$  и детализации  $D_{m-1,k}$  разделяют коэффициенты  $C_{m,k}$  предшествующего уровня декомпозиции сигнала на аппроксимированную (низкочастотную) и флюктуационную (высокочастотную) части.

Отсюда следует, что ряды коэффициенты  $C_{m-1,k}$  и  $D_{m-1,k}$  (количество точек  $2^{m-1}$  в каждом ряде) содержат полную информацию, адекватную информации в  $C_{m,k}$  предшествующего уровня (количество точек  $2^m=2^{m-1}+2^{m-1}$ ), что позволяют полностью восстановить значения коэффициентов более высокого уровня  $m$ :

$$C_{m,2k} = (1/\sqrt{2}) (C_{m-1,k} + D_{m-1,k}), \quad C_{m,2k+1} = (1/\sqrt{2}) (C_{m-1,k} - D_{m-1,k}), \quad (3.1.8)$$

а, следовательно, и восстановить исходный дискретный сигнал. Для восстановления значений в исходных интервалах при  $m=4$ , значение аппроксимирующего коэффициента на первой половине интервала при  $m=3$  складывается с детализирующим коэффициентом, а на второй – вычитается. Для математического отображения этой операции введем функцию  $\psi$  (3.1.2), форма которой приведена на рис. 3.1.1, и обеспечим ее сдвиг по координате синхронно со скейлинг-функцией. Эта функция является ортонормированным базисом разложения детализирующих коэффициентов. Именно она и получила название вейвлета (вейвлетной или детализирующей функции). С ее использованием уравнение (3.1.4) с входящими в него уравнениями (3.1.8) приводятся к следующей форме (с уровня  $m=3$ ,  $2^{m-1}=8$ ):

$$s_r(3, x) = \sum_{k=0}^7 C_{3,k} \varphi_{3,k}(x) + \sum_{k=0}^7 D_{3,k} \psi_{3,k}(x). \quad (3.1.9)$$

Как и значения коэффициентов  $C_{m,k}$ , значения детализирующих коэффициентов могут вычисляться непосредственно по формуле (3.1.3) с заменой скейлинг-функции на вейвлет-функцию.

Аналогичным образом операция разделения на аппроксимирующие и детализирующие коэффициенты может быть продолжена над значениями коэффициентов  $C_{3,k}$  по уровню  $m=2$ , с выделением коэффициентов аппроксимации  $C_{2,k}$  и детализации  $D_{2,k}$ , и далее по уровням  $m=1$  и  $m=0$ . На последнем уровне  $m=0$  получаем только 1 коэффициент аппроксимации  $C_0$  и детализации  $D_0$  по всему интервалу задания сигнала  $0 \leq x \leq 1$ . Применяя последовательно, начиная с  $m=0$ , функцию "сборки" сигнала (3.1.9), получаем общую формулу реконструкции сигнала:

$$s_r(x) = C_0 \cdot \varphi_0(x) + D_0 \cdot \psi_0(x) + \sum_{k=0}^1 D_{1,k} \cdot \psi_{1,k}(x) + \sum_{k=0}^3 D_{2,k} \cdot \psi_{2,k}(x) + \sum_{k=0}^7 D_{3,k} \cdot \psi_{3,k}(x). \quad (3.1.10)$$

**Свойства преобразования.** Отметим на этом примере характерные и очевидные особенности нового представления сигнала:

- Общее количество коэффициентов разложения равно количеству отсчетов исходного

сигнала (условие необходимости и достаточности сохранения в новом математическом представлении исходного объема информации).

- Вейвлет и его скейлинг-функция должны иметь однозначную связь. Это определяется тем, что разложение сигнала может быть выполнено с использованием только скейлинг-функции, а детализирующие коэффициенты определяться по разности  $m$  и  $m+1$  уровней аппроксимации, и наоборот.

- Значение  $C_0$  представляет собой среднее значение исходного сигнала по интервалу его задания. Для центрированных сигналов это значение равно нулю. При выполнении разложения без скейлинг-функции (с вейвлетом в (3.1.2)) картина детальных особенностей нецентрированных сигналов остается без изменений, но полная реконструкция сигнала невозможна. Без значения  $C_0 \cdot \varphi_0(x)$  при полном разложении сигнал центрируется, при реконструкции с других масштабов декомпозиции искажается за счет отсутствия соответствующих коэффициентов  $C_{m,k}$ .

- Увеличение масштабного значения  $m$  разложения соответствует возрастанию временного разрешения сигнала ( $1/2^m$ ). Коэффициенты вейвлет-преобразования вскрывают флюктуационную структуру сигнала на разных масштабах и в разных временных точках. В областях "гладких" значений сигнала коэффициенты детализации близки к нулевым и ими можно пренебречь, что позволяет осуществлять сжатие информации для хранения.

- Реконструкция сигнала возможна с любого масштабного уровня декомпозиции, причем все особенности сигнала сохраняются без искажений с временным разрешением первого вейвлета (с минимальной шириной окна).

### 3.2. МАТЕМАТИЧЕСКИЕ ОСНОВЫ КРАТНОМАСШТАБНОГО АНАЛИЗА

Разложение сигнала на сумму аппроксимирующих и детализирующих составляющих производится с использованием ортогональных и биортогональных вейвлетов. На таких вейвлетах выполняется быстрое вейвлет-преобразование. При выполнении КМА пространство сигналов  $L^2(\mathbb{R})$  представляется в виде системы вложенных подпространств  $V_m$ , отличающихся друг от друга перемасштабированием независимой переменной.

Исходные условия ортогонального кратномасштабного анализа.

В качестве пространства сигналов будем рассматривать  $L^2(\mathbb{R})$  – пространство функций  $s(t)$  с конечной энергией. В этом пространстве определено скалярное произведение и норма функций:

$$\langle s(t), g(t) \rangle = \int_{-\infty}^{\infty} s(t) g^*(t), \quad \|s(t)\| = \sqrt{\langle s(t), s(t) \rangle}.$$

Базисом в пространстве  $V \subset L^2(\mathbb{R})$  называется такая система функций  $\{v_n(t)\}$ , что любая функция  $v(t) \in V$  единственным образом записывается в виде  $v(t) = \sum_n c_n v_n(t)$ . Базис называется ортонормированным, если  $\langle v_i(t), v_j(t) \rangle = \delta_{ij}$ . В этом случае  $c_n = \langle v(t), v_n(t) \rangle$ .

Под кратномасштабным анализом понимается описание пространства  $L^2(\mathbb{R})$  через иерархические вложенные подпространства  $V_m \subset L^2(\mathbb{R})$ ,  $m = 0, \pm 1, \pm 2, \dots$ , которые не пересекаются и объединение которых в пределе дает  $L^2(\mathbb{R})$ . Система подпространств должна удовлетворять следующим условиям.

1. Условие вложенности:  $V_m \subset V_{m+1}$ .

Все пространство сигналов  $L^2(\mathbb{R})$  в целом может быть представлено в виде последовательности вложенных друг в друга замкнутых подпространств соответствующих уровней  $m$  декомпозиции сигнала:

$$\dots \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots \subset V_m \subset V_{m+1} \dots$$

"Размеры" подпространств непрерывно расширяются по мере роста значения  $m$ , объединение всех подпространств в пределе дает пространство  $L^2(\mathbb{R})$ .

2. Условие полноты и плотности разбиения:

$$\bigcup_{m \in I} V_m = L^2(\mathbb{R}). \quad (3.2.1)$$

3. Условие ортогональности подпространств:

$$\bigcap_{m \in I} V_m = \{0\}. \quad (3.2.2)$$

4. Условие сохранения в подпространстве при сдвигах функций:

$$v(t) \in V_m \Leftrightarrow v(t+1) \in V_m.$$

5. Для любой функции  $v(t) \in V_m$  ее масштабное преобразование по аргументу в 2 раза перемещает функцию в соседнее подпространство:

$$v(t) \in V_m \Leftrightarrow v(2t) \in V_{m+1}, \quad v(t) \in V_m \Leftrightarrow v(t/2) \in V_{m-1} \quad (3.2.3)$$

6. Для пространства  $V_0$  существует  $\phi$ -функция  $\phi(t) \in V_0$ , целочисленные сдвиги которой по аргументу образуют ортонормированный базис пространства  $V_0$ :

$$\phi_{0,k} = \phi(t-k), \quad k \in I \quad (k=0, \pm 1, \pm 2, \dots). \quad (3.2.4)$$

Функция  $\phi(t)$  называется скейлинг-функцией (scaling function). Условие нормирования скейлинг-функции:

$$\int_{-\infty}^{\infty} \phi(t) dt = 1.$$

Из этих условий следует, что если подпространство  $V_0$  имеет ортонормированный базис  $\phi_{0,k}$ , то и все остальные подпространства также имеют ортонормированные базисы, которые образуются масштабным преобразованием базиса  $\phi_{0,k}$ :

$$\phi_{m,k}(t) = a^{m/2} \phi(a^m t - k), \quad m, k \in I. \quad (3.2.5)$$

Стандартное значение параметра 'a' в кратномасштабном анализе равно 2. Так, например, если  $\phi_{0,k}(t)=1$  на интервале  $[0,1)$  и  $\phi_{0,k}(t)=0$  вне этого интервала, то целочисленные сдвиги этой функции попарно ортогональны, и пространство  $V_0$  состоит из функций, имеющих постоянные значения на интервалах вида  $[k, k+1)$ ,  $V_1$  – из функций, постоянных на интервалах  $[k/2, (k+1)/2)$ ,  $V_{-1}$  – из функций, постоянных на интервалах  $[2k, 2k+1)$ , и т.д.

Если сигнал  $v(t)$  принадлежит пространству  $V_m$ , то одновременно он входит и в пространство  $V_{m+1}$ , и вместе с ним в этом пространстве находится и сигнал  $v(2t)$ . Увеличение номера пространства позволяет изучать все более и более мелкие детали и особенности сигнала с более высокочастотными компонентами (как под микроскопом).

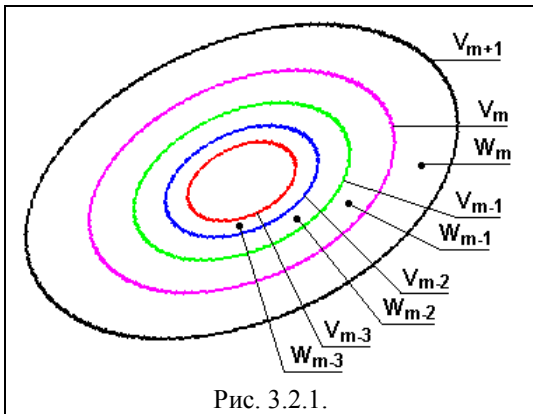


Рис. 3.2.1.

На рис. 3.2.1 чисто в условном виде показано разбиение пространства  $L^2(\mathbb{R})$  на систему вложенных подпространств. Сигнал внутри каждого подпространства  $V_m$  (ограниченного на рисунке окружностью) представляет собой произвольный вектор числовых значений  $v_m(t) \in V_m$ . Все подпространства не пересекаются и вложены друг в друга таким образом, что объем каждого последующего (меньшего по номеру) подпространства  $V_{m-1}$ , вложенного в

подпространство  $V_m$ , при  $a=2$  в два раза меньше объема подпространства  $V_m$ , при этом четные числовые отсчеты вектора  $v_m(t) \in V_m$  входят в подпространство  $V_{m-1}$ , а нечетные отсчеты остаются в подпространстве, ограниченном границами  $[V_{m-1}, V_m)$  (отмечены на рисунке индексами  $W_m$ ).

Все условия в совокупности позволяют разложить произвольный сигнал  $s(t) \in L^2(\mathbb{R})$  по подпространствам  $V_m$ , т.е. на множество последовательных разномасштабных и



ортогональных друг другу функций  $v_m(t) \in V_m$ , объединение которых дает исходный сигнал  $s(t)$ , или аппроксимирует сигнал с определенной точностью в зависимости от ограничения количества значений масштабирующего коэффициента  $m$  (и, соответственно, количества подпространств  $V_m$ ). Функции  $v_m(t)$  являются ортогональными проекциями сигнала  $s(t)$  на подпространства  $V_m$ . Отсюда появляется возможность анализа функции или сигнала на различных уровнях разрешения, или масштаба. Переменная  $m$  называется масштабным коэффициентом, или уровнем анализа. Если значение  $m$  мало, то функция  $v_m(t)$  есть грубая аппроксимация  $s(t)$ , в которой отсутствуют детали. При увеличении значений  $m$  точность аппроксимации повышается.

Кратность КМА, равную 2, в принципе, можно заменить любым целым числом, большим 1, но использование двоичной кратности оптимально и позволяет использовать быстрое вейвлет-преобразование. Выбор функции  $\varphi_0$  также произволен, но желательно стремиться к тому, чтобы спектр функций подпространства  $V_0$  был сконцентрирован в интервале  $(-\pi, \pi)$ .

**Масштабирующая функция.** Для того чтобы задать КМА, достаточно знать только одно из подпространств  $V_m$ , остальные определяются уравнением (3.2.5). Допустим, что это подпространство  $V_0$ , состоящее из сигналов, заданных "с разрешением 1". Тогда в пространстве  $V_m$  задаются сигналы с разрешением  $2^m$ , и оно отличается от  $V_0$  только перемасштабированием базисной функции в соответствии с (3.2.5). Так, если пространство  $V_0$  имеет скейлинг-функцию  $\varphi_0(t)$ , то соответствующее уравнение для скейлинг-функции  $\varphi_1(t)$  пространства  $V_1$  определяется выражением  $\sqrt{2} \varphi(2t-k)$ .

Поскольку  $V_0 \subset V_1$ , то функцию  $\varphi_0(t)$  можно представить линейной комбинацией сдвигов функции  $\varphi_1(t)$  (с учетом ее более компактного носителя) с определенными весовыми коэффициентами перемасштабирования  $h_k$ . Так, для скейлинг-функции Хаара, имеющей прямоугольное окно, каждая функция  $\varphi_{m-1}(t)$  образуется суммой двух последовательных функций  $\varphi_m(t)$  с соответствующим коэффициентом для сохранения единичной нормы. В общем случае, носитель функции может иметь произвольный размер с числом отсчетов  $2M$  (в единицах  $k$ ), при этом уравнение линейной связи базисных функций пространств, которое обычно называют функциональным уравнением масштабирования (уравнение рескейлинга), записывается в следующем виде:

$$\varphi(t/2) = \sqrt{2} \sum_{k=0}^{2M-1} h_k \varphi(t-k), \quad (3.2.6)$$

или, в эквивалентной форме:

$$\varphi(t) = \sqrt{2} \sum_{k=0}^{2M-1} h_k \varphi(2t-k). \quad (3.2.6')$$

Это уравнение называется масштабирующим. Решение этого уравнения и дает скейлинг-функцию, которую иногда называют "отцовским" вейвлетом. Уравнение масштабирования может иметь и несколько иные формы записи. В частности, в пакете Wavelet Toolbox Matlab оно используется в виде:

$$0.5\varphi(t/2) = \sum_k w_k \varphi(t-k), \quad (3.2.6'')$$

где вместо коэффициентов  $h_k$  используются коэффициенты  $w_k = h_k / \sqrt{2}$  и более удобная для реализации на ПК нормировка вейвлетов.

Значения  $h_k$  определяются из условия для ортонормальных базисов:

$$h_k = \sqrt{2} \int_t \varphi(t) \varphi^*(2t-k) dt. \quad (3.2.7)$$

При дискретных значениях параметров сдвига масштабирующий вейвлет также дискретен и при задании функции  $\varphi(t)$  на конечном интервале имеет конечное число

коэффициентов  $h_k$ , отличных от нуля. Условие нормировки масштабирующих коэффициентов:

$$\int_{-\infty}^{\infty} \varphi(t) dt = 1, \quad (3.2.8)$$

откуда следует:

$$\sum_k (h_k)^2 = 1. \quad (3.2.9)$$

Простейший и самый короткий вейвлет соответствует  $M=1$ . Так, для скейлинг-функции Наара при определении коэффициентов  $h_k$  уравнение рескейлинга (3.2.6) отображает растяжение вдвое исходного интервала более высокого уровня, число ненулевых коэффициентов  $h_k$  равно  $2M = 2$ , при этом, с учетом коэффициента нормировки  $\sqrt{2}$ , значения коэффициентов  $h_0 = h_1 = 1/\sqrt{2}$ . Подставляя значения коэффициентов в (3.2.6), получаем:

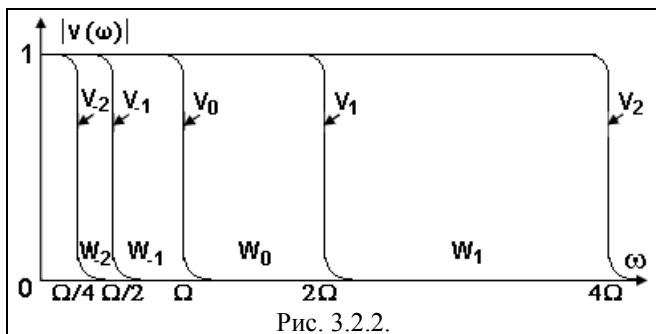
$$\varphi(t/2) = \varphi(t) + \varphi(t-1).$$

Решение этого функционального уравнения:

$$\varphi(t) = \theta(t) + \theta(1-t),$$

где  $\theta(t)$  – функция Хевисайда:  $\theta(t) = 1$  при  $t \geq 0$ ,  $\theta(t) = 0$  при  $t < 0$ .

Из совокупности исходных условий кратномасштабного анализа и уравнения (3.2.5) следует, что перевод сигнала из пространства  $V_{m+1}$  с более высоким разрешением в пространство  $V_m$ , по существу, представляет собой нормированную децимацию сигнала - двукратное прореживание, с соответствующим уменьшением в 2 раза числа отсчетов сигнала. Это эквивалентно низкочастотной фильтрации сигналов  $v_{m+1}(k) \in V_{m+1}$  оператором  $h_k$  с частотой среза, равной половине частоты Найквиста сигналов  $v_{m+1}(k)$ , с автоматическим сокращением (за счет прореживания) главного частотного диапазона децимированного сигнала  $v_m(k)$  в 2 раза.

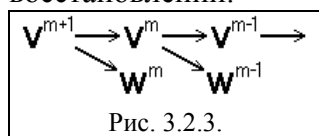


На рис. 3.2.2 приведено отображение такой операции в частотной области представления пространств  $V_m(\omega)$ , которое, в отличие от рис. 3.2.1, имеет вполне конкретный физический смысл разделения спектров сигналов (и самих сигналов при их восстановлении из спектров) на низкочастотную  $V_{m-1}(\omega)$  и высокочастотную  $W_{m-1}$  части.

Для исключения потерь высокочастотной информации, которая может потребоваться при восстановлении сигнала, она должна "переводиться" и сохраняться в новые подпространства  $W_m$ , ортогональные подпространствам  $V_m$ , такие, что

$$V_{m+1} = V_m \oplus W_m. \quad (3.2.10)$$

Подпространства  $W_m$  называются *детализирующими* в том смысле, что именно они содержат ту дополнительную информацию (не пересекающуюся с пространством  $V_m$ ), необходимую для повышения уровня разрешения сигнала с  $V_m$  на  $V_{m+1}$  при его восстановлении.



Этот процесс нагляднее рассматривать в обратном порядке, как это показано на рис. 3.2.3. Так как наиболее детальный уровень разрешения сигнала соответствует максимальным значениям  $m$  (при  $m \Rightarrow \infty$  сигнал становится непрерывным), то разложение сигнала по подпространствам начинается с максимальных

значений  $m$ . На каждом цикле разделения, при переходе из пространства  $V_{m+1}$  в пространство  $V_m$ , от пространства  $V_{m+1}$  отделяется подпространство  $W_m$ , в которое отфильтровывается высокочастотная информация пространства  $V_{m+1}$ , а остающаяся информация более "грубого" разрешения перемещается в пространство  $V_m$  и представляет собой аппроксимацию данных пространства  $V_{m+1}$ . В пределе, с учетом свойства ортогональности пространств:

$$V_{m+1} = \bigoplus_{m=-\infty}^m W_m, \quad \bigoplus_{m=-\infty}^{\infty} W_m = L^2(\mathbb{R}). \quad (3.2.11)$$

Таким образом, физический смысл процесса разложения пространств достаточно прост. Исходное пространство ( $V_{m+1}$  на рисунке) является пространством сигналов и функций с определенным частотным диапазоном. При разложении сигнала в пространство  $W_m$  отделяются высокочастотные составляющие пространства  $V_{m+1}$ , а в пространстве  $V_m$  остаются его низкочастотные составляющие.

**Базисный вейвлет.** Детализирующие подпространства  $W_m$  в совокупности также образуют взаимно ортогональный набор и имеют свой ортонормальный базис при любом заданном значении  $m$ . Если скейлинг-функция установлена, то базисная функция детализирующего пространства, которую называют вейвлетом ("материнским"), должна иметь определенную связь со скейлинг-функцией.

Уравнение (3.2.6), по существу, представляет собой нормированное уравнение свертки, где  $h_k$  представляет собой оператор низкочастотного фильтра:

$$\varphi(t/2) = \sqrt{2} h_k \otimes \varphi(t-k). \quad (3.2.12)$$

При переходе в частотную область:

$$\varphi(2\omega) = H(\omega) \varphi(\omega). \quad (3.2.13)$$

Уравнение (3.2.13) является масштабирующим уравнением в частотной области и полностью определяется  $2\pi$  периодической функцией  $H(\omega)$ . Отсюда следует, что в пространстве  $W_m$  должна сохраняться высокочастотная часть информации сигнала, что может выполняться квадратурным обращением низкочастотного фильтра:

$$G(\omega) = -\exp(-j\omega) H^*(\omega+\pi). \quad (3.2.14)$$

Фурье-образ искомого вейвлета:

$$\Psi(2\omega) = G(\omega) \varphi(\omega). \quad (3.2.15)$$

При переходе во временную область:

$$\psi(t/2) = \sqrt{2} g_k \otimes \varphi(t-k) = \sqrt{2} \sum_{k=0}^{2M-1} g_k \varphi(t-k). \quad (3.2.16)$$

При этом значения  $g_k$  могут быть вычислены и непосредственно во временной области соответствующим квадратурным обращением оператора  $h_k$ :

$$g_k = (-1)^k h_{2M-1-k}. \equiv (-1)^k (h_k)_{\text{rev}}. \quad (3.2.17)$$

где  $(h_k)_{\text{rev}}$  – реверсированный массив оператора  $h_k$ , записанный в обратном порядке. Соответственно, для вейвлета Наара эти коэффициенты равны:  $g_0 = 1/\sqrt{2}$ ,  $g_1 = -1/\sqrt{2}$ . Именно этот вейвлет и известен, как вейвлет Хаара (рис. 3.1.1). В функциональном анализе он применяется с 1910 года. Масштабированные и смещенные версии скейлинг-функции и вейвлета:

$$\varphi_{m,k} = 2^{m/2} \varphi(2^m t - k).$$

$$\psi_{m,k} = 2^{m/2} \psi(2^m t - k).$$

Вейвлет Хаара знакопеременен, при этом

$$\int_{-\infty}^{\infty} \psi(t) dt = 0.$$

Условие знакопеременности является общим условием для всех "материнских"

вейвлетов, которое обеспечивает безусловную устойчивость базиса при восстановлении исходного сигнала.

**Разложение функций на вейвлетные ряды** на заданном уровне разрешения  $m$  выполняется по формуле:

$$s(t) = \sum_k C_{m,k} \varphi_{m,k} + \sum_{m,k} D_{m,k} \psi_{m,k}. \quad (3.2.18)$$

Значения коэффициентов (которые обычно называют суммами и разностями):

$$C_{m,k} = \int_t s(t) \varphi_{m,k}(t) dt. \quad (3.2.19)$$

$$D_{m,k} = \int_t s(t) \psi_{m,k}(t) dt. \quad (3.2.20)$$

На практике значения коэффициентов определяются с помощью быстрого вейвлет-преобразования, которое будет рассмотрено ниже.

Первая сумма в (3.2.18) содержит усредненные (с весовыми функциями  $\varphi_{m,k}$ ) значения функции  $s(t)$  по диадным интервалам  $[k \cdot 2^{-m}, (k+1) \cdot 2^{-m}]$ , вторая – значения флюктуаций на данных интервалах. По мере возрастания значения  $m$  длина интервалов уменьшается и уровень детализации (разрешения) функции  $s(t)$  увеличивается. На самом детальном уровне  $m = m_{\max} = M$  ряд представлен только скейлинг-функцией и, в пределах точности разложения, практически совпадает с исходной функцией:

$$s(t) = \sum_k C_{M,k} \varphi_{M,k}.$$

На низшем уровне разрешения (на наиболее широких интервалах) первая сумма ряда (3.2.18) содержит всего одно усредненное взвешенное значение сигнала, а вторая сумма показывает флюктуации на всех без исключения уровнях. Числовой ряд на каждом из уровней является "истинным" представлением сигнала с тем же объемом информации, но только в другом (вейвлетном) математическом представлении. Вейвлет-преобразованием сигнала очень часто называют полную комбинацию рядов только второй суммы (3.2.18), дающей представление о локальных особенностях и флюктуациях сигнала на всех уровнях разрешения, которые обычно и являются предметом изучения.

Таким образом, выражение (3.2.18) показывает возможность аппроксимации любой произвольной функции  $s(t)$  набором простых локальных функций  $\varphi_{m,k}(t)$  и  $\psi_{m,k}(t)$ , ортогональных на разных уровнях значений  $m$  и полностью покрывающих пространство  $L^2(\mathbb{R})$  за счет смещений  $k$ . Первая сумма выражения (3.2.18) дает "сглаженные средние" значения функции  $s(t)$  на разных масштабных уровнях, вторая сумма вейвлетных функций добавляет к "грубой" аппроксимации сигнала все более подробные детали на все меньших масштабных интервалах.

**Вычисление вейвлетных рядов.** Допустим, что сигнал  $s(t)$  известен с разрешением 1, т.е. задан числовым рядом  $s_n \equiv C_{m,n}$  в пространстве  $V_0$  с масштабом  $m=0$ , который можно рассматривать, как результат его разложения по сдвигам скейлинг-функции:

$$s(t) = \sum_n s_n \varphi(t-n).$$

Версия масштаба  $m-1$  ортогональной проекции  $s(t)$  на подпространство  $V_{-1}$  будет задаваться набором скалярных произведений  $s(t)$  с функциями из базиса  $V_{-1}$ :

$$C_{m-1,n} = \langle s(t), (1/\sqrt{2})\varphi(t/2-n) \rangle.$$

Из уравнения (3.2.6) и условий ортогональности следует:

$$C_{m-1,n} = \sum_s h_s C_{m,2n-s}.$$

Другими словами, вычисление вейвлетных коэффициентов аппроксимации масштаба  $m-1$  может осуществляться путем свертки коэффициентов предыдущего масштаба ( $m$ ) с низкочастотным фильтром  $h_s$  и прореживания вдвое, которое «встроено» в эту формулу через индекс  $2n-s$ .

В качестве деталей сигнала  $s(t)$ , исчезающих при переходе к новому масштабу  $m-1$ ,

следует взять компоненту  $s(t)$ , ортогональную к сигналам масштаба  $m-1$  пространства  $V_{-1}$ , и спроецировать ее на базис пространства детализирующих коэффициентов  $W_{-1}$ :

$$D_{m-1,n} = \langle s(t), (1/\sqrt{2})\psi(t/2-n) \rangle.$$

Или, с использованием уравнения (3.2.16):

$$D_{m-1,n} = \sum_s g_s C_{m,2n-s}.$$

Т.е. действует та же схема свертки аппроксимационных коэффициентов предыдущего масштаба с оператором высокочастотного фильтра и прореживания вдвое.

Схемы последовательного масштабного разложения сигналов действуют на любых масштабах. Эту процедуру вычисления вейвлетных рядов называют быстрым вейвлет-преобразованием (БВП) или алгоритмом Малла по фамилии его автора.

### 3.3. БЫСТРОЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ

**Принцип преобразования.** Любую функцию  $s(t)$  можно рассматривать на любом  $m'$ -уровне разрешения. Для разделения функции на этом уровне между ее усредненными значениями и флюктуациями вокруг средних значений преобразуем формулу 3.3.1 к следующему виду:

$$s(t) = \sum_{k=-\infty}^{\infty} C_{m',k} \varphi_{m',k}(t) + \sum_{m=m'}^{\infty} \sum_{k=-\infty}^{\infty} D_{m,k} \psi_{m,k}(t). \quad (3.3.1)$$

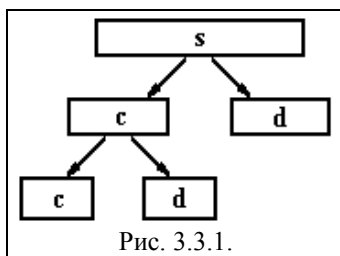
При бесконечных пределах первая сумма в этом выражении стремится к нулю и может быть опущена, давая "чистое" вейвлет-преобразование. В общем случае коэффициенты  $C_{m,k}$  и  $D_{m,k}$  можно вычислять непосредственно по формулам (3.2.19) и (3.2.20). На практике мы обычно имеем дело с цифровыми данными в виде конечного набора отсчетов, а, соответственно, наилучший уровень разрешения определен интервалом, содержащим один отсчет, и суммирование выполняется в конечных пределах. Значение  $m=0$  обычно принимается для этого наилучшего уровня разрешения. Для принятой нами формы вейвлетов  $\varphi_{m,k} = 2^{m/2} \varphi(2^m t - k)$  усреднение отсчетов (расширение размеров вейвлетов) происходит при уменьшении значений  $m$ , т.е. при  $m = 0, -1, -2, \dots$ . Для исключения использования отрицательных индексов масштабирования знак "минус" обычно вводится непосредственно в функции вейвлетов, т.е.  $\varphi_{m,k} = 2^{-m/2} \varphi(2^{-m} t - k)$ , при этом вейвлет-коэффициенты вычисляются для  $m > 0$ .

**Алгоритм Малла.** Кратномасштабный анализ при последовательном увеличении значений  $m$  приводит к естественной форме быстрых итерационных вычислений:

$$C_{m+1,k} = \sum_n h_n C_{m,2k+n}, \quad (3.3.2)$$

$$D_{m+1,k} = \sum_n g_n C_{m,2k+n}, \quad (3.3.3)$$

$$C_{0,k} = \int_{k\Delta t}^{(k+1)\Delta t} s(t) \varphi(t-k) dt. \quad (3.3.4)$$



Уравнения обеспечивают пирамидальный алгоритм вычисления вейвлет-коэффициентов (алгоритм Малла), приведенный на рис. 3.3.1. Явный вид вейвлета требуется только для расчета коэффициентов  $h_n$  и  $g_n$ , при самом преобразовании используются значения коэффициентов  $h_n$  и  $g_n$ . Уравнение (3.3.4) применяется при известной аналитической форме функции  $s(t)$ . Для цифровых данных в качестве значений

$C_{0,k}$  принимаются исходные значения данных, т.е.  $C_{0,k} = s(k)$ .

Сущность операций, выполняемых формулами (3.3.2) и (3.3.3), заключается в следующем. На первом этапе преобразования цифровой фильтр  $h_n$  из сигнала  $s_k = C_{0,k}$  выделяет низкие частоты  $|\omega| \leq \pi/2$ , а октавный фильтр  $g_n$  выделяет верхние частоты  $\pi/2 \leq |\omega| \leq \pi$ . Поскольку на выходе фильтра  $h_n$  отсутствует верхняя половина частот, то частота

дискретизации выходного сигнала может быть уменьшена в 2 раза, т.е. выполнена децимация выходного сигнала, что производится в формуле (3.3.2) сдвигами  $(2k+n)$  через 2 отсчета по входному сигналу. На выходе фильтра  $g_n$  освобождается место в области низких частот, и аналогичное прореживание выходного сигнала приводит к транспонированию верхних частот на освободившееся место. Таким образом, каждый из выходных сигналов несет информацию о своей половине частот, при этом выходная информация представлена таким же количеством отсчетов, что и входная.

**Реконструкция сигналов.** Поскольку в формулах (3.3.2, 3.3.3) вместо базисных функций используются фильтры, то обратные преобразования, т.е. последовательную сборку сигнала от больших  $m$  к малым и реконструкцию сигналов по значениям его вейвлет-коэффициентов с любого уровня разрешения, имеет смысл также выразить через фильтры реконструкции:

$$C_{m-1} = \sum_{n \in I} C_{m,n} h_{k-2n}^r + \sum_{n \in I} D_{m,n} g_{k-2n}^r, \quad (3.3.5)$$

Алгоритм вычислений по (3.3.5) обратен алгоритму декомпозиции, т.е. представляет собой аппроксимацию коэффициентов  $C_m$  и  $D_m$  на новый, в 2 раза меньший, шаг дискретизации с двукратным увеличением частоты Найквиста и восстановлением спектра коэффициентов  $C_m$  в низкочастотную часть нового главного диапазона спектра  $C_{m-1}$ , а спектра коэффициентов  $D_m$  в высокочастотную часть спектра  $C_{m-1}$ . Это выполняется расстановкой нулевых значений между коэффициентами  $C_m$  и  $D_m$  (увеличение в 2 раза числа отсчетов), фильтрацией полученных массивов низкочастотным  $h^r(k)$  и высокочастотным  $g^r(k)$  фильтрами реконструкции, и сложением результатов фильтрации. Модули частотных характеристик фильтров  $h^r(k)$  и  $g^r(k)$  должны повторять модули частотных характеристик фильтров  $h(k)$  и  $g(k)$ . Но фильтры декомпозиции  $h(k)$  и  $g(k)$  являются односторонними и фазосдвигающими, и при реконструкции коэффициентов  $C_{m-1}$  этот сдвиг фазы должен ликвидироваться. Последнее достигается реверсом значений коэффициентов фильтров декомпозиции, т.е.:

$$h^r(k) = \text{reverse}(h(k)), \quad g^r(k) = \text{reverse}(g(k)). \quad (3.3.6)$$

Точность реконструкции сигналов зависит от потерь информации при выполнении прореживания спектров, причем эти потери наблюдаются на срезах полос пропускания фильтров низких и высоких частот, крутизна которых зависит от порядка фильтров, их согласованности, и типа вейвлетных функций.

Обязательным условием преобразования сигнала является его задание количеством точек (отсчетов), равном  $N=2^m$ , где значение  $m \geq 1$  определяет максимально возможное число уровней декомпозиции сигнала при целочисленных значениях кратности сдвигов операторов фильтров количеству отсчетов вейвлетных коэффициентов на каждом уровне декомпозиции. Для выполнения этого условия количество отсчетов сигнала, как правило, дополняется до ближайшего большего значения  $N$  методами, известными из практики задания начальных/конечных условий свертки (нулями, конечными значениями сигналов, четными или нечетными значениями относительно конечных отсчетов, периодическим продолжением и т.п.). Может применяться также передискретизация исходного сигнала до необходимого количества отсчетов.

**Пакетные вейвлеты.** Основная информация обычно заключена в низкочастотной части сигнала, разложение которой может быть продолжено вплоть до нулевого уровня. Но аналогичная операция может применяться и к любой высокочастотной части разложения. Это соответствует замене вейвлета  $\psi(t)$  на два новых вейвлета

$$\psi_1(t) = \sum_n h_n \psi(t-n), \quad \psi_2(t) = \sum_n g_n \psi(t-n),$$

которые тоже локализованы в пространстве, но на вдвое более широком интервале, чем исходный вейвлет. Бинарное дерево разложения (рис. 3.3.1) "расщепляется" и для коэффициентов 'D' любого уровня. Такое расщепление является адаптивным и легко

приспосабливается к индивидуальным особенностям сигналов. Функции адаптивного преобразования называют вейвлет-пакетом.

### 3.4. ФИЛЬТРЫ ДУАЛЬНОЙ ДЕКОМПОЗИЦИИ И РЕКОНСТРУКЦИИ СИГНАЛОВ

Рассмотренные математические основы дуального вейвлет-преобразования показывают, что основную роль в реализации вейвлетных преобразований играют низкочастотные и высокочастотные фильтры декомпозиции и реконструкции сигналов.

**Идеальные фильтры.** Преобразование Фурье произвольной числовой последовательности

$\{s_k\}$  является  $2\pi$  - периодической функцией и определяется числовыми значениями на главном частотном диапазоне  $[-\pi, \pi]$ . При этом полагается, что шаг дискретизации данных  $\Delta t=1$ , а частота Найквиста сигнала  $s_k$  равна  $\omega_N = \pi/\Delta t = \pi$ .

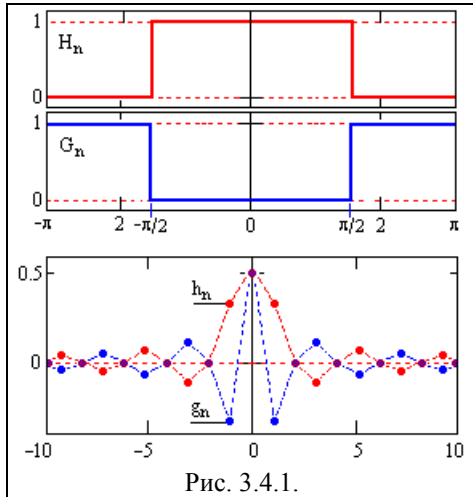


Рис. 3.4.1.

Передаточная функция  $H(\omega)$  низкочастотного фильтра  $h_n$ , удовлетворяющего условию (3.2.7), концентрируется в интервале  $[-\pi/2, \pi/2]$ . При разделении сигнала на два частотных поддиапазона с полным сохранением исходной информации должно выполняться условие (рис. 3.4.2):

$$H(\omega) + G(\omega) = 1. \quad (3.4.1)$$

Отсюда следует, что передаточная функция  $G(\omega)$  высокочастотного фильтра  $g_n$ , сосредоточена на  $[-\pi, -\pi/2]$  и  $[\pi/2, \pi]$ . Соответственно, идеальные фильтры  $H(\omega)$  и  $G(\omega)$  в пределах главного частотного диапазона задаются выражениями:

$$H(\omega) = \begin{cases} 1, & \omega \in [-\pi/2, \pi/2] \\ 0, & \omega \notin [-\pi/2, \pi/2] \end{cases}, \quad G(\omega) = \begin{cases} 0, & \omega \in [-\pi/2, \pi/2] \\ 1, & \omega \notin [-\pi/2, \pi/2] \end{cases}. \quad (3.4.2)$$

Коэффициенты фильтров (обратное преобразование Фурье, рис. 3.4.2):

$$h_0 = 0.5; \quad h_{2k} = 0, \quad k \neq 0; \quad h_{2k+1} = (-1)^k / (\pi(2k+1)); \quad k=0, \pm 1, \pm 2, \pm 3, \dots \quad (3.4.3)$$

$$g_0 = 0.5; \quad g_{2k} = 0, \quad k \neq 0; \quad g_{2k+1} = (-1)^{k+1} / (\pi(2k+1)); \quad k=0, \pm 1, \pm 2, \pm 3, \dots \quad (3.4.4)$$

Связь значений коэффициентов:

$$g_n = (-1)^n h_n. \quad (3.4.5)$$

Разложение сигнала  $s(k)$  на низкочастотную и высокочастотную части в спектральной и временной области:

$$S(\omega) = H(\omega)S(\omega) + G(\omega)S(\omega) = S_h(\omega) + S_g(\omega). \quad (3.4.6)$$

$$s(k) = h(n) \otimes s(k-n) + h_g(n) \otimes s(k-n) = s_h(k) + s_g(k). \quad (3.4.7)$$

Так как носитель функции  $S_h(\omega)$  находится на интервале  $[-\pi/2, \pi/2]$ , то  $S_h(\omega)$  можно разложить в ряд Фурье, как  $\pi$  - периодическую функцию с частотой Найквиста  $\pi/2$ :

$$S_{h2\downarrow}(\omega) \leftrightarrow 2 \sum_k s_h(2k) \exp(-j 2k\omega), \quad \omega \in [-\pi/2, \pi/2], \quad (3.4.8)$$

т.е. функция  $s_h(k)$  избыточна по количеству отсчетов и может быть децимирована. Соответственно, не требуется вычислять свертку по всем нечетным значениям  $s(k)$ . Двукратная децимация обычно обозначается индексом  $2\downarrow$ :

$$s_{h2\downarrow}(m) = s_h(2k) = h(n) \otimes s(2k-n) = C(m),$$

где  $m$  – последовательная нумерация четных отсчетов  $s_h(2k)$  ( $m = \text{int}(k/2)$ ).

Учитывая периодичность частотных функций, спектр  $S_g(\omega)$  можно рассматривать на интервале  $[0, 2\pi]$ , где ненулевые отсчеты  $S_g(\omega)$  находятся на интервале  $[\pi/2, 3\pi/2]$ . При аналогичном разложении  $S_g(\omega)$  в этом интервале, как  $\pi$  - периодической функции:

$$S_{g2\downarrow}(\omega) \leftrightarrow 2 \sum_k s_g(2k) \exp(-j 2k\omega), \quad \omega \in [\pi/2, 3\pi/2], \quad (3.4.9)$$

$$s_{g2\downarrow}(m) = s_g(2k) = g(n) \otimes s(2k-n) = D(m).$$

Для обратного преобразования спектров  $S_{h2\downarrow}(\omega)$  и  $S_{g2\downarrow}(\omega)$  в спектры  $S_h(\omega)$  и  $S_g(\omega)$  главный диапазон спектров нужно увеличить в 2 раза дополнением нулями. Во временной области эта операция может быть выполнена передискретизацией значений  $s_h(2k)$  и  $s_g(2k)$  с шага  $2\Delta k$  на шаг  $\Delta k$  рядом Котельникова-Шеннона. Альтернативная более быстрая операция - обратная децимация массивов  $C(m) \rightarrow C_2\uparrow(k)$  и  $D(m) \rightarrow D_2\uparrow(k)$ , которая выполняется дополнением массивов нулями между всеми отсчетами (обозначается индексом  $2\uparrow$ ), с последующей фильтрацией фильтрами  $2h(n)$  и  $2g(n)$ :

$$s_h(k) = 2h(n) \otimes C_2\uparrow(k-n), \quad s_g(k) = 2g(n) \otimes D_2\uparrow(k-n), \quad (3.4.10)$$

что обеспечивает восстановление исходного сигнала:

$$s(k) = s_h(k) + s_g(k). \quad (3.4.11)$$

**Реальные фильтры.** Операторы идеальных фильтров, заданные в частотной области прямоугольными импульсами (3.4.2), имеют бесконечные импульсные характеристики (3.4.3, 3.4.4) и убывают достаточно медленно. При усечении таких операторов на частотной характеристике проявляется явление Гиббса, что увеличивает погрешности декомпозиции и реконструкции сигналов. С практической точки зрения целесообразно использовать фильтры с плавным переходом от полосы пропускания в полосу подавления, которые имеют конечное число ненулевых коэффициентов. При задании таких низкочастотных  $H(\omega)$  и высокочастотных  $G(\omega)$  фильтров, удовлетворяющих условию (3.4.1), разложение сигнала с децимацией остается без изменений. Добавляя к операторам фильтров декомпозиции индекс  $d$ , получаем:

$$\begin{aligned} s(k) \rightarrow s_h(k) &= h_d(n) \otimes s(k-n) \rightarrow 2\downarrow \rightarrow C(m) = s_h(2k), \\ s(k) \rightarrow s_g(k) &= g_d(n) \otimes s(k-n) \rightarrow 2\downarrow \rightarrow D(m) = s_g(2k). \end{aligned} \quad (3.4.12)$$

Однако точное восстановление сигналов по формулам (3.4.10, 3.4.11) возможно только для взаимно ортогональных фильтров. Для ограниченных перекрывающихся по спектру фильтров для постановления сигналов необходимы парные к ним фильтры реконструкции, компенсирующие возможные искажения восстановления. Для упрощения выражений для числовых рядов перейдем в  $z$ -область представления сигналов.

$$H_d(z) = \sum_n h_d(n) z^n, \quad G_d(z) = \sum_n g_d(n) z^n, \quad S(z) = \sum_n s_d(n) z^n,$$

где  $z = \exp(-j\omega)$  – комплексная переменная.

Отфильтрованные низкочастотный и высокочастотный сигналы:

$$C_d(z) = H_d(z) S(z), \quad D_d(z) = G_d(z) S(z). \quad (3.4.13)$$

Децимация сигналов в  $z$ -области выполняется простыми выражениями:

$$C(z^2) = 0.5 (C_d(z) + C_d(-z)), \quad D(z^2) = 0.5 (D_d(z) + D_d(-z)). \quad (3.4.14)$$

Обозначим фильтры реконструкции сигналов индексами  $r$ . Уравнение реконструкции:

$$S(z) = 2[H_r(z) C(z^2) + G_r(z) D(z^2)]. \quad (3.4.15)$$

Подставляя в это выражение функции (3.4.14) и (3.4.13), получаем:

$$S(z) = [H_r(z)H_d(z)+G_r(z)G_d(z)]S(z) + [H_r(z)H_d(-z)+G_r(z)G_d(-z)]S(-z). \quad (3.4.16)$$

Отсюда следует, что искомые фильтры должны удовлетворять системе уравнений:

$$\begin{aligned} H_r(z)H_d(z)+G_r(z)G_d(z) &= 1, \\ H_r(z)H_d(-z)+G_r(z)G_d(-z) &= 0. \end{aligned} \quad (3.4.17)$$

В матричной форме:

$$\begin{pmatrix} H_d(z) & G_d(z) \\ H_d(-z) & G_d(-z) \end{pmatrix} \begin{pmatrix} H_r \\ G_r \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (3.4.17')$$

Решение системы существует, если определитель отличен от нуля всюду на единичной окружности  $z=\exp(-j\omega)$ :



$$H_d(z)C_d(-z) - H_d(-z)G_d(z) \neq 0.$$

**Сопряженные квадратурные фильтры.** Дополнительной практической задачей при определении комплекта фильтров декомпозиции и реконструкции сигналов является задание такого фильтра  $H_d(z)$  с соответствующими коэффициентами оператора  $h_d(n)$ , из которого следует определение всех других фильтров в форме:

$$G_d(z) = -z^{-1} H_d(-z^{-1}), \quad H_r(z) = H_d(z^{-1}), \quad G_r(z) = C_d(z) = -z H(-z), \quad (3.4.18)$$

при этом выполняется второе условие (3.4.7), а первое принимает вид:

$$H_d(z^{-1})H_d(z) + H_d(-z)H_d(-z^{-1}) = 1. \quad (3.4.19)$$

Поскольку коэффициенты  $h_d(n)$  вещественные, то при значениях  $z = \exp(-j\omega)$  и  $-z = \exp(-j\omega) = \exp(-j(\omega + \pi))$ , имеем:

$$H(z^{-1}) = H^*(\omega), \quad H(-z) = H(\omega + \pi).$$

При этом условие (3.4.19) и функции (3.4.18) принимают вид:

$$(H_d(z))^2 + (H_d(\omega + \pi))^2 = 1. \quad (3.4.20)$$

$$G_d(\omega) = -\exp(j\omega) H_d^*(\omega + \pi), \quad H_r(\omega) = H_d^*(\omega), \quad G_r(\omega) = -\exp(-j\omega) H_d(\omega + \pi). \quad (3.4.21)$$

Тем самым система (3.4.17) сводится к одному уравнению (3.4.20).

### 3.5. ОРТОГОНАЛЬНЫЕ И БИОРТОГОНАЛЬНЫЕ ВЕЙВЛЕТЫ

**Коэффициенты вейвлета.** Значения коэффициентов  $h_k$  и  $g_k$  в рамках КМА определяются на основании общих свойств скейлинг-функций и вейвлетов. Уравнения функций:

$$\varphi(t) = \sqrt{2} \sum_{k \in I} h_k \varphi(2t - k), \quad \psi(t) = \sqrt{2} \sum_{k \in I} g_k \varphi(2t - k).$$

Из свойства ортогональности масштабных функций следует первое уравнение на значения коэффициентов  $h_k$ :

$$\int_t \varphi(t) \varphi(t - x) dt = \delta_x, \quad \sum_k h_k h_{k+2x} = \delta_x. \quad (3.5.1)$$

Из условий нормировки скейлинг-функции следует второе уравнение:

$$\int_t \varphi(t) dt = 1, \quad \sum_k h_k = \sqrt{2}. \quad (3.5.2)$$

Из ортогональности вейвлетных и масштабных функций следует уравнение, решением которого являются значения коэффициентов  $g_k$ :

$$\int_t \psi(t) \varphi(t - n) dt = 0, \quad \sum_k h_k g_{k+2x} = 0. \quad (3.5.3)$$

$$g_k = (-1)^k h_{2M-1-k}. \quad (3.5.4)$$

Точность и масштабная разрешающая способность аппроксимации анализируемых функций вейвлетами зависит от их гладкости, т.е. от порядка дифференцируемости. При использовании вейвлет-преобразования для сжатия информации (отбрасыванием малозначимых коэффициентов разложения) вейвлеты с высокой гладкостью обеспечивают более точную реконструкцию сигналов. Для обеспечения знакопеременности и заданной гладкости до степени  $M-1$  вейвлеты должны быть ортогональны полиномам соответствующих степеней:

$$\int_t t^m \psi(t) dt = 0, \quad m = 0, 1, \dots, M-1, \quad \sum_k k^m g_k = 0. \quad (3.5.5)$$

$$\sum_k (-1)^k k^m h_k = 0. \quad (3.5.6)$$

**Пример расчета** коэффициентов выполним при  $M=2$ .

Запишем уравнения (3.5.1, 2 и 6) в явном виде:

$$\begin{aligned} h_0 h_2 + h_1 h_3 &= 0, & h_0 + h_1 + h_2 + h_3 &= \sqrt{2}, \\ h_0 - h_1 + h_2 - h_3 &= 0, & -h_1 + 2h_2 - 3h_3 &= 0. \end{aligned}$$

Решение этой системы уравнений:

$$\begin{aligned} h_0 &= 2^{-3/2} - h_3, & h_1 &= 2^{-1/2} - h_3, \\ h_2 &= 2^{-3/2} + h_3, & h_3 &= 2^{-5/2} (1 \pm \sqrt{3}). \end{aligned} \quad (3.5.7)$$

Примем для коэффициента  $h_3$  знак минус в скобках, при этом:

$$\begin{aligned} h_0 &= 2^{-5/2} (1 + \sqrt{3}) = 0.483, & h_1 &= 2^{-5/2} (3 + \sqrt{3}) = 0.837, \\ h_2 &= 2^{-5/2} (3 - \sqrt{3}) = 0.224, & h_3 &= 2^{-5/2} (1 - \sqrt{3}) = -0.129. \end{aligned}$$

Соответственно, значения коэффициентов  $g_k$ , вычисленные по (3.5.4):

$$g_0 = -0.129, \quad g_1 = -0.224, \quad g_2 = 0.837, \quad g_3 = -0.483.$$

Спектры коэффициентов  $h_k$  и  $g_k$  приведены на рис. 3.5.1.

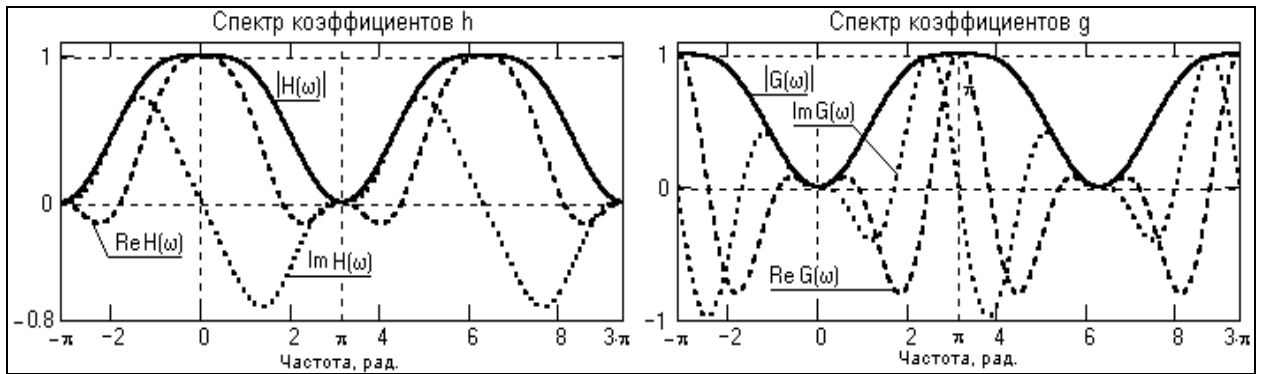


Рис. 3.5.1.

Из рисунка 3.5.1 можно видеть, что спектры коэффициенты  $h_k$  и  $g_k$  представляют собой передаточные функции односторонних согласованных цифровых фильтров, низкочастотного и высокочастотного соответственно.

Выбор знака в (3.5.7) для  $h_3$  несколько изменяет форму вейвлета, особенно для вейвлетов более высокого порядка. При повышении порядка гладкость вейвлета повышается и, как правило, несколько увеличивается область его определения  $(2M-1)$ .

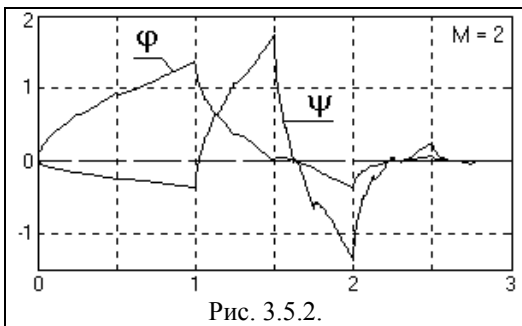


Рис. 3.5.2.

вейвлет второго порядка  $D_2$  из семейства ортонормальных вейвлетов Добеши, приведенный на рис. 3.5.2. Область задания вейвлетов Добеши шире, чем вейвлетов Хаара, но при этом они обеспечивают при вейвлет-преобразовании большее количество малозначимых коэффициентов разложения и, при отбрасывании последних, более сильное сжатие данных.

**Вейвлет Добеши.** Существуют семейства ортогональных вейвлетов, которые вообще не имеют аналитического выражения и определяются только фильтрами. К ним относятся вейвлеты Добеши. Скейлинг-функции и вейвлеты Добеши – это непрерывные функции, не тождественные нулю и недифференцируемые на конечном отрезке. Вычисленные выше при  $M=2$  коэффициенты  $h_k$  и  $g_k$  определяют простейший

ортонормальных вейвлетов Добеши, приведенный на рис. 3.5.2. Область задания вейвлетов Добеши шире, чем вейвлетов Хаара, но при этом они обеспечивают при вейвлет-преобразовании большее количество малозначимых коэффициентов разложения и, при отбрасывании последних, более сильное сжатие данных.

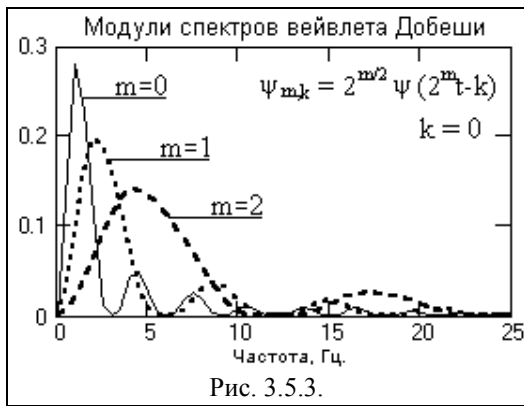


Рис. 3.5.3.

На рис. 3.5.3 приведены спектры вейвлета Добеши  $D_2$  при  $k=0$  в трех последовательных масштабах, отображающих частотную локализацию вейвлета и сдвиг по частоте при изменении масштабирования.

Вид скейлинг-функции и вейвлета Добеши вычисляются методом итераций (можно посмотреть в Matlab), а в практических приложениях используются только вейвлет-

коэффициенты  $h_k$  и  $g_k$  без вычисления конкретной формы вейвлетов.

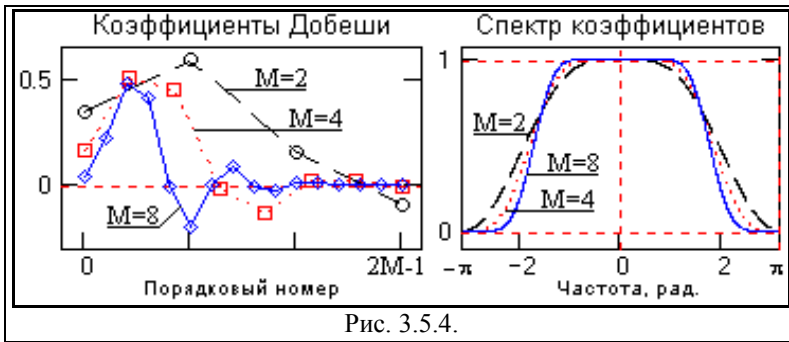


Рис. 3.5.4.

На рис. 3.5.4 приведены три вектора коэффициентов Добеши из системы Matlab (db2, db4 и db8) и спектры коэффициентов. Как видно из рисунка, по мере увеличения порядка векторов крутизна среза их частотных характеристик увеличивается, а, соответственно, качество разложения сигналов и их реконструкции также будут

улучшаться.

За исключением вейвлетов Хаара, все вещественные ортогональные вейвлеты асимметричны. Наиболее близкую к симметричной форму имеет вейвлет Коифлетса, форма которого приведена на рис. 3.5.5. Частотная характеристика коэффициентов  $h_k$  вейвлета по сравнению с коэффициентами Добеши такого же порядка имеет более высокую крутизну среза полосы пропускания.

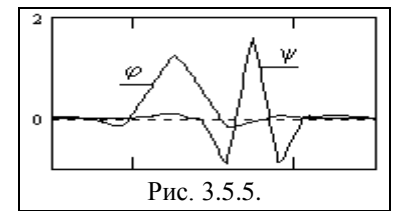


Рис. 3.5.5.

**Биортогональные вейвлеты.** Требование ортогональности вейвлетного базиса является сильным ограничением. Например, невозможно построить ортонормированный базис для вейвлетов с гладким симметричным носителем. Для обеспечения симметрии при точной реконструкции сигналов применяются биортогональные вейвлеты. Для вейвлетов  $\psi_{m,k}(t)$  с базисом Рисса можно определить дуальный вейвлет  $\psi_{m,k}^\#(t)$ , который образует биортогональную пару с вейвлетом  $\psi_{m,k}(t)$ , удовлетворяющую требованию ортогональности их скалярного произведения:

$$\langle \psi_{m,k}(t), \psi_{m,k}^\#(t) \rangle = \delta_{m,k}.$$

При использовании биортогональной пары декомпозиция сигналов может производиться вейвлетом  $\psi_{m,k}(t)$ , а реконструкция парным вейвлетом  $\psi_{m,k}^\#(t)$ , или наоборот. Разложение функций с биортогональными вейвлетами может производиться в двух эквивалентных формах:

$$s(t) = \sum_{m,k} \langle s(t), \psi_{m,k}^\#(t) \rangle \psi_{m,k}. \quad (3.5.8)$$

$$s(t) = \sum_{m,k} \langle s(t), \psi_{m,k}(t) \rangle \psi_{m,k}^\#. \quad (3.5.8')$$

Свойства регулярности биортогональных вейвлетов могут заметно отличаться. Если один из них обладает гладкостью порядка  $n$ , то дуальный ему вейвлет может иметь, по крайней мере,  $n$  нулевых моментов. Большое число нулевых моментов дает хорошие результаты при сжатии информации, а большая степень гладкости вейвлета обеспечивает более точную реконструкцию сигналов. При этом оба вейвлета можно выполнить симметричными.

Ортогональное преобразование можно рассматривать, как частный случай

биортогонального при  $\psi_{m,k}(t) = \psi_{m,k}^\#(t)$ . Для биортогонального преобразования действительны все свойства ортогонального. В частности, возможна реализация быстрого вейвлет-преобразования с применением операторов фильтров  $h_k$  и  $g_k$ . Отличие будет только в том, что при реконструкции сигналов по формуле (3.3.5) в ней должны применяться коэффициенты  $h_k^\#$  и  $g_k^\#$  парного вейвлета  $\psi_{m,k}^\#(t)$ , при этом должно выполняться условие (3.3.6) в следующем варианте:

$$h(z) h^\#(z^{-1}) + g(z) g^\#(z^{-1}) = 2, \quad h(z) h^\#(-z^{-1}) + g(z) g^\#(-z^{-1}) = 0. \quad (3.5.9)$$

При этом:

$$h^\#(z^{-1}) = \text{reverse}(h(z)), \quad g^\#(z^{-1}) = \text{reverse}(g(z)). \quad (3.5.10)$$

### 3.6. ДВУМЕРНЫЕ ВЕЙВЛЕТЫ

Многомерные вейвлет-преобразования являются расширением одномерных преобразований. Вейвлетными функциями такого преобразования являются тензорные произведения одномерных функций по размерности преобразования.

Двумерные вейвлеты определяются функциями двух переменных  $(x, y)$  в двумерном пространстве  $V(x, y) \in L^2(\mathbb{R}^2)$ , при этом параметры  $a$  и  $b$  могут быть индивидуальными для каждой переменной. В общей форме для двумерного непрерывного вейвлета:

$$\psi_{a1,b1; a2,b2}(x,y) = (a1 \cdot a2)^{-1/2} \psi_0[(x-b1)/a1, (y-b2)/a2]. \quad (3.6.1)$$

При двумерном КМА двумерный ортонормальный базис пространства  $L^2(\mathbb{R}^2)$  имеет четыре порождающих функции, которые строятся на основе одномерного ортонормального вейвлет-базиса  $\psi(t)$  и скейлинг-функции  $\phi(t)$ . Для скейлинг-функции:

$$\phi\phi(x, y) = \phi(x) \phi(y). \quad (3.6.2)$$

Тензорное произведение для вейвлетных функций:

$$\phi\psi(x, y) = \phi(x) \psi(y), \quad \psi\phi(x, y) = \psi(x) \phi(y), \quad \psi\psi(x, y) = \psi(x) \psi(y). \quad (3.6.3)$$

Если масштабирование по обеим переменным производится синхронно, то все остальные функции двумерных базисов определяются выражением:

$$\Phi_{m,k,n}(x, y) = 2^m \Phi(2^m x - k, 2^m y - n), \quad m, k, n \in I, \quad (3.6.4)$$

где функция  $\Phi(x,y)$  – соответствующие выражения (3.6.2 – 3.6.3).

Если система (3.6.3) является ортонормированным базисом в  $L^2(\mathbb{R}^2)$ , то прямое вейвлет-преобразование сигнала  $s(x, y)$  соответственно выполняется по формулам:

$$\begin{aligned} cc_{m,k,n} &= \langle s(x,y), \phi\Phi_{m,k,n}(x, y) \rangle, \quad cd_{m,k,n} = \langle s(x,y), \phi\psi_{m,k,n}(x, y) \rangle, \\ dc_{m,k,n} &= \langle s(x,y), \psi\Phi_{m,k,n}(x, y) \rangle, \quad dd_{m,k,n} = \langle s(x,y), \psi\psi_{m,k,n}(x, y) \rangle. \end{aligned}$$

При этом можно считать, что двумерный сигнал в плоскости  $(x, y)$  анализируется по горизонталям, вертикалям и диагоналям с одинаковым разрешением.

Обратное преобразование с уровня  $m$ :

$$\begin{aligned} s(x, y) &= \sum_{m,k,n} cc_{m,k,n} \phi\Phi_{m,k,n}(x, y) + \\ &+ \sum_{m,k,n} [cd_{m,k,n} \phi\psi_{m,k,n}(x, y) + dc_{m,k,n} \psi\Phi_{m,k,n}(x, y) + dd_{m,k,n} \psi\psi_{m,k,n}(x, y)]. \end{aligned}$$

При использовании быстрого преобразования дискретного сигнала, представленного матрицей  $s_{i,j}$ , которую (как и в одномерном случае) можно считать максимальным разрешением исходного непрерывного сигнала  $s(x,y)$ , преобразование начинается с фильтрации строк матрицы низкочастотным  $h_k$  и высокочастотным  $g_k$  фильтрами, аналогичными одномерному преобразованию, в результате которой вычисляются матрицы следующего уровня разложения НЧ и ВЧ. Полученные матрицы фильтруются этими же фильтрами по столбцам, в результате чего из НЧ формируются матрицы НЧНЧ и НЧВЧ, а из ВЧ - матрицы ВЧНЧ и ВЧВЧ. Таким образом, реализация двумерного БВП является четырехканальной, а общее количество отсчетов во всех четырех новых матрицах равно количеству отсчетов в исходной матрице разложения, что сохраняет в них полный объем информации, заключенный в исходной матрице.

В общем случае  $n$ -мерного пространства ортонормальный базис образуют  $2n-1$  функций, при помощи которых осуществляется МКА любой функции их  $L^2(\mathbb{R}^n)$  пространства, при этом нормировочный множитель равен  $2^{nm/2}$ .

## ТЕМА 22. НЕПРЕРЫВНОЕ И ДИАДНОЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ

В систему Mathcad программные средства для работы с вейвлетами были введены одними из первых, но в ограниченной форме – только для прямого и обратного вейвлет-преобразования временных рядов вейвлетами Добеши db4. Дальнейшее развитие программного обеспечения выполнялось в рамках специальных пакетов расширения (сначала Numerie Recipes, а затем Wavelets Extension), которые устанавливаются в систему обычным порядком.

В настоящей теме рассматриваются непрерывное вейвлет-преобразование, выполняемое типовыми средствами в системе Mathcad вейвлетами произвольной формы, задаваемой пользователем, а также простое диадное вейвлет-преобразование вейвлетом Добеши db4.

### 22.1. НЕПРЕРЫВНОЕ ВЕЙВЛЕТ - ПРЕОБРАЗОВАНИЕ.

Под непрерывным вейвлет-преобразованием (CWT), как правило, понимается преобразование сигналов и функций в аналитической форме. Однако при обработке физических данных на ЭВМ они представляются в дискретной форме - числовыми рядами результатов измерений. В принципе, если дискретизация выполнена корректно и максимальные частоты в данных не превышают частоты Найквиста, то всегда возможно восстановление аналоговой формы сигналов. Однако обработка данных в аналоговой форме аналитическими методами не имеет каких-либо преимуществ перед цифровыми методами. Практически, чаще используется замена аналитических методов дискретными аналогами с достаточно малым шагом дискретизации. Применительно к CWT, если нас интересует только анализ сигнала (без последующего его синтеза), то дискретизация может выполняться с любой частотой, даже не удовлетворяющей критерию Найквиста. Дискретный аналог CWT с достаточно малым шагом дискретизации данных по масштабу вейвлета и его сдвигам будем называть непрерывно-дискретным (CDWT).

**CWT типовыми средствами Mathcad.** Качественный анализ сигналов, особенно модельных сигналов на основе гладких финитных функций, удобно производить симметричными аналитическими вейвлетами. Так, например, базовый вейвлет МХАТ конструируется на основе второй производной функции Гаусса:

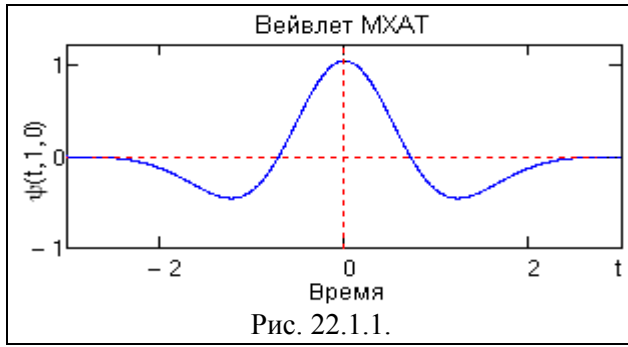
$$n := 2 \quad f(t) := (-1)^{n+1} \cdot \frac{d^n}{dt^n} \exp(-t^2) \rightarrow 2 \cdot e^{-t^2} - 4 \cdot t^2 \cdot e^{-t^2}$$

Отсюда:

$$\psi(t, a, b) := \frac{K}{\sqrt{a}} \cdot \exp\left[-\left(\frac{t-b}{a}\right)^2\right] \cdot \left[1 - 2 \cdot \left(\frac{t-b}{a}\right)^2\right], \quad (22.1.1)$$

где значения 'a' и 'b' – параметры масштаба вейвлета и сдвига, а значение коэффициента K определяется приведением нормы вейвлетной функции  $\|\psi(t,a,b)\|$  к 1 при нулевом среднем значении:

$$\int_{-\infty}^{\infty} \psi(t,a,b) dt \rightarrow 0, \quad \int_{-\infty}^{\infty} \psi(t,a,b)^2 dt \rightarrow 1.$$



Для МХАТ-вейвлета значение  $K = 1.0314$ . Форма базового вейвлета при  $a=1$  и  $b=0$  приведена на рис. 22.1.1. Аналогично могут задаваться вейвлеты нечетного типа и вейвлеты любой другой формы.

Для демонстрации СWT зададим модельный сигнал в виде функции Гаусса на интервале  $0-T$ ,  $T=100$  (рис. 22.1.2):

$$s(t) = 4 \exp(-(t-50)^2/10).$$

Формула прямого непрерывного

вейвлет-преобразования представляет собой скалярное произведение сигнала и вейвлета:

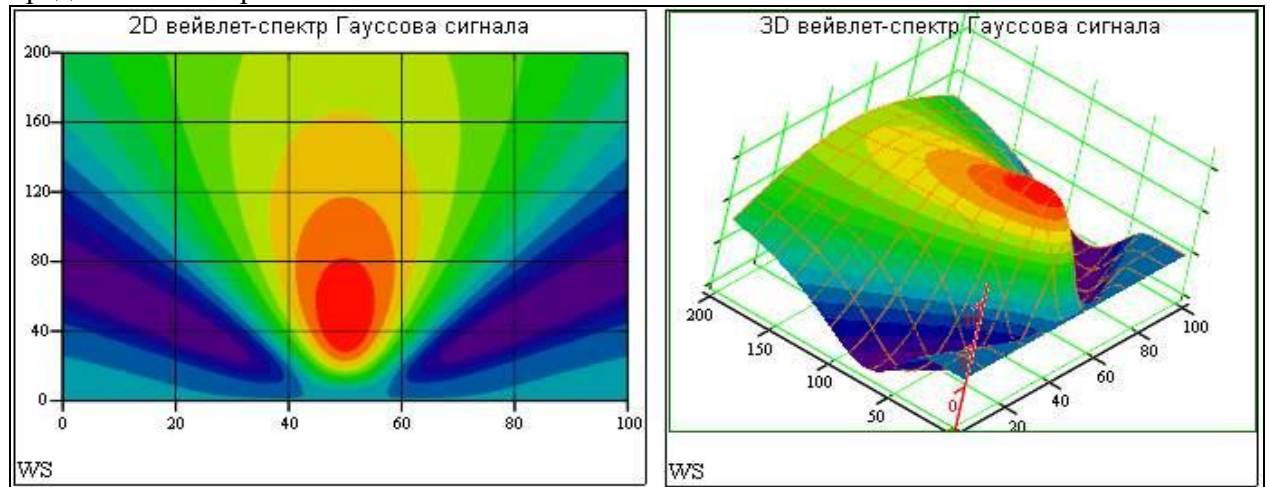
$$S(a,b) = \langle s(t), \psi(t,a,b) \rangle = \int_{-\infty}^{\infty} s(t) \psi(t,a,b) dt. \quad (22.1.2)$$

Если сигнал  $s(t)$  является финитным и выходит практически на нулевые значения по концам интервала задания, то пределы интегрирования в (22.1.2) можно задавать непосредственно по интервалу задания сигнала.

Для графического просмотра результатов преобразования следует задать шаг дискретизации функции  $S(a,b)$  по параметрам 'a', 'b' и выполнить дискретизацию. Например, для заданного интервала  $0-T$  функции  $s(t)$ :

$$\Delta b = 1, \quad i = 0..T, \quad \Delta a = 0.5, \quad j = 1..200, \quad WS_{i,j} = S(j\Delta a, i\Delta b).$$

Графическое представление вейвлетного спектра сигнала в двух вариантах представлено на рис. 22.1.3.



**Логарифмическая шкала масштабов.** Вейвлетный спектр по шкале масштабов (обратна частотной шкале) удобно представлять в логарифмической шкале. Переход на логарифмическую шкалу (с определенным множителем) можно выполнять непосредственно при дискретизации сигнала  $S(a,b)$ :

$$j := 0..25 \quad \Delta a := 0.25 \quad WSL_{1,j} := S[(\sqrt{2})^j \cdot \Delta a, i \cdot \Delta b]$$

Спектр сигнала в логарифмической шкале масштабов приведен на рис. 2.1.4.

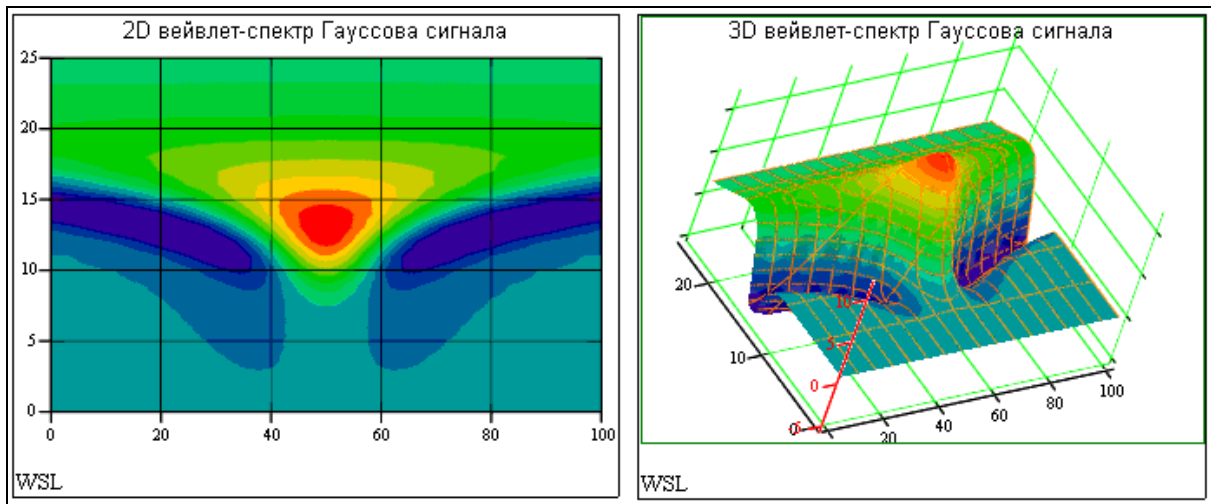


Рис. 22.1.4.

На рис. 22.1.5 приведен пример преобразования более сложного ЛЧМ сигнала.

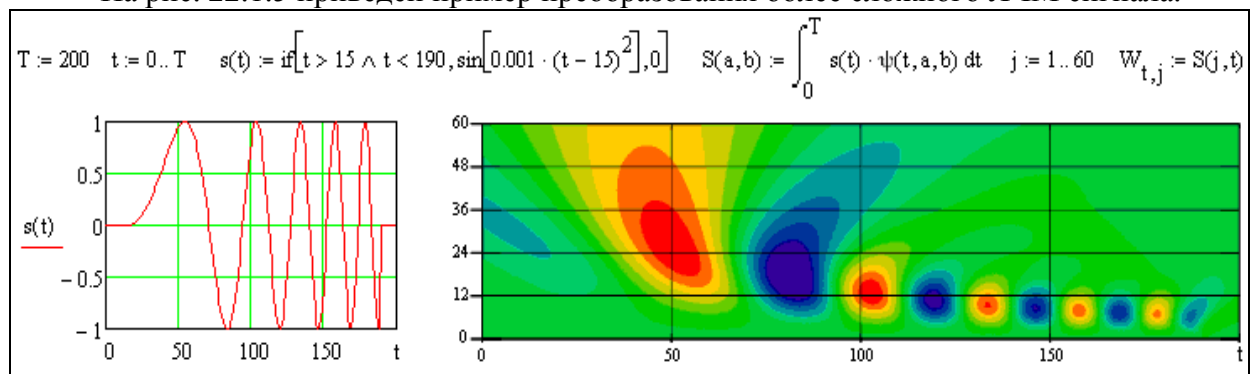


Рис. 22.1.5.

**Связь шкалы масштабов с частотой.** При практическом анализе частотных неоднородностей в сигнале масштабная шкала вейвлетного спектра несколько непривычна для визуального восприятия, но она всегда может быть заменена шкалой частот. Для перехода к шкале частот следует определить среднюю частоту вейвлета  $f_0$  на единичном масштабе ( $a=1$ ). Выполнить это можно непосредственно во временной области по максимуму функции взаимной корреляции:

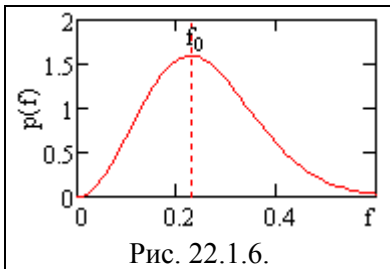


Рис. 22.1.6.

$$p(f) = \int_{-\infty}^{\infty} \psi(t, 1, 0) \cos(2\pi ft) dt.$$

После определения значения  $f_0$  (рис. 22.1.6) частотная шкала вычисляется по масштабной шкале трансформацией  $f_0/a \rightarrow f$ .

**CDWT типовыми средствами Mathcad.** При вычислении вейвлетных спектров в дискретной форме существенное значение для наглядности и выразительности спектров имеет шаг дискретизации параметров 'a' и 'b'. Шаг дискретизации  $\Delta b$ , как правило, принимается равным  $\Delta t$  анализируемых сигналов, т.е. равен 1, как и условное значение  $\Delta t$ , при этом временной масштаб вейвлетного спектра соответствует временному масштабу сигнала и удобен для локализации особенностей в сигнале.

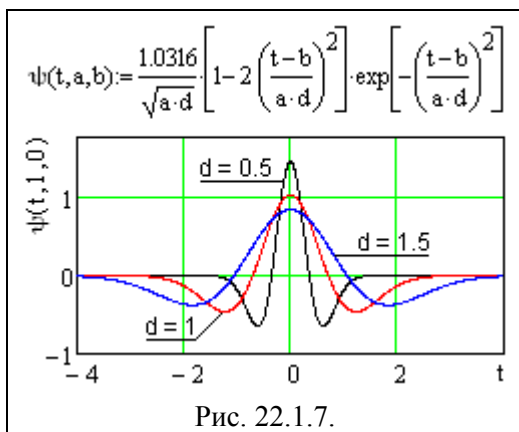


Рис. 22.1.7.

Для обеспечения возможности изменения базового размера вейвлета в его формулу множителем к параметру 'a' вводится постоянный масштабный коэффициент 'd', как это показано на рис. 22.1.7 для вейвлета МХАТ. Средняя



частота спектра базового вейвлета при  $a=1$  не должна превышать частоты Найквиста, но должна быть больше максимальной частоты в спектре сигнала.

**Размер вейвлета** в единицах  $\Delta t$  (целочисленные значения) задается параметром  $L$ . Так как значение  $L = 1$  (размер вейвлета  $2\Delta t$ ) соответствует частоте Найквиста, а частота «короткой волны» в пределах этого размера будет больше частоты Найквиста, то минимальное значение  $L$  равно 2, что обеспечивает максимальную разрешающую способность по времени. Под это значение  $L$  и рекомендуется подобрать такое максимальное значение коэффициента 'd', при котором обеспечивается (с определенной точностью) выполнение условий:

$$\int_{-L}^L \psi(t,1,0) dt \rightarrow 0, \quad \int_{-L}^L \psi(t,1,0)^2 dt \rightarrow 1. \quad (2.1.3)$$

Сохранение найденного значения  $\text{const} = L \cdot d$  обеспечивает выполнение условий (2.1.3) для любой пары значений  $L$  и  $d$ . Минимальное значение  $L=2$  действительно для SWT, но не рекомендуется для CDWT, т.к. дискретизация вейвлета порождает периодизацию его частотного спектра и перекрытие главного диапазона с боковыми, что наглядно видно на рис. 22.1.8.

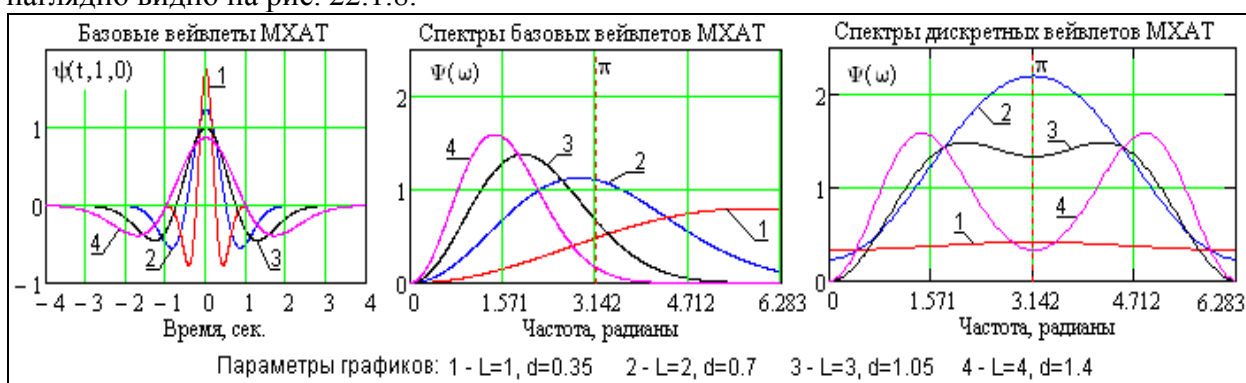


Рис. 22.1.8.

Для дискретного вейвлета МХАТ при  $L=2$  и 3 форма спектра вейвлета в области высоких частот существенно искажена, что приведет к искажению строк вейвлет-спектра. Только начиная с  $L=4$ , при средней частоте вейвлета порядка  $\pi/2$  перекрытие с боковыми диапазонами становится не столь значимым.

**Процесс преобразования** начинается с задания интервалов и шага дискретизации по масштабу и сдвигу вейвлетов. Для наглядного сравнения с SWT повторим преобразование сигнала, приведенного на рис. 22.1.5, методом CDWT с заданием параметров дискретизации:

$$T = 200, \quad V = T, \quad b = 0..V, \quad \Delta b = 1, \quad A = 60, \quad a = 1..A, \quad \Delta a = 1.$$

Дискретизации вейвлетной функции не требуется, она выполняется автоматически при вычислении вейвлетного спектра сигнала. Скалярное произведение дискретных функций требует задания в сигнале начальных (при одностороннем вейвлете) и конечных (при двустороннем вейвлете) условий. Особенностью CDWT является изменение интервала задания этих условий в прямой зависимости от масштаба вейвлета. Это означает, что продление произвольных сигналов должно производиться на длину  $LA$ , где  $A$  - максимальный заданный масштаб вейвлета. Для финитных сигналов выполнение этого требования (с заданием нулевых начальных и конечных условий) труда не представляет, как это и выполнено ниже. Формула преобразования:

$$C_{b,a} = \sum_{t=b-L \cdot a}^{b+L \cdot a} (\psi(t, a \cdot \Delta a, b) \cdot \text{if}(t < 0 \vee t > T, 0, s_t)) \quad (2.1.4)$$



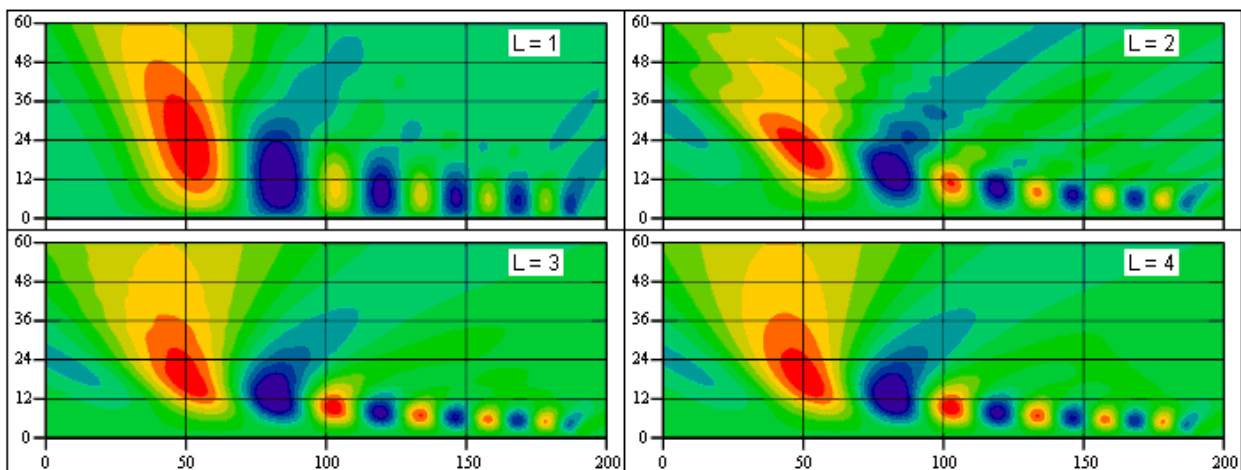


Рис. 22.1.9.

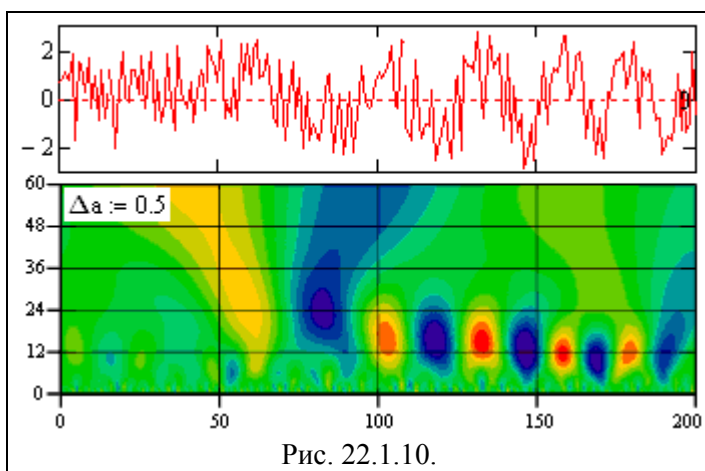


Рис. 22.1.10.

Примеры результатов преобразования приведены на рис. 22.1.9 при различных значениях параметра  $L$  для наглядного представления искажений, которые вносятся в дискретный спектр при неправильном выборе размеров базового вейвлета. При  $L=4$  сравнением с рис. 22.1.5 можно убедиться в практической идентичности спектров CWT и CDWT.

На рис. 22.1.10 приведен пример преобразования ЛЧМ сигнала (рис. 22.1.5), на который наложен статистический шум, мощность которого в 3 раза больше мощности самого сигнала. Однако вид спектра хорошо фиксирует как общую картину спектра ЛЧМ сигнала, так и распределение во времени локальных неоднородностей.

## 22.2. ДИАДНОЕ ВЕЙВЛЕТ - ПРЕОБРАЗОВАНИЕ.

Диадное вейвлет-преобразование (DWT) одномерных сигналов в системе Mathcad выполняется вейвлетом Добеши четвертого порядка db4 функциями прямого преобразования  $S := \text{wave}(s)$  и обратного  $s := \text{iwave}(S)$ . Значения параметров  $a$  и  $b$  задаются в виде степенных функций:

$$a = a_0^{-m}, \quad b = k \cdot a_0^{-m}, \quad a_0 = 2, \quad m, \quad k \in I,$$

где  $I$  – пространство целых чисел,  $m$  – параметр масштаба,  $k$  – параметр сдвига. Базис пространства  $L^2(\mathbb{R})$  в дискретном представлении:

$$\psi_{mk}(t) = |a_0|^{-m/2} \psi(a_0^{-m}t - k), \quad m, k \in I, \quad \psi(t) \in L^2(\mathbb{R}). \quad (22.2.1)$$

Вейвлет-коэффициенты прямого преобразования:

$$C_{mk} = \int_{-\infty}^{\infty} s(t) \psi_{mk}(t) dt. \quad (22.2.2)$$

Вейвлет Добеши относится к числу ортогональных вейвлетов и обратное преобразование для непрерывных сигналов при нормированном ортогональном вейвлетном базисе пространства:

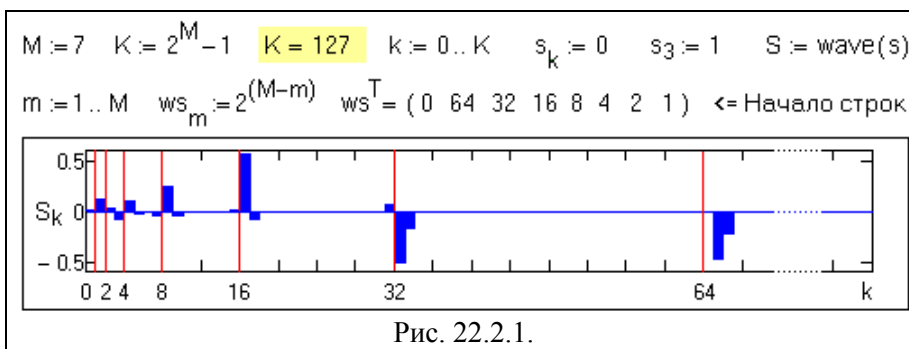
$$s(t) = \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_{mk} \psi_{mk}(t). \quad (22.2.3)$$

**Уровень декомпозиции.** Число использованных вейвлетов по масштабному коэффициенту  $m$  задает уровень декомпозиции сигнала, при этом за нулевой уровень ( $m =$

0) принимается уровень максимального временного разрешения сигнала, т.е. сам сигнал, а последующие уровни ( $m > 0$ ) образуют *вейвлет-дерево* (от коротких вейвлетов к длинным, или, по средней частоте вейвлетов, от высоких частот к низким).

Учитывая диадность преобразования, размер  $T$  (в единицах  $\Delta t$ ,  $t := 0..T-1$ ) дискретного массива  $s_t$  должен задаваться кратным  $2^M$ , где  $M$  – число уровней декомпозиции сигнала ( $m := 1..M$ ), при этом отсчет уровней детализирующих коэффициентов спектра начинается с  $m=1$  (минимальный масштабный коэффициент вейвлета  $a=2$ ), до  $m=M$  ( $a=2^M$ ). На последнем  $M$ -уровне записывается 1 детализирующий коэффициент и дополнительно конечный аппроксимирующий коэффициент. Общее количество отсчетов спектра равно количеству отсчетов сигнала. Если количество отсчетов сигнала на удовлетворяет условию  $2^M$ , то выполняется либо передискретизация сигнала с уменьшением шага  $\Delta t$ , либо дополнение интервала  $T$  сигнала до ближайшего большего значения  $2^N$ . Обычно используется второй способ, т.к. с учетом односторонней формы вейвлета Добеши для исключения искажений спектра требуется задание начальных условий. Методы дополнения сигналов не отличаются от задания начальных условий при выполнении свертки сигналов.

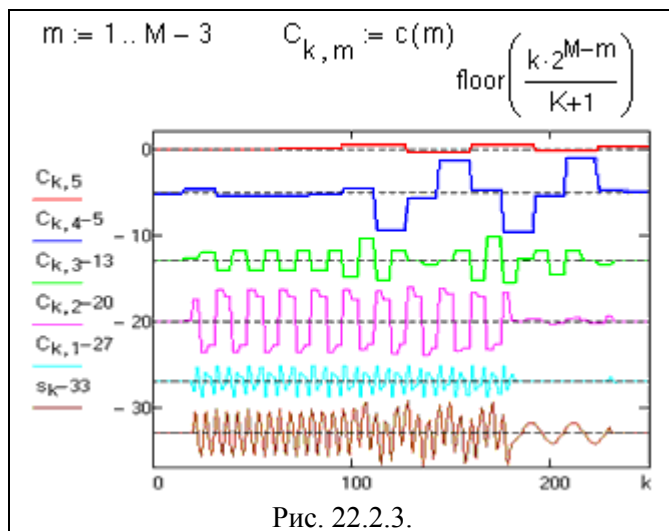
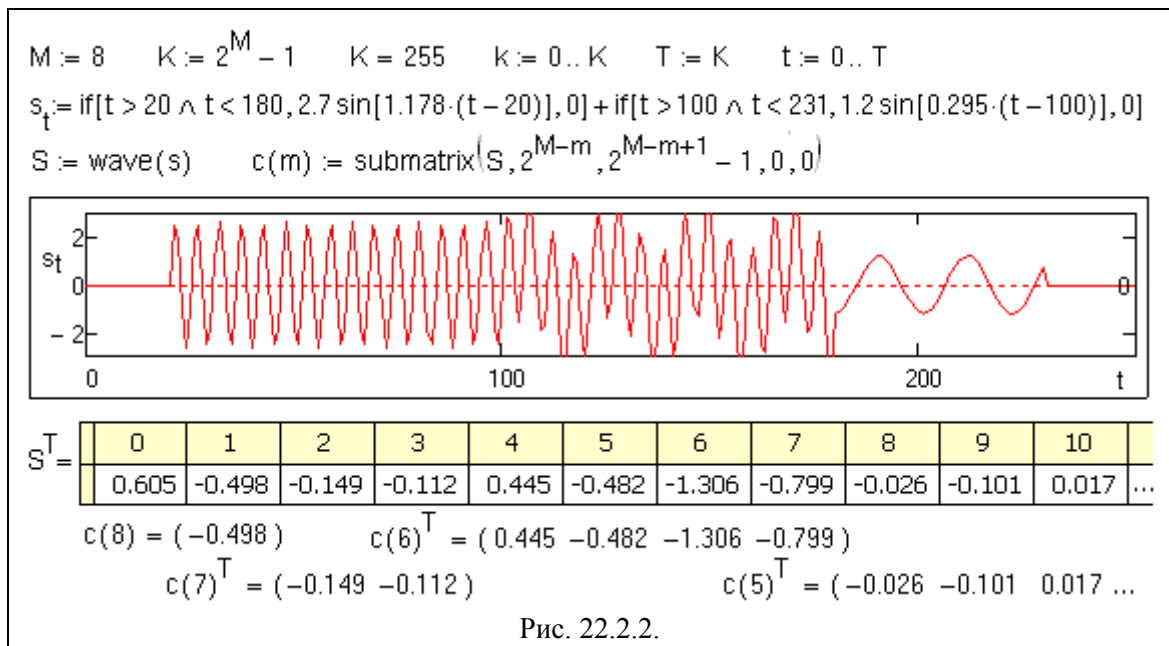
**Структура записи спектра.** Все коэффициенты спектра записываются в массив спектра в следующем порядке. Размер каждой строки (уровня декомпозиции  $m = 1..M$ ) занимает  $2^M/2^m$  отсчетов. Каждая  $m$  – строка коэффициентов уровня начинается в массиве спектра с номера  $2^{M-m}$ .



На рис. 22.2.1. приведен график вейвлетного разложения импульса Кронекера в точке  $t=3$  массива размером 128 отсчетов.

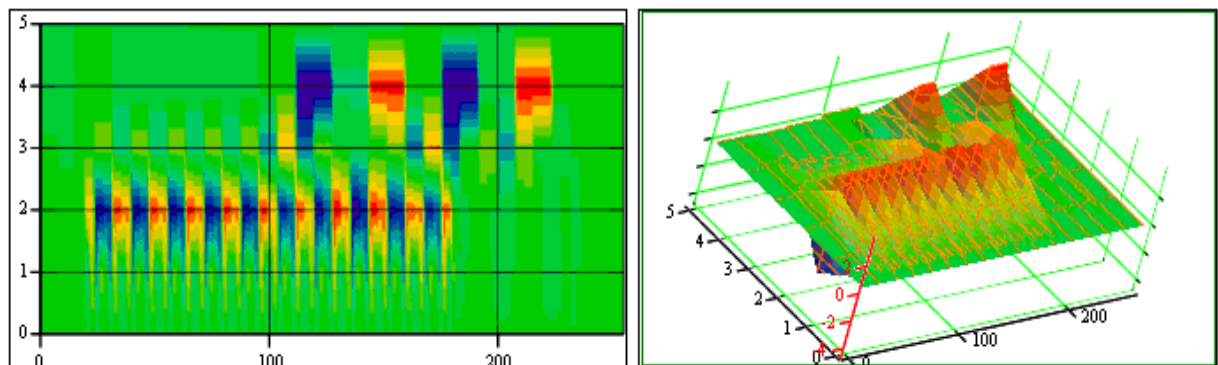
На рис. 22.2.2 приведен график модельного сигнала  $s_t$  с двумя перекрывающимися

во времени синусоидами. Для просмотра коэффициентов вейвлет-преобразования  $S$  сигнала отдельно по уровням декомпозиции  $m$  можно использовать функцию формирования субматрицы  $c(m) = \text{submatrix}(S, r1, r2, p1, p2)$ , которая записывает в строки  $m$  из вектора  $S$  отсчеты с номера  $r1$  по  $r2$  (из столбцов с  $p1$  по  $p2$ , в данном случае столбцов нет). Пример применения функции можно видеть на рисунке.



**Визуализация спектра.** При визуализации картины коэффициентов субматрица переводится в двумерный массив с приведением к единой числовой оси входного сигнала (растягивание коэффициентов по оси сдвигов без изменения их значений), как это показано на рис. 22.2.3. Это позволяет выводить в графической форме только наиболее информативные уровни разложения, а также представлять вейвлетный спектр в 2D и в 3D форме (рис. 22.2.4).

В силу диадности преобразования выразительность вейвлетного спектра DWT существенно уступает SWT и DCWT, но сохраняет все частотно-временные особенности сигналов и, что наиболее существенно, позволяет производить изменения (определенную обработку) сигнала на разных уровнях декомпозиции и выполнять обратное преобразование без потерь информации.



Качество визуализации спектра может быть улучшено при переводе субматрицы на единую временную ось с применением методов интерполяции. В качестве примера приводится листинг подпрограммы интерполяции коэффициентов спектра с

использованием кубического сплайна. Результаты применения интерполяции можно наглядно видеть на рис. 22.2.5 при сопоставлении с рисунками 22.2.3 и 22.2.4.

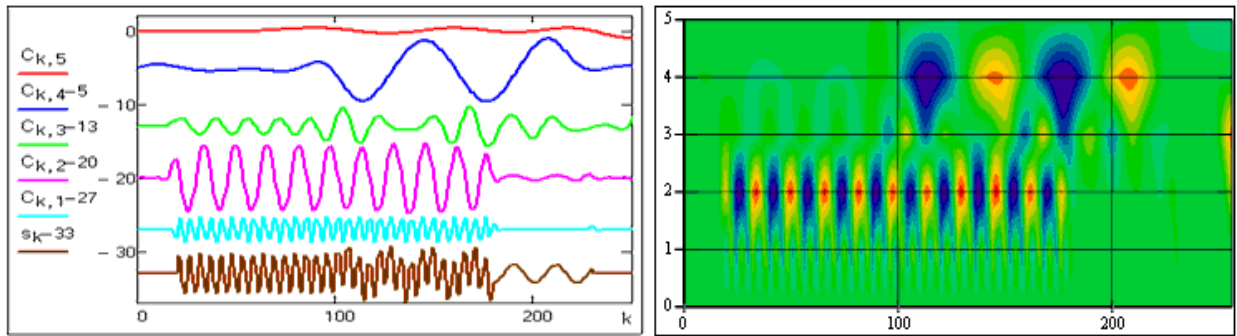


Рис. 22.2.5.

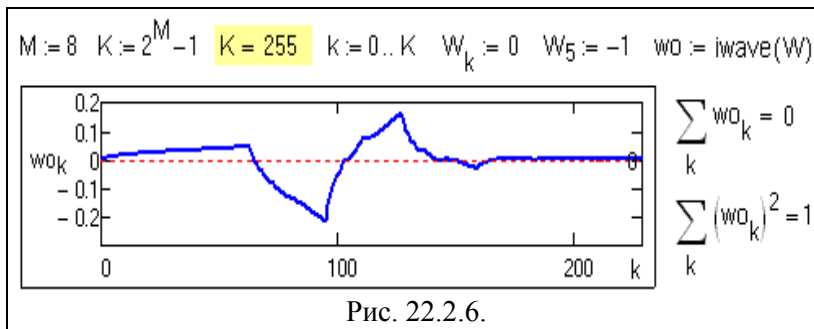


Рис. 22.2.6.

**Форму вейвлета db4** и средние частоты уровней декомпозиции (масштаб вейвлета) можно определить обратным преобразованием импульса Кронекера, задаваемого на какой-либо уровень декомпозиции спектра, при обнулении всех

других уровней спектра. Пример операции приведен на рис. 22.2.2.

Для более точного вычисления формы вейвлета число уровней декомпозиции следует устанавливать достаточно большим ( $N = 8-12$ ), а импульс Кронекера в массиве спектра задавать на 2-3 уровня меньше  $M$  и на 2-3 точку интервала отсчетов этого уровня. Восстановленный при обратном преобразовании массив отсчетов вейвлета можно с использованием БПФ перевести в спектральную область и определить среднюю частоту вейвлета (с учетом вывода вейвлета на  $K/2$  точек временной оси). При известной средней частоте вейвлета на  $m$ -уровне декомпозиции  $\omega_m$ , частотная шкала разложения при переходе на 1 уровень ниже возрастает в 2 раза, а на 1 уровень выше – в 2 раза уменьшается. Соответственно, частотная шкала 1-го уровня (максимальная частота декомпозиции) по измерениям средней частоты вейвлета на  $m$ -уровне определяется выражением:

$$\omega_1 = \omega_m 2^{m-1},$$

и для вейвлета db4 составляет 2.356 радиан. Соответственно, на любом другом  $i$ -уровне:

$$\omega_i = \omega_1 / 2^i.$$

Аналогичным образом, при занесении единицы на нулевую точку вейвлетного спектра  $S$  может быть вычислена смещенная форма скейлинг-функции в ненормированном виде.

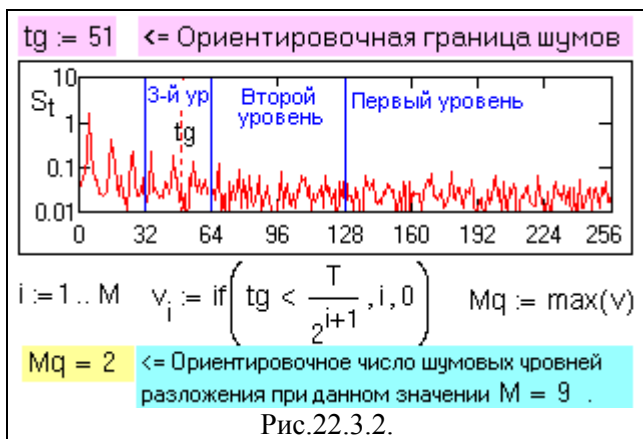
### 22.3. ВЕЙВЛЕТНАЯ ОЧИСТКА СИГНАЛОВ ОТ ШУМОВ.

Вейвлетная очистка сигналов от шумов может использоваться для любых типов сигналов, но особенно эффективна для сигналов, имеющих в своем составе крутые перепады значений (скачки), местоположение которых является информацией, подлежащей сохранению. Для наглядности использования в этом процессе типовых функций



преобразования  $\text{wave}(s)$  и  $\text{iwave}(S)$  в качестве примера выполним операцию очистки зашумленного меандра, представленного на рис. 22.3.1 на 500 точках координатной оси.

**Подготовка преобразования.** Для подготовки вейвлет-преобразования определяется количество уровней полной декомпозиции сигнала  $M$ , и массив дополняется (в данном случае нулями) до требуемой величины  $2^M$ .



Для оценки количества возможных уровней декомпозиции, в которых будет преобладать шум, можно выполнить БПФ массива. На рис. 22.3.2 представлен спектр массива в интервале  $0-\pi$  (256 отсчетов), т.е. на первой половине главного диапазона БПФ. При диадном делении спектра на каждом уровне декомпозиции, первый уровень детализирующих коэффициентов будет сформирован из высокочастотной части спектра сигнала от  $\pi/2$  до  $\pi$  (в односторонней физической шкале частот). Вторая часть спектра от  $0$

до  $\pi/2$  конвертируется в аппроксимирующие коэффициенты. На втором уровне декомпозиции аппроксимирующие коэффициенты диапазона  $0-\pi/2$  также будут разделены пополам с преобразованием диапазона  $\pi/4-\pi/2$  в детализирующие, а диапазона  $0-\pi/4$  в аппроксимирующие коэффициенты 2-го уровня декомпозиции, и т.д. Это позволяет непосредственно по частотному спектру сигнала установить ориентировочную границу шумов и, соответственно уровни декомпозиции, в которых мощность шумов соизмерима и выше мощности сигнала.

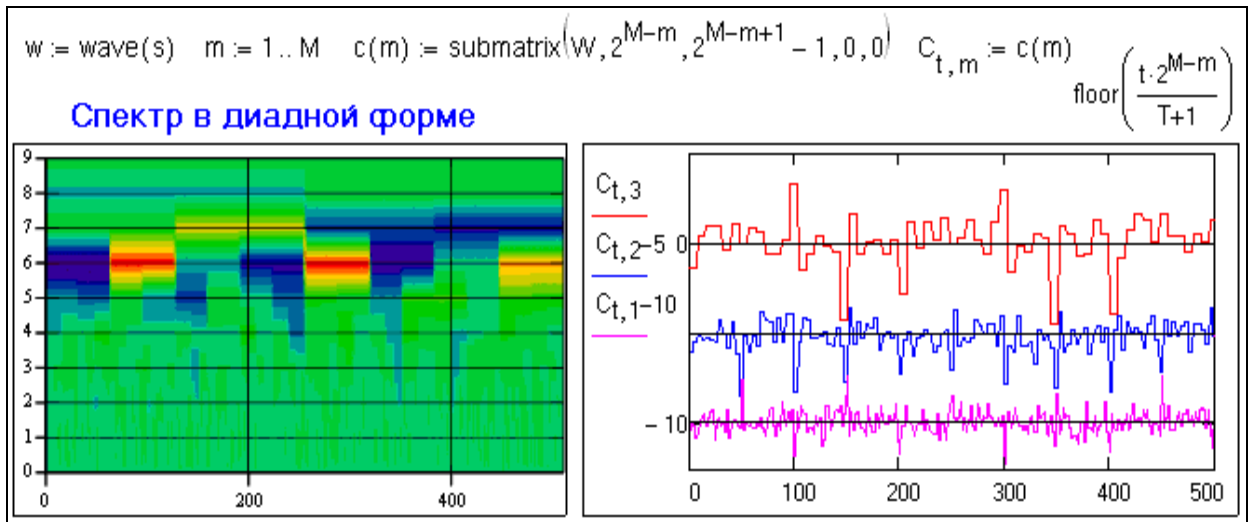


Рис. 22.3.3.

**Анализ шумов по вейвлетному спектру.** На рис. 22.3.3 приведен вейвлетный спектр и графики первых трех уровней декомпозиции модельного сигнала. Можно видеть, что сделанное выше на рис. 22.3.2 заключение по шумам подтверждается графиками вейвлетного спектра. Для сигналов без скачков первые два уровня декомпозиции явно шумовые и должны быть обнулены. Для сигналов со скачками и с высокочастотными компонентами может быть выполнена более тонкая последовательная очистка с оценкой статистики распределения коэффициентов уровней.



Рис. 22.3.4.

На рис. 22.3.4. приведена гистограмма коэффициентов первого уровня. Судя по этой гистограмме, основное шумовое распределение находится в интервале от  $-2.0\sigma$  до  $2.0\sigma$ , где  $\sigma$  - абсолютное среднеквадратическое отклонение шумовых импульсов от среднего значения (стандарт). На гистограмме наблюдаются также «хвосты», явно не входящие в основное распределение шумов, и обязанные своим происхождением скачкам и крутым перепадам в сигнале.

шумов, и обязанные своим происхождением скачкам и крутым перепадам в сигнале.

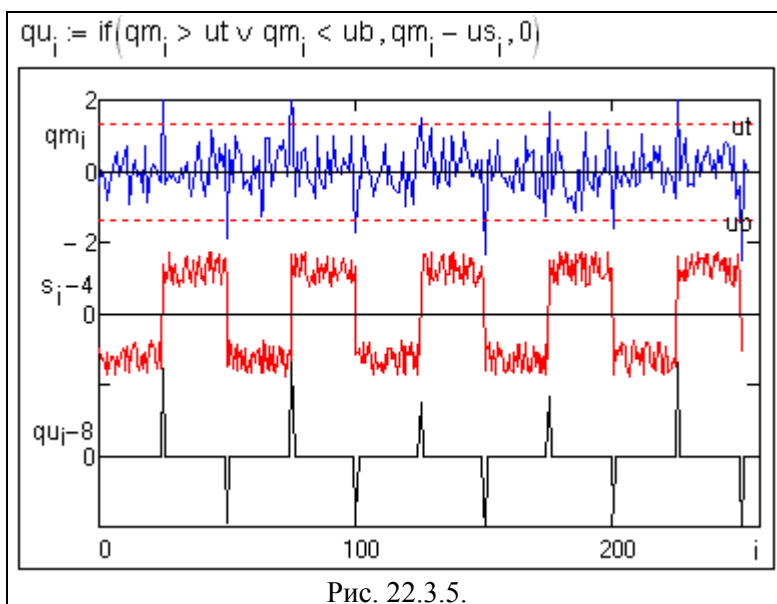


Рис. 22.3.5.

Такое предположение подтверждается и рис. 22.3.5, где приведено сопоставление графика коэффициентов уровня с исходным сигналом преобразования, приведенным к масштабу уровня. При формировании новой строки уровня qm коэффициенты, превышающие установленные пороги ub и ut шумового распределения, целесообразно сохранить полностью или с небольшим занижением значений.

После формирования новой строки qu данного уровня

декомпозиции вейвлетного спектра можно заменить этой строкой соответствующий уровень разложения в полном массиве коэффициентов и визуально оценить результаты операции. Пример приведен на рис. 22.3.6.

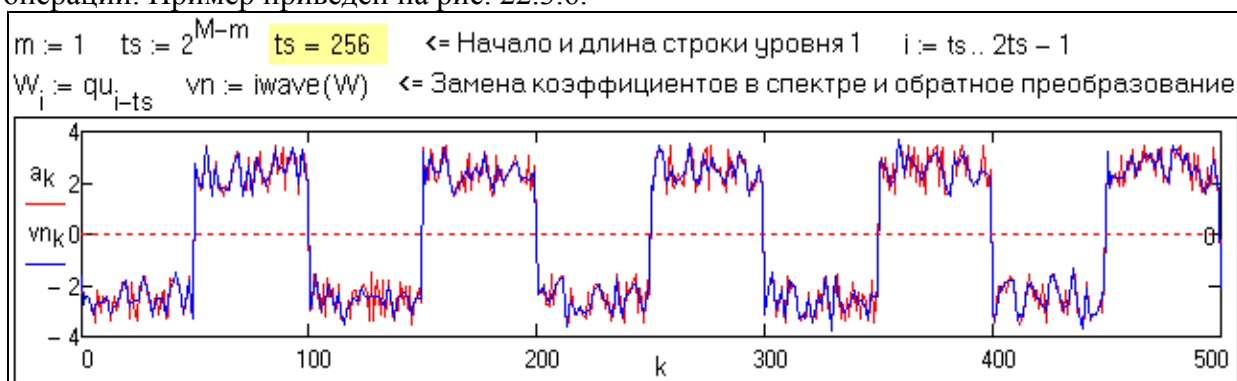


Рис. 22.3.6.

После проведения аналогичной операции на втором уровне декомпозиции получаем результат, приведенный на рис. 22.3.7. Он является конечным, т.к. попытка повторения операции на третьем уровне с данной моделью сигнала результатов не принесла.



Рис. 22.3.7.

Попутно отметим, что аналогичная методика может применяться для выделения из сигнала шумовых распределений и оценки их статистических характеристик.

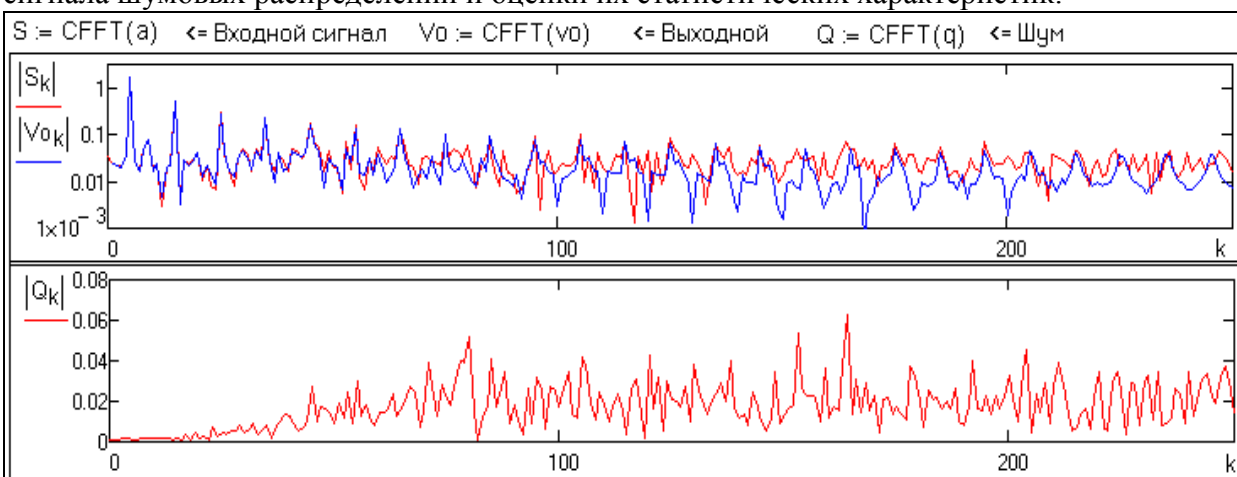


Рис. 22.3.8.

На рис. 22.3.8 приведены спектральные характеристики входного и очищенного от шумов сигнала и спектр выделенного шума. На спектре сигналов можно видеть, что вейвлетная очистка сохранила по всему частотному диапазону характерные пики меандра, формирующие периодические скачки значений сигнала, что не может выполнить практически никакой линейный частотный фильтр.



## Тема 24. ПРЕОБРАЗОВАНИЕ ГИЛЬБЕРТА-ХУАНГА

Под преобразованием Гильберта-Хуанга (Hilbert-Huang transform – ННТ) понимается метод эмпирической модовой декомпозиции (EMD) нелинейных и нестационарных процессов и Гильбертов спектральный анализ (HSA). ННТ представляет собой частотно-временной анализ данных (сигналов) и не требует *априорного* функционального базиса преобразования. Функции базиса получаются адаптивно непосредственно из данных процедурами отсеивания функций «эмпирических мод». Мгновенные частоты вычисляются от производных фазовых функций Гильбертовым преобразованием функций базиса. Заключительный результат представляется в частотно-временном пространстве /1/.

EMD-HSA был предложен Норденом Хуангом в 1995 в США (NASA) для изучения поверхностных волн тайфунов, с обобщением на анализ произвольных временных рядов коллективом соавторов в 1998 г. /2/. В последующие годы, по мере расширения применения EMD-HSA для других отраслей науки и техники, вместо термина EMD-HSA был принят более короткий термин преобразования ННТ.

Традиционные методы анализа данных предназначены, как правило, для линейных и стационарных сигналов и систем, и только в последние десятилетия начали активно развиваться методы анализа нелинейных, но стационарных и детерминированных систем, и линейных, но нестационарных данных (вейвлетный анализ, распределение Wagner-Ville и др.). Между тем, большинство естественных материальных процессов, реальных физических систем и соответствующих этим процессам и системам данных в той или иной мере являются нелинейными и нестационарными, и при анализе данных используются определенные упрощения, особенно в отношении априорно устанавливаемого базиса преобразования данных в новые, удобные для обработки и анализа метрические пространства.

Необходимое условие корректного представления нелинейных и нестационарных данных заключается в том, чтобы иметь возможность формирования адаптивного базиса, функционально зависящего от содержания самих данных. Такой подход и реализуется в методе ННТ, хотя на данный момент без достаточно строгих математических обоснований. Хорошие результаты применения метода для решения многих практических задач позволяют надеяться, что за разработкой строгой теории метода дело не станет.

### 24.1. ПРЕОБРАЗОВАНИЕ ГИЛЬБЕРТА И АНАЛИТИЧЕСКИЙ СИГНАЛ

**Определение и формула.** Преобразование Гильберта  $\text{TH}[x(t)]$  действительной функции  $x(t)$ ,  $-\infty < t < \infty$ , - есть действительная функция, определенная как

$$\begin{aligned}\tilde{x}(t) &= \text{TH}[x(t)] = x(t) * (1/\pi t), \\ \tilde{x}(t) &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau.\end{aligned}\quad (24.1.1)$$

Функция  $1/(t-\tau)$  называется *ядром преобразования* Гильберта. Преобразование имеет особую точку при  $t-\tau \Rightarrow 0$ , в которой при вычислении используется главное значение интеграла по Коши. Функции  $x(t)$  и  $\tilde{x}(t)$  обычно называют сопряженными по Гильберту.

Физически, преобразование Гильберта может быть интерпретировано как естественный  $\pi/2$  фазовращатель, который при прохождении через систему сигнала  $x(t)$  изменяет фазу всех частотных составляющих сигнала на  $\pi/2$ , и тем самым делает сигнал  $\tilde{x}(t)$  ортогональным сигналу  $x(t)$ . Это позволяет сформировать из сигналов  $x(t)$  и  $\tilde{x}(t)$  комплексный аналитический сигнал  $z(t)$ , как



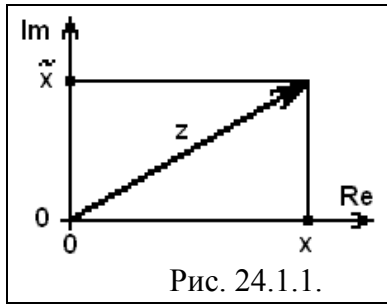


Рис. 24.1.1.

$$z(t) = x(t) + j \tilde{x}(t), \quad (24.1.2)$$

где  $z(t)$  представлен вектором на комплексной плоскости с проекциями на действительной и мнимой оси соответственно  $x(t)$  и  $\tilde{x}(t)$  (рис. 24.1.1). Преимущество этого представления состоит в том, что возникает возможность однозначно определять текущие временные параметры сигнала  $z(t)$ , а именно – мгновенные значения его амплитуды и фазы.

Допустим, что имеем зарегистрированный радиосигнал  $x(t)$  с несущей частотой  $\omega_0$ , который содержит определенную информацию, заключенную в огибающей сигнала  $u(t)$  и его фазе  $\varphi(t)$ :

$$x(t) = u(t) \cos(\omega_0 t + \varphi(t)).$$

В другой форме:

$$x(t) = a(t) \cdot \cos(\omega_0 t) + b(t) \cdot \sin(\omega_0 t),$$

$$a(t) = u(t) \cos \varphi(t), \quad b(t) = u(t) \sin \varphi(t), \quad u(t) = \sqrt{a^2(t) + b^2(t)}, \quad \text{tg } \varphi(t) = b(t)/a(t).$$

**Мгновенная амплитуда, фаза и частота сигнала.** С использованием преобразования Гильберта из сигнала  $x(t)$  сформируем аналитически сопряженный сигнал  $\tilde{x}(t)$ . С учетом сдвига фазы на  $\pi/2$ :

$$\tilde{x}(t) = a(t) \cdot \sin(\omega_0 t) - b(t) \cdot \cos(\omega_0 t).$$

$$z(t) = x(t) + j \cdot \tilde{x}(t).$$

Квадрат модуля сигнала  $z(t)$ :

$$|z(t)|^2 = x^2(t) + \tilde{x}^2(t) = a^2(t)[\cos^2(\omega_0 t) + \sin^2(\omega_0 t)] + b^2(t)[\cos^2(\omega_0 t) + \sin^2(\omega_0 t)] = u^2(t).$$

Отсюда, огибающая  $u(t)$  и мгновенная фаза  $\varphi(t)$  сигнала  $x(t)$ :

$$u(t) = \sqrt{x^2(t) + \tilde{x}^2(t)}. \quad (24.1.3)$$

$$\phi(t) = \omega_0 t + \varphi(t) = \text{arctg}[\tilde{x}(t)/x(t)]. \quad (24.1.4)$$

$$\varphi(t) = \phi(t) - \omega_0 t.$$

Мгновенная частота сигнала определяется по скорости изменения мгновенной фазы:

$$d\varphi(t)/dt = [\tilde{x}'(t)x(t) - x'(t)\tilde{x}(t)] / (\tilde{x}^2(t) + x^2(t)). \quad (24.1.5)$$

Определения (24.1.3) и (24.1.5) подразумевают, что в каждый текущий момент времени в сигнале существует единственное значение амплитуды и частоты. Однако физическая интерпретация понятия «мгновенности» неоднозначна и требует определенных ограничений.

Действительно, в стационарных моногармонических сигналах и в сигналах с непрерывным гладким изменением частоты понятие «мгновенности» имеет вполне определенный физический смысл, поскольку точно фиксирует положение каждой текущей точки в частотно-временном пространстве, что можно наглядно видеть на рис. 24.1.2 и 24.1.3.

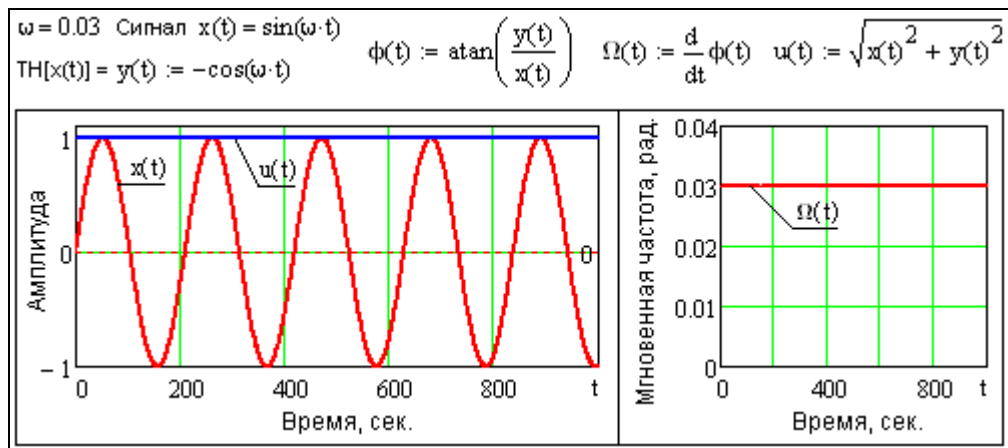


Рис. 24.1.2. Мгновенная амплитуда и частота гармонического сигнала.

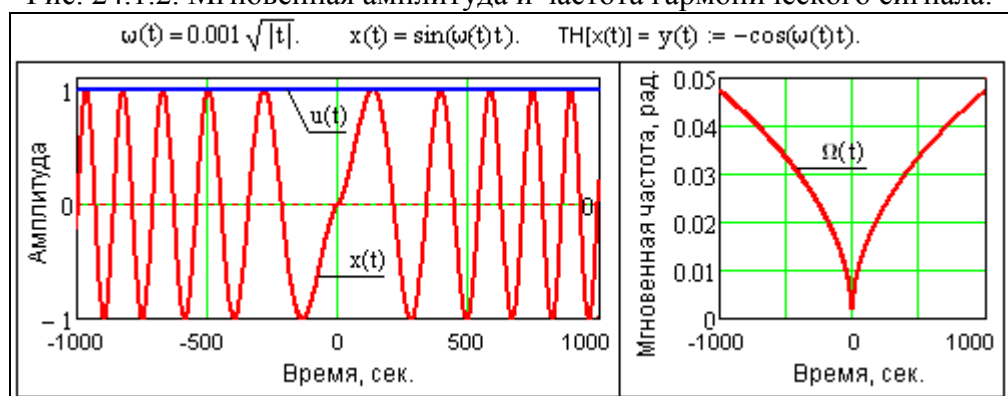


Рис. 24.1.3. Мгновенная амплитуда и частота гладкого непрерывного сигнала.

Однако уже для суммы двух гармонических сигналов, приведенных на рис. 24.1.4, положение усложняется. Физически в каждой текущей точке сигнала постоянно присутствуют две частоты с определенной амплитудой колебаний. Мгновенная амплитуда преобразования Гильберта в этом случае отображает не сумму значений гармоник в каждый текущий момент времени, а огибающую интерференции этих гармоник, при этом максимальные мгновенные значения огибающей, равные сумме амплитуд гармоник, фиксируются в точках максимумов по модулю суммы первых производных гармоник, а минимальные значения, равные разности амплитуд гармоник, в точках минимума суммы модулей первых производных гармоник. Это обеспечивает симметричность верхней и нижней огибающих относительно временной оси. Все вышеизложенное действительно и для любых многотональных сигналов.

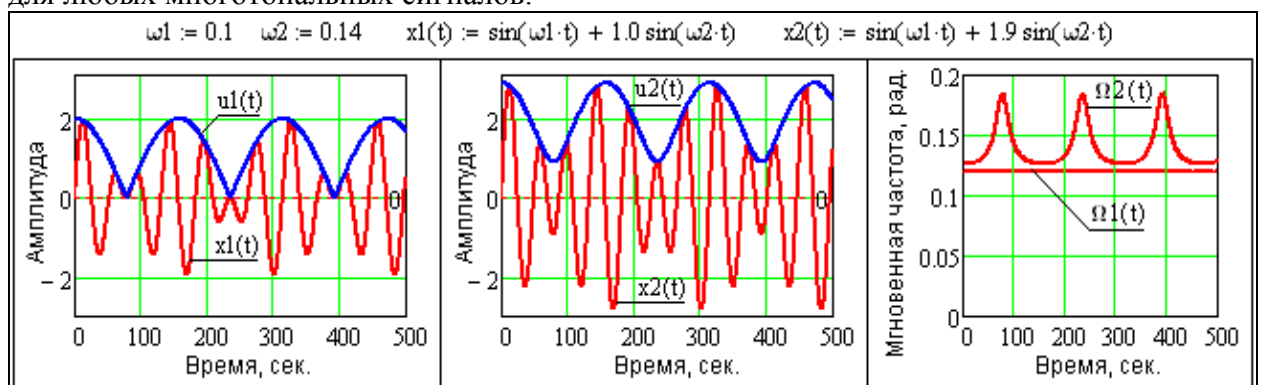


Рис. 24.1.4. Мгновенные амплитуды и частоты двутональных сигналов.

Что касается мгновенных значений частоты, то при равной амплитуде гармоник мгновенная частота во всех точках сигнала равна среднему значению частот гармоник. При неравной амплитуде гармоник функция мгновенной частоты сдвигается в сторону частоты гармоники с большей амплитудой и приобретает пульсирующий характер, при

этом пики экстремумов частотных пульсаций соответствуют минимумам огибающих пульсаций и также направлены в сторону частоты гармоники с большей амплитудой.

**Симметрия огибающих сигнала** имеет существенное значение и ее нарушение, например, в случае наличия в сигнале постоянной составляющей или тренда, существенно изменяет результаты преобразования.

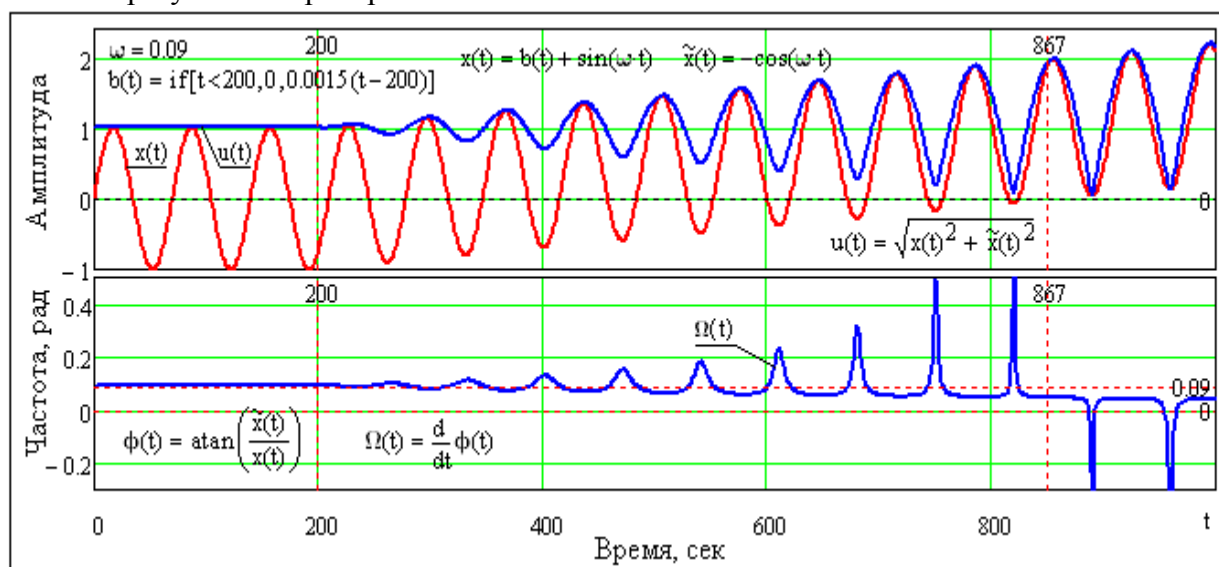


Рис. 24.1.5. Преобразование Гильберта сигнала с трендом.

На рис. 24.1.5 приведен сигнал, представленный монотонной гармоникой с амплитудой  $a=1$  в сумме с трендом  $b(t)$ , значения которого изменяются линейно от нуля до 1.2 в интервале  $200 < t < 1000$ .

При отсутствии тренда ( $0 < t < 200$ ) значения мгновенной амплитуды и частоты соответствуют амплитуде и частоте гармоники. При появлении в сигнале тренда и его значениях  $b(t) < a$  в значениях мгновенной амплитуды появляются пульсации, синхронные с гармоникой сигнала, полный размах которых (от максимума до минимума) нарастает пропорционально  $b(t)/(2a)$ , и становится практически постоянным порядка  $2a$  при  $b(t) > a$ .

Отношение  $\alpha = b(t)/a$  в интервале  $0 \leq \alpha \leq 1$  в первом приближении может считаться коэффициентом локальной асимметрии формы сигнала относительно временной оси. Асимметрия формы сигнала существенно влияет на вычисления мгновенных частот, что наглядно видно на частотно-временном графике рис. 24.1.5 в интервале  $200 < t < 867$ . В интервалах минимумов мгновенных амплитуд появляются частотные пики максимумов, амплитуда которых пропорциональна  $\alpha$ . При  $\alpha > 1$  колебания становятся стоячими и частотные пики меняют знак, уходя в область отрицательных частот, а, следовательно, мгновенные частоты в этом случае вообще не имеют смысла.

Таким образом, можно констатировать, что для простых гармонических сигналов физически значимая мгновенная частота может быть определена только для функций локально симметричных относительно нулевого среднего уровня. Аналогичное заключение может быть распространено и на многокомпонентные сигналы, при условии, что каждый из таких сигналов симметричен относительно нулевого уровня.

В принципе, любой произвольный сигнал можно рассматривать в виде суммы колебательных процессов, удовлетворяющих условию симметричности и наложенных на тренд произвольной формы. В этом случае правомочна и обратная задача разложения произвольного сигнала на эти составляющие компоненты и остаточный тренд. Метод решения этой задачи был предложен Н.Хуангом и получил название эмпирической модовой декомпозиции (EMD) сигналов.

## 24.2. МЕТОД ЭМПИРИЧЕСКОЙ МОДОВОЙ ДЕКОМПОЗИЦИИ СИГНАЛОВ

EMD (Empirical Mode Decomposition) - метод разложения сигналов на функции, которые получили название внутренних или «эмпирических мод». Метод представляет

собой адаптивную итерационную вычислительную процедуру разложения исходных данных (непрерывных или дискретных сигналов) на эмпирические моды или внутренние колебания.

**Огибающие сигналов.** У каждого сигнала имеются *локальные экстремумы*: чередующиеся локальные максимумы и локальные минимумы с произвольным расположением по координатам (независимым переменным) сигналов. По этим экстремумам с использованием методов аппроксимации можно построить две огибающие сигналов: *нижнюю* - построенную по точкам локальных минимумов, и *верхнюю* - построенную по точкам локальных максимумов, а также функцию «*среднего значения огибающих*», которой отвечает срединная линия, расположенная в точности между нижней и верхней огибающими.

**Функции внутренних мод сигналов.** Модовая декомпозиция сигналов основана на предположении, что любые данные состоят из различных внутренних колебаний (intrinsic mode functions, IMF). В любой момент времени данные могут иметь множество сосуществующих внутренних колебаний - IMFs. Каждое колебание, линейное или нелинейное, представляет собой модовую функцию, которая имеет экстремумы и нулевые пересечения. Кроме того, колебания в определенной степени «симметричны» относительно локального среднего значения. Конечные сложные данные образуются суммой модовых функций, наложенных на региональный тренд сигнала.

Эмпирическая мода - это такая функция, которая обладает следующими свойствами:

1. Количество экстремумов функции (максимумов и минимумов) и количество пересечений нуля не должны отличаться более чем на единицу.

2. В любой точке функции среднее значение огибающих, определенных локальными максимумами и локальными минимумами, должно быть нулевым.

IMF представляет собой колебательный режим, но вместо постоянной амплитуды и частоты, как в простой гармонике, у IMF могут быть переменная амплитуда и частота, как функции независимой переменной (времени, координаты, и пр.). Первое свойство гарантирует, что локальные максимумы функции всегда положительны, локальные минимумы соответственно отрицательны, а между ними всегда имеют место пересечения нулевой линии. Второе свойство гарантирует, что мгновенные частоты функции не будут иметь нежелательных флуктуаций, являющихся результатом асимметричной формы волны.

Любую функцию и любой произвольный сигнал, изначально содержащие произвольную последовательность локальных экстремумов (минимум 2), можно разделить на семейство функций IMFs и остаточный тренд. Если данные лишены экстремумов, но содержат точки перегиба («скрытые» экстремумы наложения модовых функций и крутых трендов), то для открытия экстремумов может использоваться дифференцирование сигнала.

Допустим, что имеется произвольный сигнал  $y(t)$ . Сущность метода EMD заключается в последовательном вычислении функций эмпирических мод  $c_j(t)$  и остатков  $r_j(t) = r_{j-1}(t) - c_j(t)$ , где  $j = 1, 2, 3, \dots, n$  при  $r_0 = y(t)$ . Результатом разложения будет представление сигнала в виде суммы модовых функций и конечного остатка:

$$x(t) = \sum_{j=1}^n c_j(t) + r_n(t),$$

где  $n$  — количество эмпирических мод, которое устанавливается в ходе вычислений.

Для наглядности методику реализации EMD рассмотрим на примере разложения цифрового массива модельного сигнала  $y(k)$ , представленного на рис. 24.2.1. Сигнал смоделирован суммой трех нестационарных по амплитуде гармоник различной частоты на интервале отсчетов по  $k$  от 0 до 200, и продлен на начальном и конечном участках на интервалы  $tr=4$  для задания начальных и конечных условий преобразования и устранения

ошибок преобразования на конечных интервалах обрабатываемого массива данных.

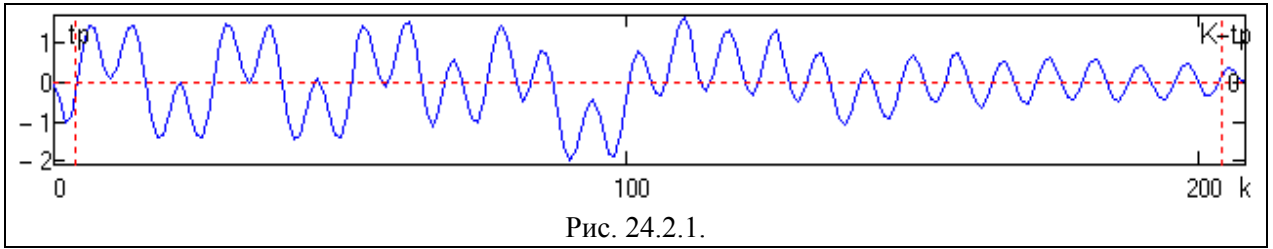


Рис. 24.2.1.

**Процесс отсеивания функций IMF.** Алгоритм эмпирической декомпозиции сигнала складывается из следующих операций его преобразования.

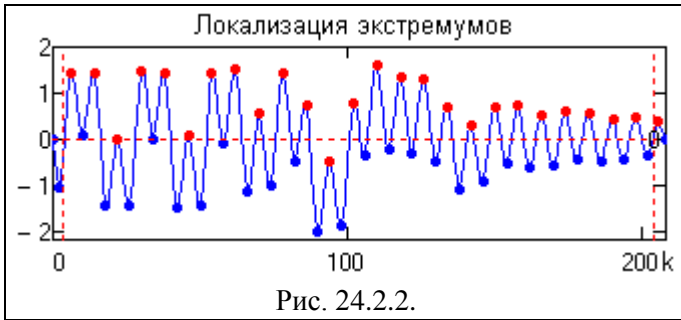


Рис. 24.2.2.

**Операция 1.** Находим в сигнале  $y(k)$  положение всех локальных экстремумов, максимумов и минимумов процесса (номера точек  $k_{i.ext}$  экстремумов), и значения  $y(k_{i.ext})$  в этих точках (рис. 24.2.2). Между этими экстремумами сосредоточена вся информация сигнала. Группируем отдельно для максимумов и для

минимумов массивы координат  $k_{i.ext}$  и соответствующих им амплитудных значений  $y(k_{i.ext})$ . Число строк в массивах максимумов и минимумов не должно отличаться более чем на 1.

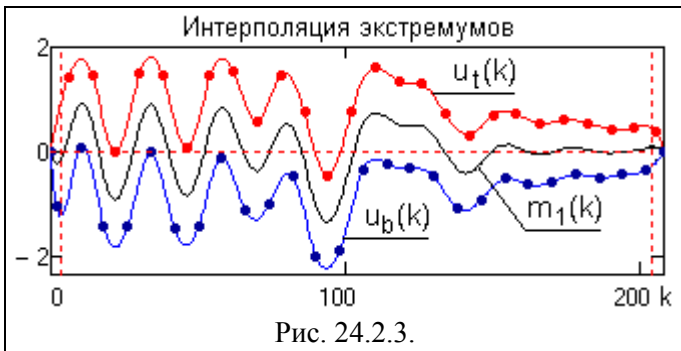


Рис. 24.2.3.

**Операция 2.** Кубическим сплайном (или каким либо другим методом) вычисляем верхнюю  $u_t(k)$  и нижнюю  $u_b(k)$  огибающие процесса соответственно, по максимумам и минимумам, как это показано на рис. 24.2.3 (красный и синий цвет). Определяем функцию средних значений  $m_1(k)$  между огибающими (черный цвет).

$$m_1(k) = (u_t(k) + u_b(k)) / 2.$$

Разность между сигналом  $y(k)$  и функцией  $m_1(k)$  дает нам первую компоненту отсеивания (*Sifting*) – функцию  $h_1(k)$ , которая является первым приближением к первой функции IMF:

$$h_1(k) = y(k) - m_1(k). \quad (24.2.1)$$

**Операция 3.** Повторяем операции 1 и 2, принимая вместо  $y(k)$  функцию  $h_1(k)$ , и находим второе приближение к первой функции IMF – функцию  $h_2(k)$ .

$$h_2(k) = h_1(k) - m_2(k).$$

**Последующие итерации** выполняются аналогично. Алгоритм итераций нахождения первой функции IMF:

$$h_i(k) = h_{i-1}(k) - m_i(k), \quad (24.2.2)$$

По мере увеличения количества итераций функция  $m_i(k)$  стремится к нулевому значению, а функция  $h_i(k)$  – к неизменяемой форме. С учетом этого, естественным критерием останова итераций является задание определенного предела по нормализованной квадратичной разности между двумя последовательными операциями приближения, определяемой как

$$\delta = \sum_k [ |h_{i-1}(k) - h_i(k)|^2 / h_{i-1}^2(k) ]. \quad (24.2.3)$$

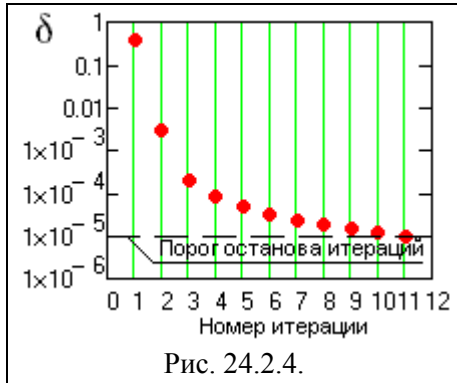


Рис. 24.2.4.

Как правило, для выполнения качественного отсеивания модовых функций достаточно 6-8 итераций. Слишком строгий критерий останова может завышать количество IMF и создавать компоненты, не несущие какой-либо полезной информации. С другой стороны, при слабом критерии возможно отсеивание IMF, не полностью удовлетворяющих свойствам модовых функций, что может приводить к появлению в этих IMF отрицательных мгновенных частот.

Останов итераций по нормализованной квадратичной разности (24.2.3) исторически был первым по применению. В 2003 г. Quek [4] предложил другой, более эффективный критерий переменной меры, определенной как

$$\delta = \sum_k |h_{i-1}(k) - h_i(k)|^2 / \sum_k h_{i-1}^2(k). \quad (24.2.3)$$

Пример изменения значений  $\delta$  в процессе итераций приведен на рис. 24.2.4.

Однако для сложных и объемных (по количеству отсчетов) сигналов в процессе итераций может изменяться количество выделяемых экстремумов (появление ранее «скрытых» экстремумов), при этом наблюдаются скачки значения  $\delta$  в большую сторону, начиная с которых снова начинается процесс уменьшения  $\delta$ , а общее количество итераций может увеличиваться до 20-30 без существенного повышения качества отсеивания. Опыт показывает, что для оптимальных отсеиваний число итераций порядка 6-8 является вполне достаточным. Все большее применение находят и другие критерии останова процесса отсеивания, ориентированные на характер и особенности обрабатываемых данных, и останов по нескольким критериям с заданием определенных логических условий по их соотношению (например, по порогу  $\delta$ , но не более J-итераций, и т.п.).

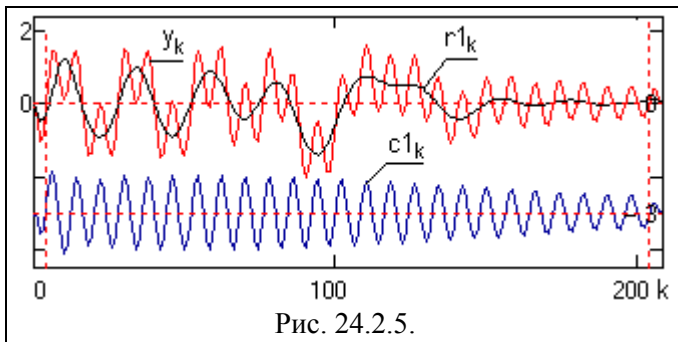


Рис. 24.2.5.

Последнее значение  $h_i(k)$  итераций принимается за наиболее высокочастотную функцию  $c_1(k) = h_i(k)$  семейства IMF, которая непосредственно входит в состав исходного сигнала  $y(k)$ . Это позволяет вычесть  $c_1(k)$  из состава сигнала и оставить в нем более низкочастотные составляющие (показано на рис. 24.2.5):

$$r_1(k) = y(k) - c_1(k). \quad (24.2.4)$$

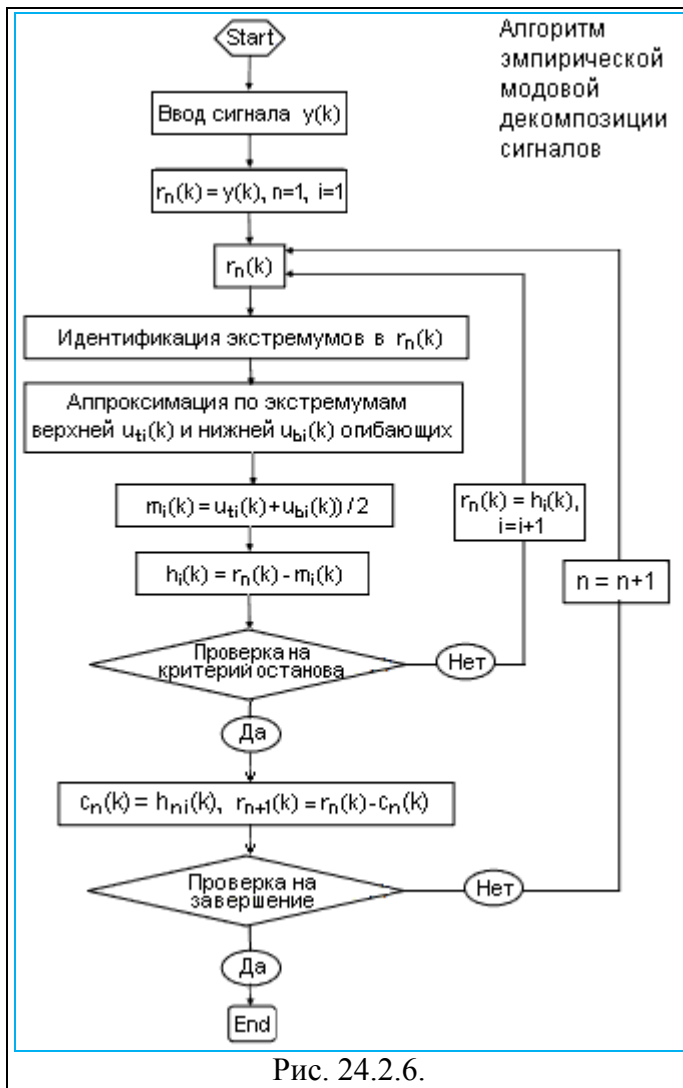
Функция  $r_1(k)$  обрабатывается как новые данные по аналогичной методике с нахождением второй функции IMF —  $c_2(k)$ , после чего процесс продолжается:

$$r_2(k) = r_1(k) - c_2(k), \text{ и т.д.} \quad (24.2.5)$$

Таким образом, достигается декомпозиция сигнала в  $n$  — эмпирическом приближении:

$$y(k) = \sum_n c_n(k) + r_n(k). \quad (24.2.6)$$





Полный алгоритм EMD приведен на рис. 24.2.6.

Критерии останова процесса декомпозиции сигнала могут быть следующими:

1. Остаток  $r_n(k)$  не содержит экстремальных точек, т.е. становится либо константой, либо монотонной функцией, из которой больше не может быть извлечено функций IMF.
2. Остаток  $r_n(k)$  во всем интервале задания сигнала становится несущественным по своим значениям по сравнению с сигналом и не представляет интереса для анализа.
3. Так как суммирование всех функций IMF (реконструкция сигнала) должно давать исходный сигнал, то можно останавливать разложение заданием относительной погрешности среднеквадратической реконструкции (без учета остатка  $r_n(k)$ ).
4. По мере увеличения количества функций IMF относительная среднеквадратическая погрешность реконструкции достаточно

сложных и протяженных сигналов уменьшается, но, как правило, имеет определенный минимум. По-видимому, это определяется попытками алгоритма разложить остаток на функции, частично компенсирующие друг друга. Соответственно, останов программы может выполняться, если следующая выделенная функция IMF увеличивает погрешность реконструкции.

Другими словами, остановка декомпозиции сигнала должна происходить при максимальном «выпрямлении» остатка, т.е. превращения его в тренд сигнала по интервалу задания с числом экстремумов не более 2-3. Даже для данных с нулевым средним значением конечный остаток может отличаться от нуля. Чтобы применять метод EMD, центрирования данных не требуется, метод нуждается только в локализациях экстремумов. Нулевая линия для каждого компонента декомпозиции будет сформирована процессом отсеивания. Извлеченные IMFs локально симметричны, имеют физически значимые функции мгновенных частот, различные IMFs не показывают ту же самую частоту в то же самое время. Каждая IMF содержит более низкие частотные составляющие, чем извлеченная перед ней.

На рис. 24.2.7 приведен пример полной декомпозиции сигнала с остановом по критерию 1. На верхнем графике рисунка приведен входной сигнал преобразования (красным) и сигнал обратной реконструкции (пунктиром) суммированием функций разложения  $c_i$  ( $c_1-c_5$ ).

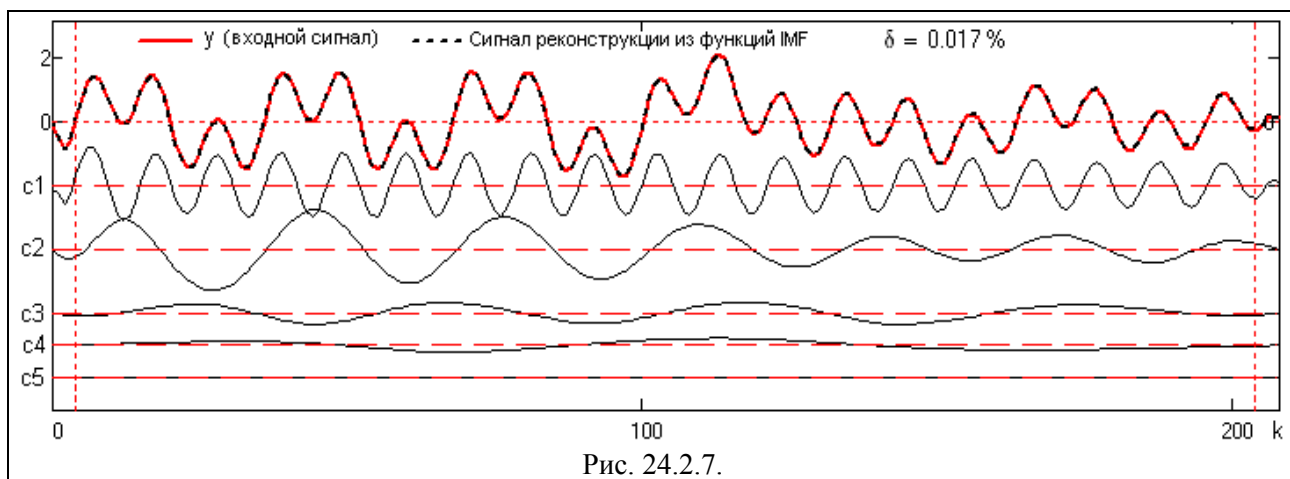


Рис. 24.2.7.

Компоненты EMD обычно физически значимы, поскольку характеристические параметры функций IMF определяются материальными данными.

**Ортогональность базиса декомпозиции.** Таким образом, входной сигнал  $y(k)$  в соответствии с выражением (24.2.6) раскладывается по базису, который, не определен аналитически, но удовлетворяет всем традиционным требованиям базиса. На основании проверки на модельных и опытных данных он является:

- законченным и сходящимся (сумма всех функций IMF и остатка равна исходному сигналу и не зависит от критериев останова итераций),
- ортогональным (все IMF и остаток ортогональны друг другу),
- единственным.

И, что самое главное – он является адаптивным, так как получен непосредственно из анализируемых данных эмпирическим методом.

Ортогональность базиса может быть проверена скалярным произведением любых пар компонентов IMF. Сумма (24.2.6) всех компонентов IMF, включая остаток, должна реконструировать входной сигнал и может использоваться для определения ошибки декомпозиции. Как правило, наибольшие локальные ошибки декомпозиции наблюдаются на концевых участках входного массива данных. Для исключения ошибок рекомендуется задавать интервалы начальных и конечных условий, а сигнал на этих интервалах формировать какой-либо функцией прогнозирования, или продлевать (четно или нечетно) функцией самого сигнала.

Н. Хуанг утверждает также, что базис разложения является единственным. Но это утверждение можно считать спорным. Эмпирический процесс разложения сигнала в силу своей адаптивности неуправляем, по крайней мере, в настоящей форме. Даже монотональные составляющие многокомпонентного сигнала при определенном влиянии дестабилизирующих факторов (шумов, импульсных помех и т.п.) и близких по частоте соседних компонент могут при декомпозиции «перетекать» на отдельных временных интервалах в модовые функции соседних IMF.

**Примеры практического применения EMD.** В качестве примера в работе [3] приводится EMD-анализ данных девиации периода вращения Земли. В результате исследований всем выделенным функциям IMF сопоставлены определенные физические процессы, которыми и вызвано их формирование (влияние штормов и тайфунов, месячных вариаций мощности приливов, явления Эль-Ниньо и прочие факторы). Ниже, без комментариев, приводятся выборки из результатов данного анализа, демонстрирующие свойства базиса.



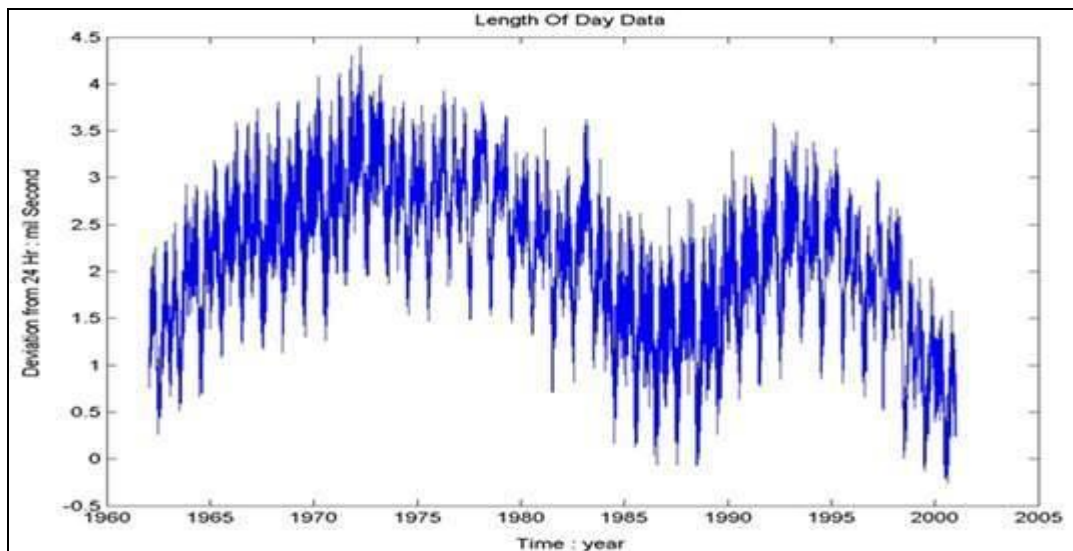


Рис. 24.2.8. Данные.

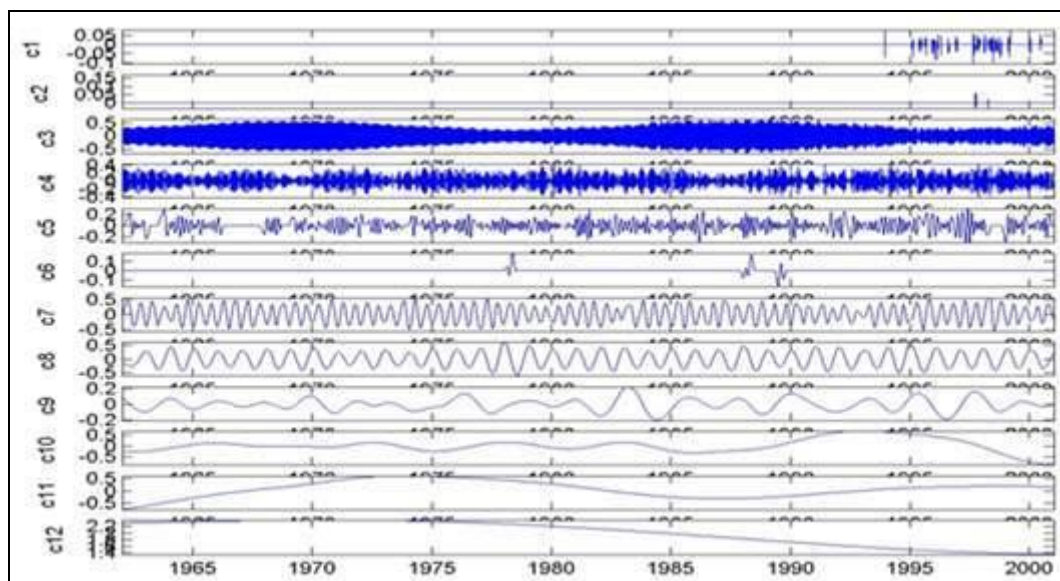


Рис. 24.2.9. Семейство IMF.

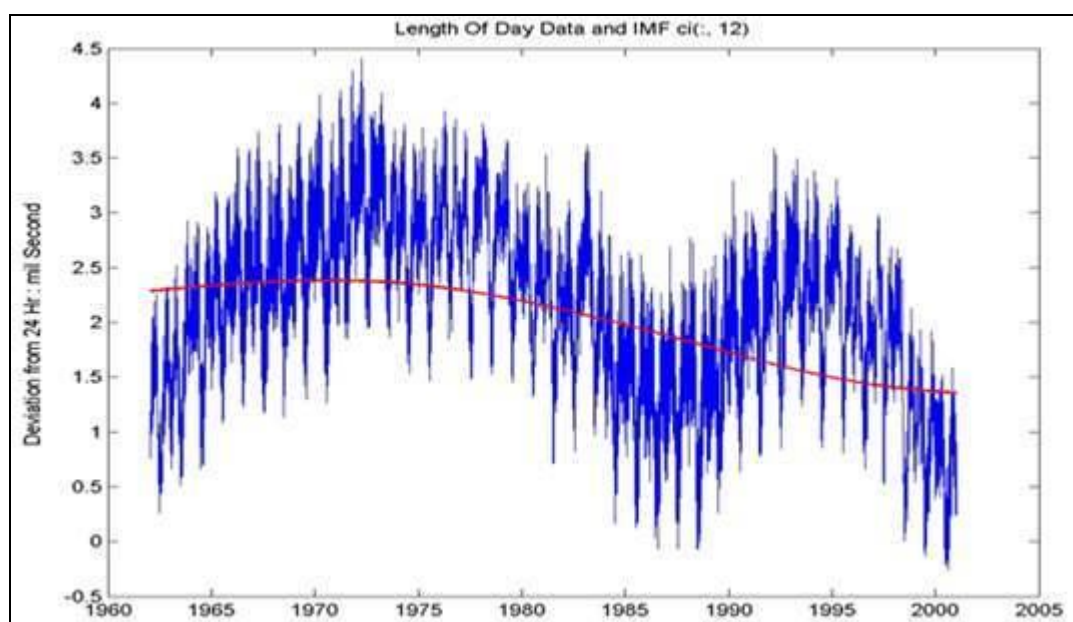


Рис. 24.2.10. Данные и c12 IMF.

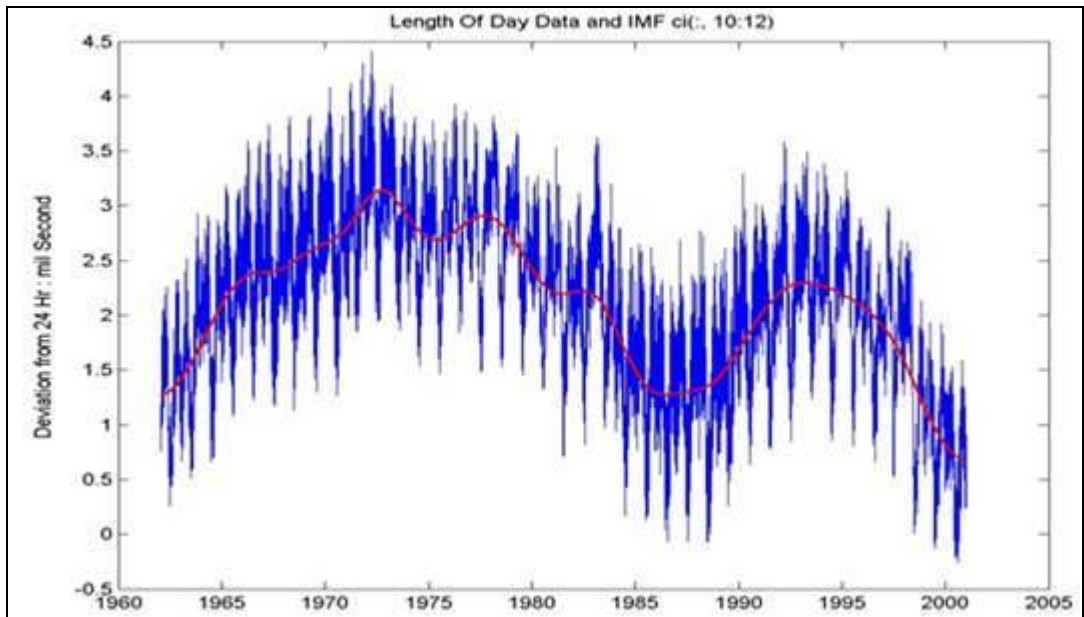


Рис. 24.2.11. Данные и сумма  $c_{10}+c_{11}+c_{12}$ .

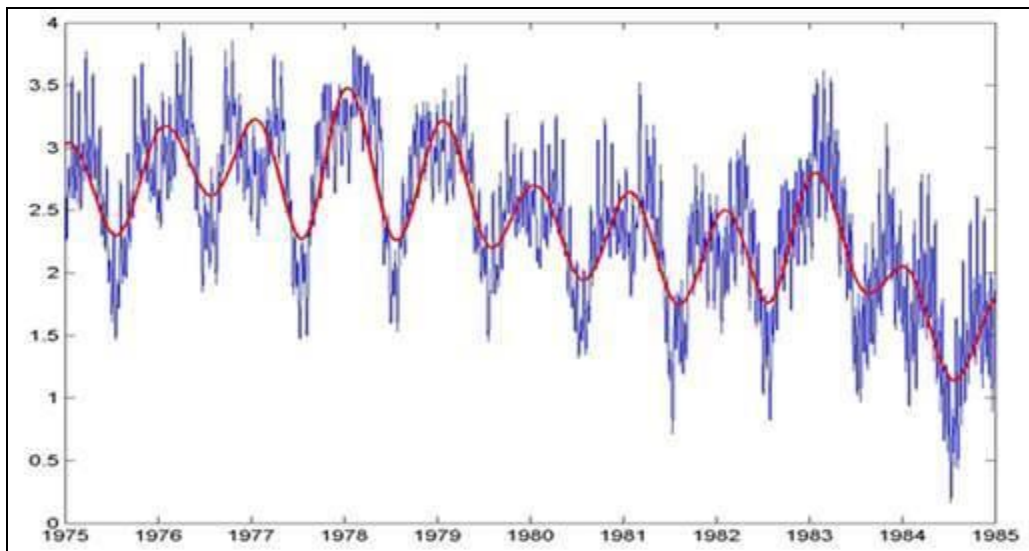


Рис. 24.2.12. Детализация данных и сумма  $c_8+c_9+c_{10}+c_{11}+c_{12}$ .

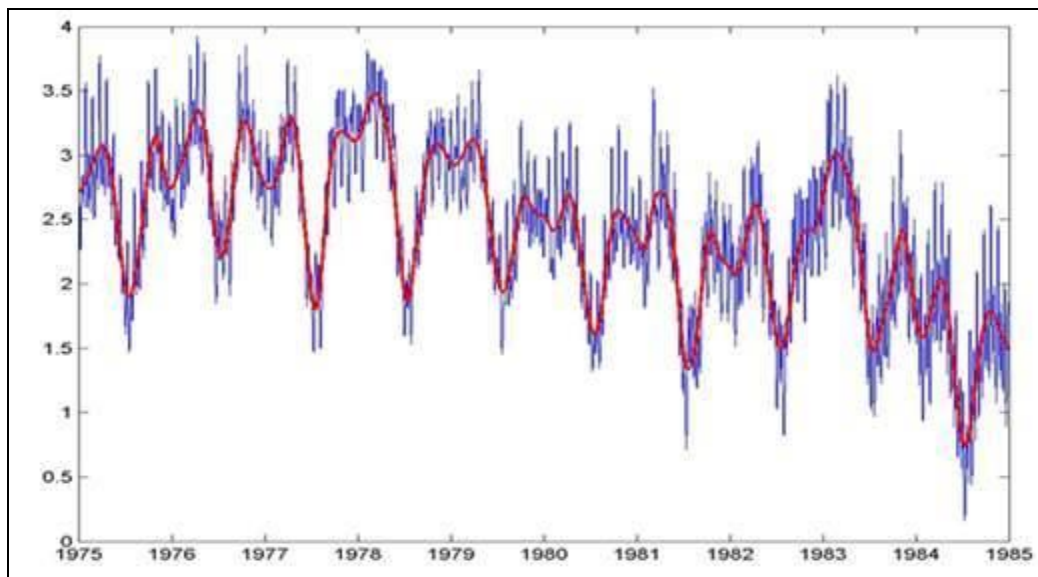


Рис. 24.2.13. Детализация данных и сумма  $c_7+c_8+c_9+c_{10}+c_{11}+c_{12}$

Еще один пример из этой же работы /3/.

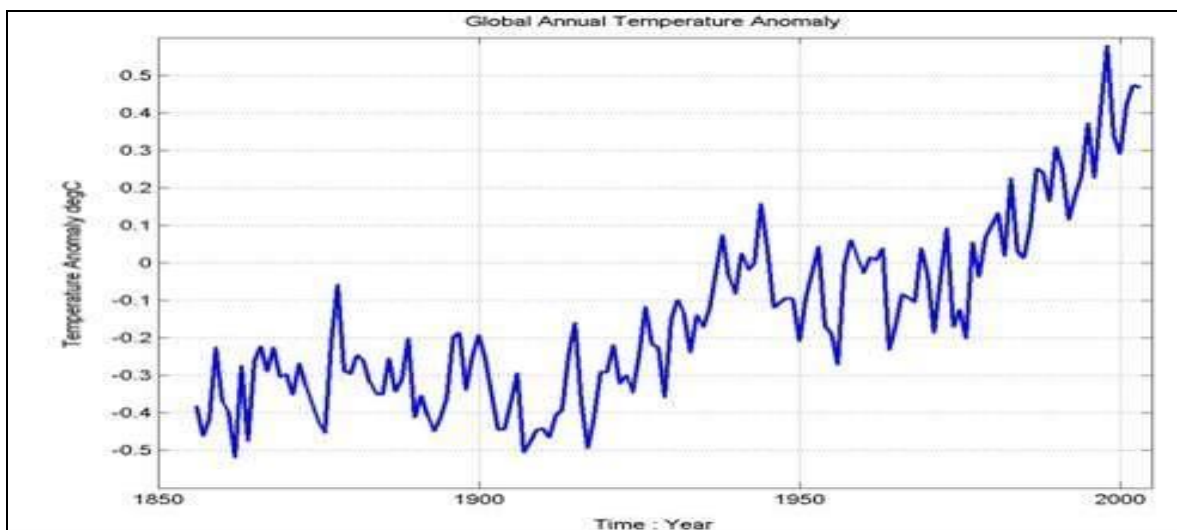


Рис. 24.2.14. Глобальная температурная аномалия. Ежегодные данные с 1856 до 2003.

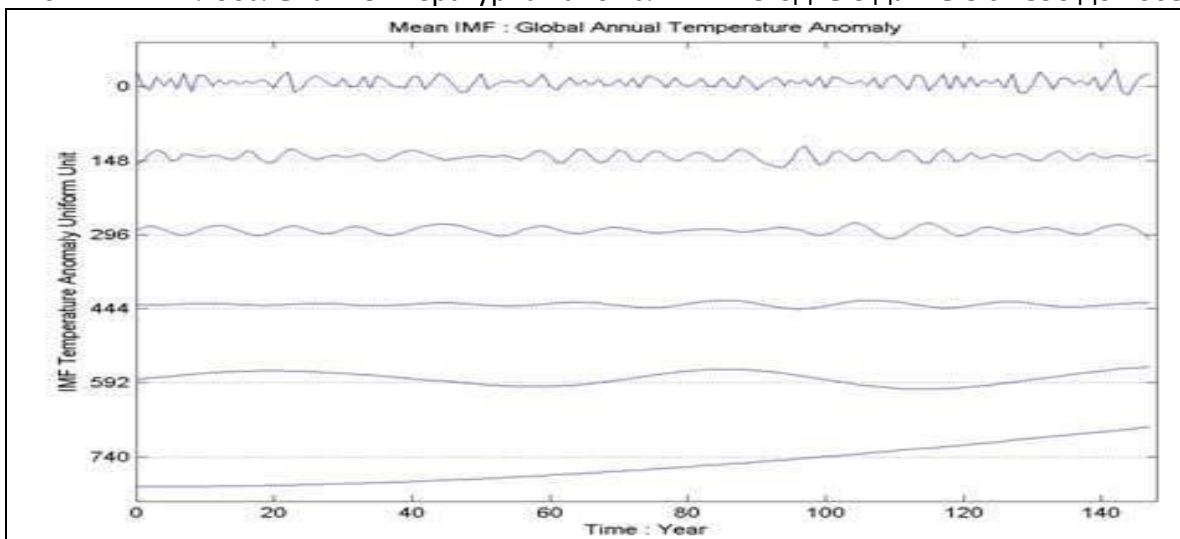


Рис. 24.2.15. Средние значения IMF за 10 просеиваний:  $CC(1000, 1)$

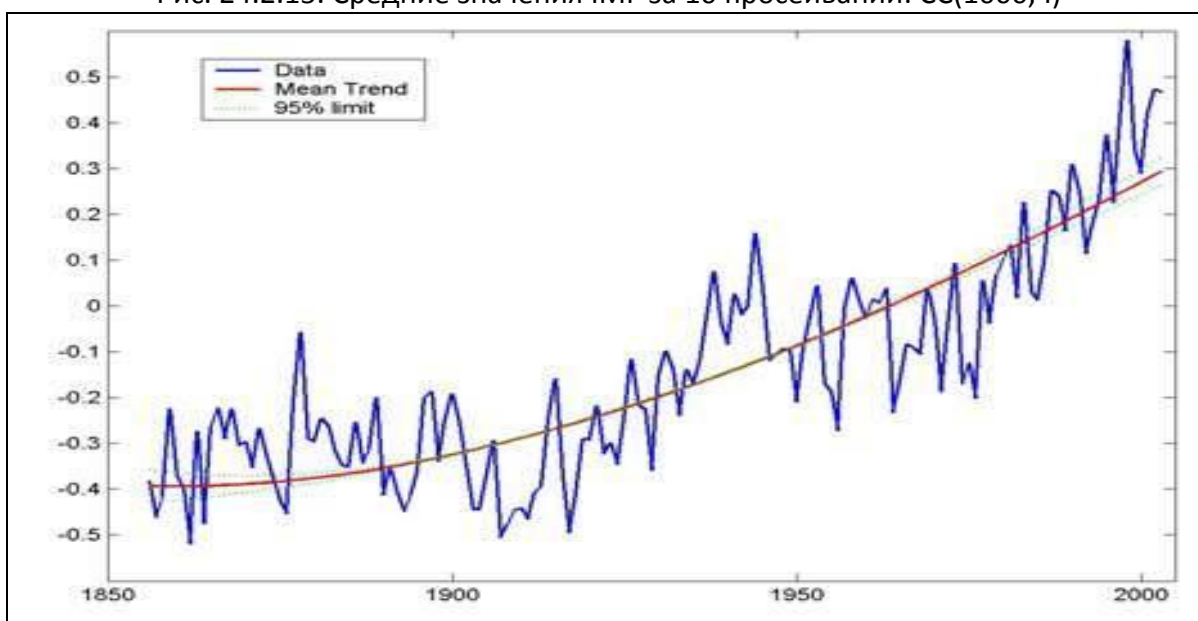


Рис. 24.2.16. Данные и тренд  $S_6$ .



### 24.3. СПЕКТРАЛЬНЫЙ АНАЛИЗ ГИЛЬБЕРТА (HSA)

IMF, определенные вышеприведенным способом, допускают вычисление физически значимых мгновенных частот, что дает возможность создать частотно-временное представление сигнала на основе преобразования Гильберта.

**Спектр мгновенных частот Гильберта.** После выполнения преобразования Гильберта на каждой компоненте IMF первоначальные данные  $x(t)$  могут быть выражены как вещественная часть комплексной формы в следующем виде:

$$x(t) = RE \left[ \sum_{j=1}^n a_j(t) * \exp(i \int \omega_j(t) dt) \right] \quad (24.3.1)$$

Здесь, остаток  $r_n$  не учтен, поскольку это постоянная или монотонная функция. В отличие от преобразования Фурье, здесь и амплитуда  $a_j(t)$ , и мгновенная частота  $\omega_j(t)$  являются функцией от времени. На рис. 24.3.1 приведено сопоставление частотно-временного представления модельного сигнала в трех представлениях.

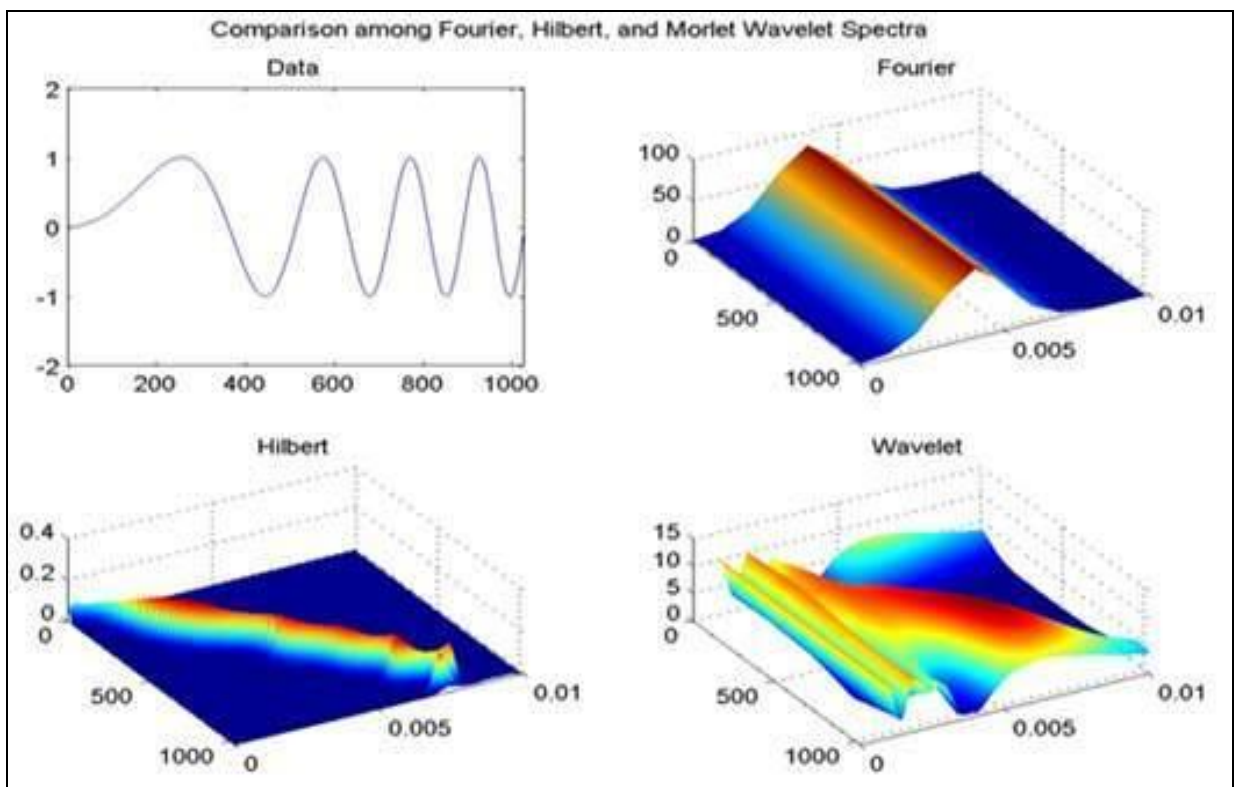


Рис. 24.3.1.

Разрешающая способность по времени спектра Гильберта определяется частотой дискретизации данных, в то время как частотная разрешающая способность произвольна. Самая низкая извлекаемая частота равна  $1/T$ , где  $T$  – интервал задания сигнала, а самая высокая частота  $1/(n\Delta t)$ , где  $\Delta t$  – интервал дискретизации данных, а  $n = 5$  минимальное число отсчетов для точного определения частоты.

**Выбор сплайна** – основной шаг в генерации модовых функций, как базиса для спектрального анализа Гильберта. Обычно используются кубические сплайны, но при их применении могут возникать серьезные проблемы на концевых участках сигналов, где у кубического сплайна могут быть большие колебания. На качество выделения модовых функций влияет также четкость выделения всех экстремумов, которая зависит от частоты дискретизации данных. Для повышения качества процесса EMD и частотной разрешающей способности анализа можно использовать сплайны более высокого порядка, но это связано с увеличением трудоемкости вычислений.

Кроме типа сплайна на точность преобразования Гильберта на конечных интервалах сигналов конечной длины оказывает влияние явление Гиббса в области конечных разрывов сигналов. Чтобы уменьшить оба типа конечных эффектов можно рекомендовать задание начальных условий в начале и в конце сигнала в виде двух волн. Другой эффективный метод - выбирать сигнал так, чтобы конечные эффекты находились вне окна анализа.

#### 24.4. EMD ШУМОВЫХ СИГНАЛОВ

Шумы, сопровождающие полезную информацию в сигнале, в принципе, не относятся к типу колебательных в прямом смысле этого понятия. Но они полностью удовлетворяют приведенным выше определениям функций IMF. При распределении во всем частотном диапазоне входного сигнала и выполнении EMD, они распределяются по всем функциям IMF. Так как информация в главном частотном диапазоне дискретных сигналов обычно является низкочастотной, то шумы «отсеиваются», в основном, в высокочастотные функции IMF. Но в эти функции могут «просачиваться» и высокочастотные спектральные составляющие информационной части сигнала в зависимости от их положения в главном частотном диапазоне. Соответственно, на первый план выдвигается задача формирования определенных критериев отбора только шумовых функций IMF (для исключения их при последующей реконструкции сигнала) и влияния на этот отбор как статистических и спектральных характеристик самих шумов, так и спектрального состава полезной информации в сигнале.

**EMD «белого шума».** Белый шум является стационарным случайным процессом  $q(t)$ , автокорреляционная функция которого описывается дельта - функцией Дирака и, соответственно, спектральная плотность мощности шумов не зависит от частоты и имеет постоянное значение  $Wq(f) = \sigma^2$ , равное дисперсии значений  $q(t)$ . По существу, это идеализированный случайный процесс с бесконечной энергией. Но в случае постоянства спектральной плотности мощности случайного процесса в конечном диапазоне частот это существенно упрощает анализ сигналов. Многие помехи в радиотехнике, в технике связи и в других отраслях техники рассматривают как белый шум, если эффективная ширина спектра сигналов много меньше эффективной ширины спектра шумов и спектральная плотность мощности шумов слабо изменяется в интервале спектра сигнала.

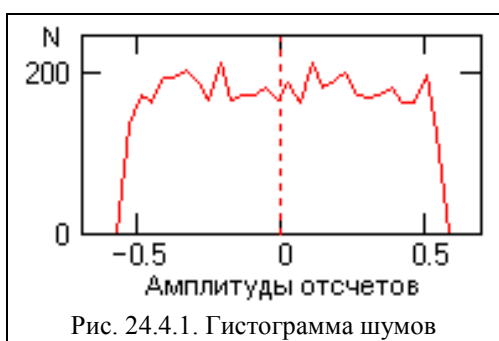


Рис. 24.4.1. Гистограмма шумов

Понятие "белый шум" определяет только спектральную характеристику случайного процесса, а, следовательно, под это понятие подпадают любые случайные процессы, имеющие равномерный энергетический спектр и различные законы распределения. На рис. 24.4.1. Приведена гистограмма единичной реализации модельного «белого шума»  $y(k)$  в системе Mathcad (5000 отсчетов) с равномерным распределением отсчетов от -0.55 до +0.55 и дисперсией 0.1. Спектральную

плотность мощности модельной реализации шума можно посмотреть на рис. 24.4.5.

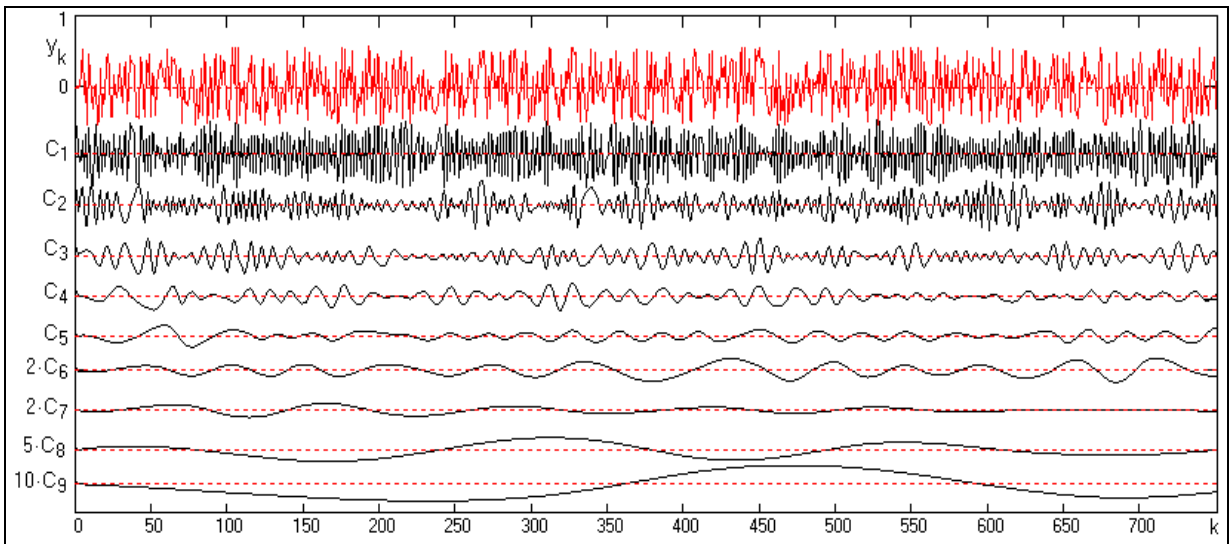


Рис. 24.4.2.

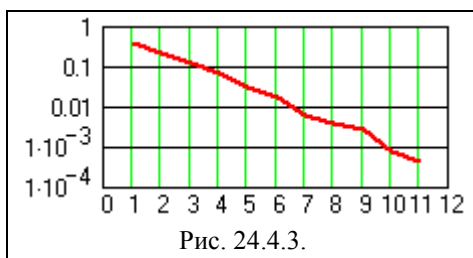


Рис. 24.4.3.

На рис. 24.4.2 приведен результат EMD модели  $y(k)$  шума на первых 750 отсчетах. Шумовой сигнал данной реализации был разложен на 12 функций IMF, первые девять из которых приведены на рисунке (4 последних функции в увеличенном масштабе). Останов процесса декомпозиции выполнен по минимуму погрешности реконструкции без учета

остатка, процесс снижения погрешности при увеличении количества функций IMF приведен на рис. 24.4.3. Количество функций IMF в различных реализациях случайного сигнала изменяется от 8 до 14. Останов итераций при вычислении каждой функции IMF был установлен по относительному расхождению между последовательными итерациями с порогом 0.01%, при этом количество итераций для первых функций IMF достигает 30 и, как правило, постепенно снижается. Погрешность реконструкции с учетом остатка практически равна нулю. Вычислением скалярного произведения любых двух функций IMF можно убедиться в их взаимной ортогональности.

На рис. 24.4.4-А приведены гистограммы первых четырех IMF в сопоставлении с гистограммой входного сигнала  $y(k)$ . Как следует из этих графиков, EMD существенно изменяет плотности распределения выходных функций. Распределение первой IMF становится двумодальным с прогибом вниз на малых (близких к нулевым) амплитудах. Это объясняется тем, что для рядом расположенных однополярных импульсов при EMD выделяются экстремумы импульсов большей амплитуды, которые и отсеиваются в первую функцию IMF. При вычитании этой функции из входного сигнала распределение оставшейся части шумов становится близким к гауссовому с нулевым средним значением и резким сокращением рядом расположенных однополярных импульсов. На отборе всех последующих IMF этот фактор уже не сказывается, и они имеют распределение, близкое к гауссовому, а рядом расположенные однополярные импульсы воспринимаются, как более низкочастотные составляющие шума. Статистика последовательной реконструкции шумового сигнала частично показана на рис. 24.4.4-В.

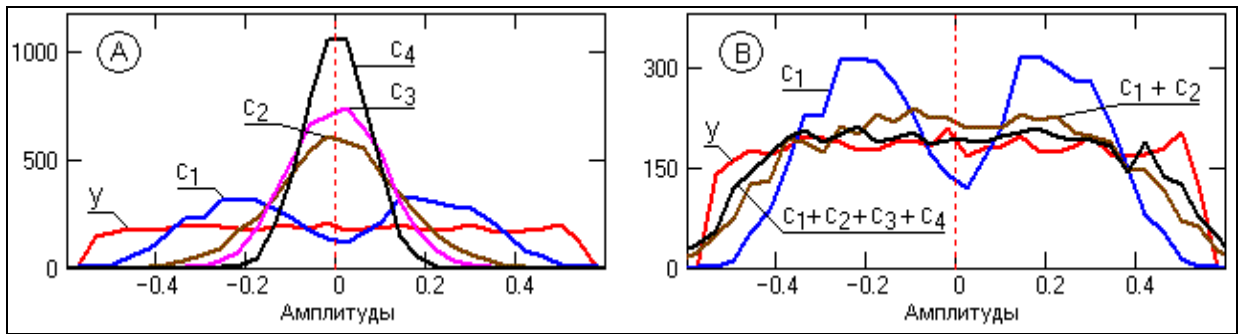


Рис. 24.4.4.

Проверка процесса EMD на шумовых сигналах с другими законами распределения (Гаусса, Пуассона и пр.) показала, что качественный характер процесса остается практически неизменным.

На рис. 24.4.5 приведены спектры плотности мощности сигнала  $y(k)$  и первых семи функций IMF в главном частотном диапазоне сигнала  $0-\pi$  (5000 отсчетов шума с  $\Delta k = 1$ , отсчеты по спектру  $0-2500$  с шагом  $\Delta\omega = \pi/2500$ ).

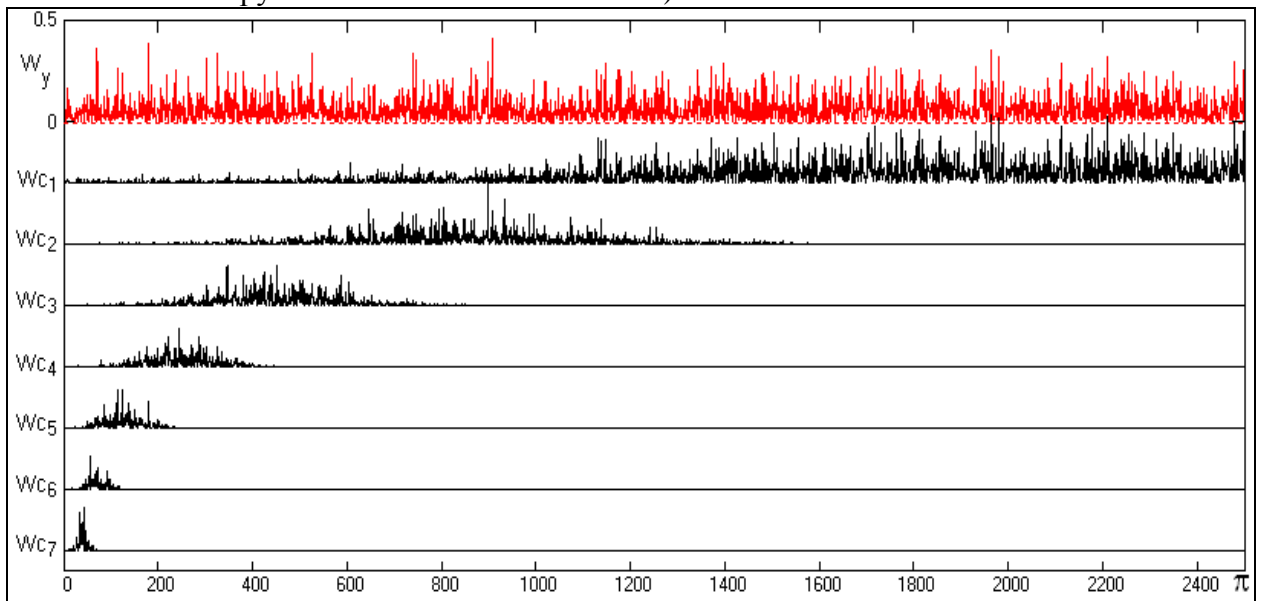


Рис. 24.4.5.

**Частотная избирательность EMD.** По рис. 24.4.5 видно, что процесс EMD обладает вполне определенной частотной избирательностью на каждом уровне EMD. Но говорить о каких-либо частотных передаточных функциях EMD, по-видимому, будет некорректным, так как любая частотная составляющая  $\omega_i$  исходного сигнала в процессе EMD может быть расщеплена по амплитуде и фазе на составляющие разных уровней функций IMF. Это можно видеть на рис. 24.4.6, где приведены графики модулей «эквивалентных» частотных передаточных характеристик разложения для первых пяти функций IMF, полученные осреднением отношения спектров функций к спектру исходного сигнала. Для получения достаточно гладких «эквивалентных» передаточных функций входной белый шум реализовался массивом из 30000 отсчетов, а сглаживание отношения спектров выполнялось в скользящем временном окне 2000 отсчетов.

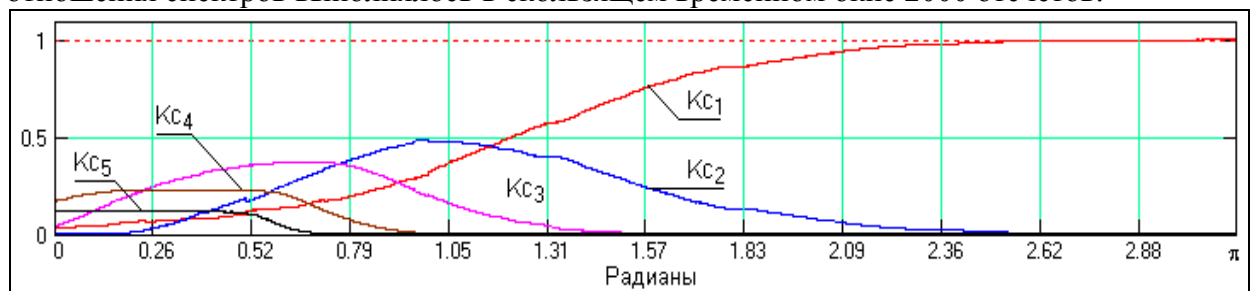


Рис. 24.4.6.

На рис. 24.4.7 приведены графики последовательного суммирования коэффициентов «эквивалентных» передаточных функций, которые показывают процесс последовательного перекрытия всего частотного диапазона входного сигнала.

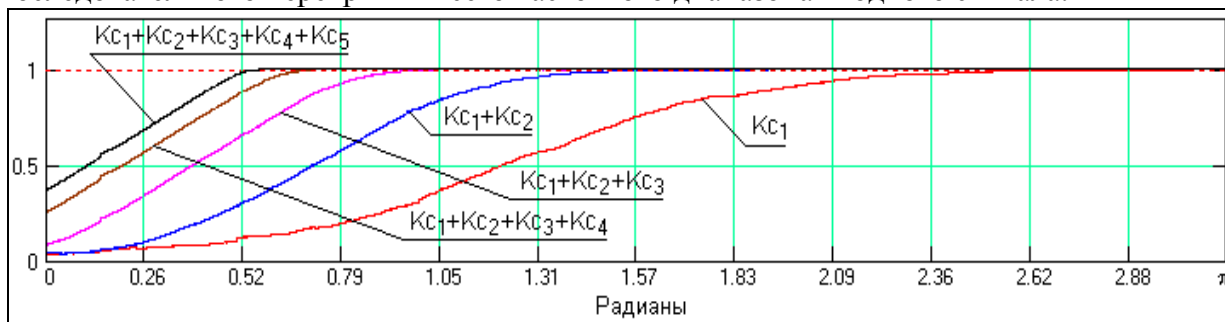


Рис. 24.4.7.

Обратным преобразованием Фурье по спектрам мощности могут быть вычислены нормированные автокорреляционные функции семейства IMF, первые 5 из которых приведены на рис. 24.4.8. Как следует из графиков, статистическая независимость отсчетов в какой-то мере сохраняется только для первой IMF. Но даже в ней появляется отрицательная (знакопеременная) корреляция между последовательными отсчетами. Во всех остальных функциях четко прослеживается появление затухающей косинусоидальной зависимости между отсчетами с последовательным увеличением интервала корреляции по мере увеличения номера IMF.

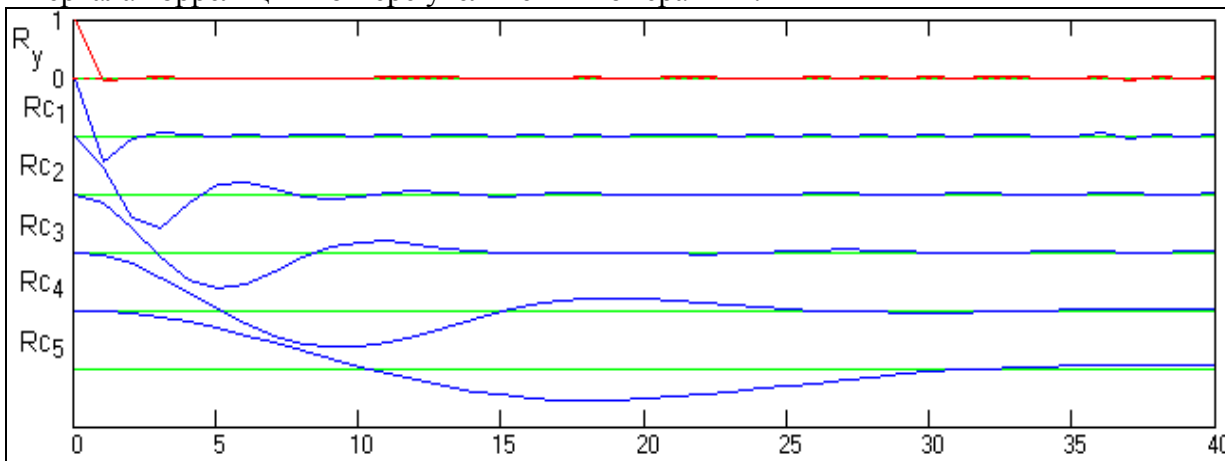


Рис. 24.4.8.

## ТЕМА 25. АНАЛИЗ И ОБРАБОТКА ГЕОФИЗИЧЕСКИХ ДАННЫХ МЕТОДОМ УПРАВЛЯЕМОЙ ЭМПИРИЧЕСКОЙ МОДОВОЙ ДЕКОМПОЗИЦИИ СИГНАЛОВ

Адаптивность формирования ортогонального базиса эмпирической модовой декомпозиции (EMD) сигналов для ряда задач анализа создает определенные трудности. В лекции рассматриваются возможности управления процессом эмпирической модовой декомпозиции при анализе геофизических данных. Для управления процессом декомпозиции используется оперативная информация по спектральному составу сигналов и внутренних модовых функций преобразования.

Задачу создания управляемой (operated) эмпирической модовой декомпозиции (OEMD) разделим на две части:

1. Очистка сигналов от статистических шумов и флуктуаций.
2. Распределение локальных колебательных процессов по уровням IMF.

Разделение вызвано тем, что результаты практически всех методов геофизических исследований природных сред отличаются большими уровнями естественных шумов и статистических флуктуаций измеряемых величин как в силу физической природы самих



регистрируемых процессов (например, в методах ядерной геофизики), так и в силу многочисленных физических микро- и макроструктурных неоднородностей геологических сред и дестабилизирующих факторов регистрации физических данных. Очистка геофизических данных от шумов относится к типовой задаче первичной обработки и подготовки данных к интерпретации. Основная задача на этом этапе – отсеять в первую (шумовую) IMF все составляющие сигнала, которые не несут полезной информации.

Что касается второй задачи, то она является естественным продолжением первой задачи применительно к методам активной геофизики (сейсмические, акустические, электромагнитные и т.п.), имеющими дело с сигналами определенной периодичности. Для решения этой задачи необходимо разработать метод управления процессом EMD, позволяющий целенаправленно изменять формирование IMF на всех уровнях разложения таким образом, чтобы получать максимально адекватное представление о процессах в анализируемых данных при сохранении ортогональности базиса разложения.

### 25.1. ЧАСТОТНОЕ УПРАВЛЕНИЕ ЭМПИРИЧЕСКОЙ МОДОВОЙ ДЕКОМПОЗИЦИЕЙ СИГНАЛОВ

**Модель сигнала.** Для рассмотрения процесса OEMD зададим (рис. 25.1.1) математическую модель зашумленного сигнала  $f_k$  с локальными нестационарными частотными составляющими (сигналами)  $f1_k - f4_k$ , которые в сумме образуют полезную информацию  $f0_k$  и которые необходимо выделить из сигнала в отдельной форме. Мощность шумов соизмерима с мощностью локальных сигналов.

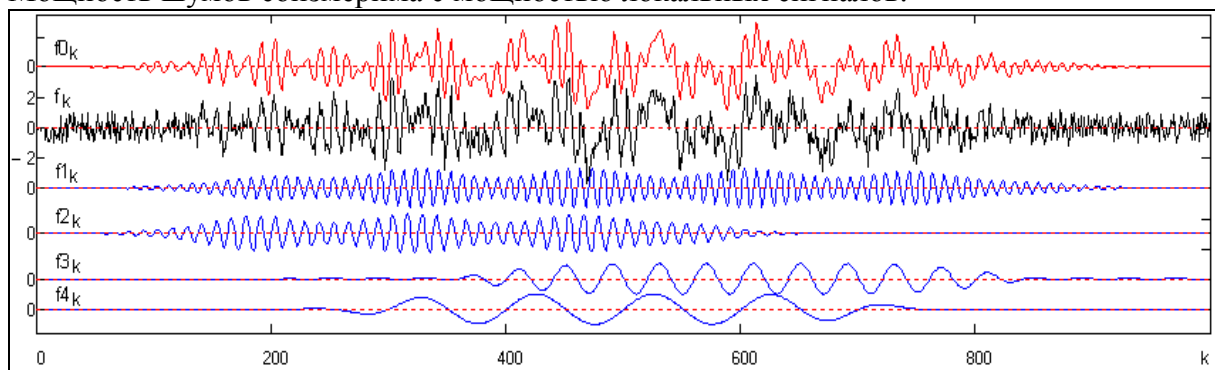


Рис. 25.1.1. Математическая модель сигнала.

Локальные сигналы представляет собой радиоимпульсы с несущими частотами  $\omega1 - \omega4$ , два первых из которых амплитудно-модулированы. Белый шум постоянной мощности распределен по всему частотному диапазону сигнала. Амплитудно-частотная характеристика (АЧХ) сигнала  $f_k$  (модуль частотного спектра в правой части главного частотного диапазона) приведена на рис. 25.1.2. Здесь и в дальнейшем для перевода функций в частотную область представления используется быстрое преобразование Фурье (БПФ).

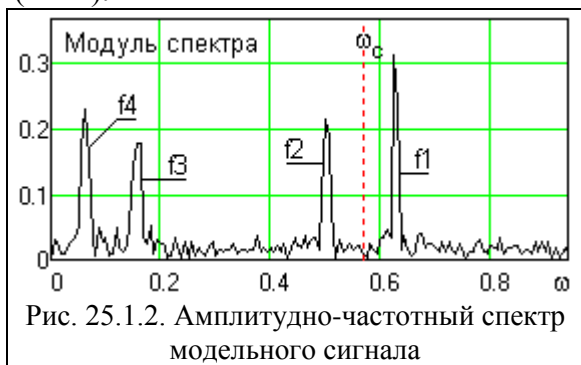


Рис. 25.1.2. Амплитудно-частотный спектр модельного сигнала

**Условия очистки.** В принципе, когда сигнал не зашумлен и в спектре сигнала хорошо выделяются все частотные составляющие (а равно и при априорно известном частотном составе полезной информации), нет необходимости в применении EMD. Разделение частотных компонент может быть выполнено полосовыми частотными фильтрами. Однако для произвольных сигналов при наличии шумов, в

спектре сигнала можно уверенно идентифицировать только главные несущие частоты, а все нестационарности сигнала и боковые модуляционные частоты маскируются шумом.

На результаты EMD влияет также отношение мощности шумов к мощности сигнала. Увеличение мощности шумов вызывает «дробление» и искажение даже монотональных функций IMF, т.к. частотные составляющие шума влияют на процесс отсеивания EMD, и на отдельных временных (координатных) интервалах возникает процесс «просачивания» из соответствующих масштабных функций IMF части главных частот на соседние IMF. При близких несущих частотах стандартное EMD может не разделять радиоимпульсы и отсеивать их в одну IMF даже при анализе чистых сигналов. В зашумленных сигналах может наблюдаться искажение всех локальных составляющих, о чем наглядно свидетельствует EMD сигнала  $f_k$  на рис. 25.1.3.

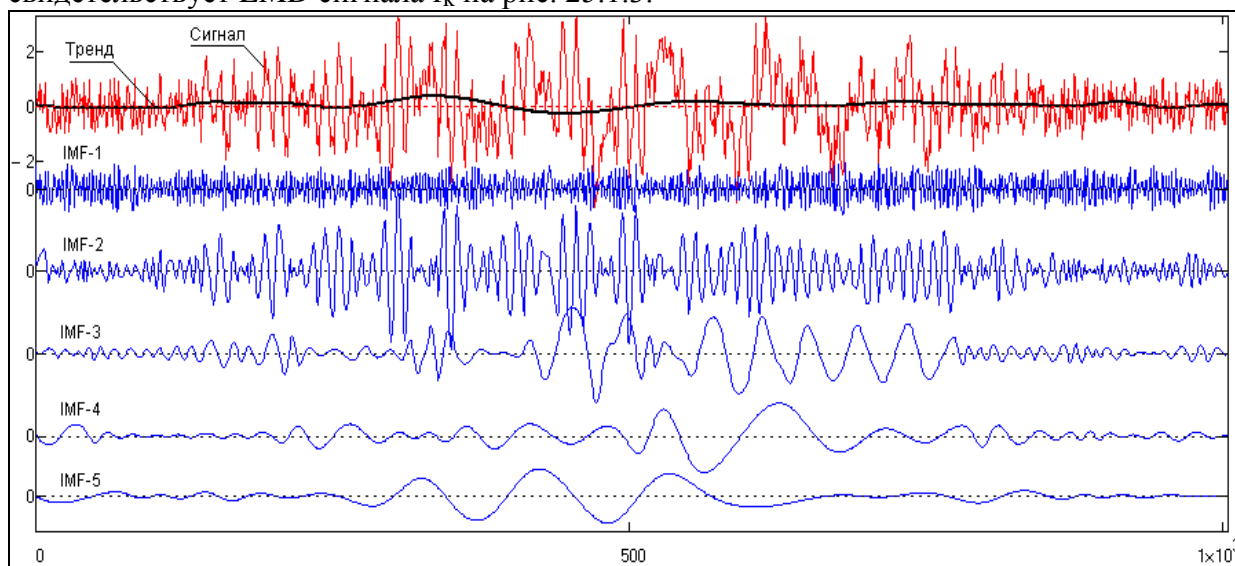


Рис. 25.1.3. Стандартное EMD модельного сигнала

**Принцип управления EMD.** Для разделения радиоимпульсов в различные IMF необходимо вмешаться в процесс EMD, что можно выполнить с использованием априорной информации о спектральном составе сигналов или непосредственно по оперативной информации о спектре сигнала и спектрах выделяемых IMF.

С учетом следования уровней IMF от высоких частот к низким, на первый уровень в IMF-1 отделяются высокочастотные шумы сигнала за пределами высоких частот полезной информации, что будет рассмотрено ниже. На второй уровень в IMF-2 в принятой нами модели сигнала необходимо отделить сигнал  $f_1$ . Для этого низкочастотным НЧ-фильтром с граничной частотой среза  $\omega_c$  ниже частот радиоимпульса  $f_1$  (рис. 25.1.2) отфильтруем все частоты сигнала в интервале  $0-\omega_c$  и используем результат в качестве начальной функции итераций выделения второй IMF [1]. Аналогично на следующий уровень IMF может быть отсеян радиоимпульс  $f_2$ , и т.д.

**Оценка качества.** На модельном сигнале оценку качества разделения сигналов можно проводить вычислением коэффициента взаимной корреляции ( $\rho$ ) между информационными функциями и их IMF-образами или угла расхождения векторов (vectors divergence angle - VDA)  $\varphi = \arccos \rho$  этих функций. Наиболее удобным для сравнения различных вариантов можно считать параметр VDA, который имеет линейный характер изменения своих значений от  $0^\circ$  при полном совпадении функций до  $90^\circ$  при их полной ортогональности (нулевой корреляции).

В этом и заключается сущность частотного управления процессом EMD (метод OEMD), эффективность которого исследуем ниже.

## 25.2. ОЧИСТКА ПЕРИОДИЧЕСКИХ СИГНАЛОВ ОТ ШУМОВ И ФЛЮКТУАЦИЙ

Шумы не относятся к типу колебательных в прямом смысле этого понятия, но они полностью удовлетворяют определениям функций IMF. При распределении спектра шумов по всему частотному диапазону сигнала частотные составляющие шумов распределяются по всем функциям IMF в зависимости от их относительной частотной избирательности. Пример распределения отсева шумов по уровням IMF в главном частотном диапазоне приведен на рис. 25.2.1.

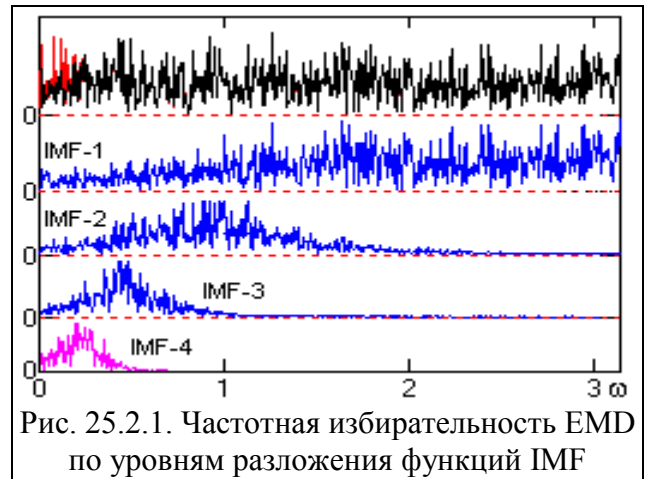


Рис. 25.2.1. Частотная избирательность EMD по уровням разложения функций IMF

**Частотные методы.** Очистка от шумов гармонических, квазигармонических и произвольных периодических сигналов при обработке числовых рядов геофизических данных обычно выполняется низкочастотной фильтрацией сигнала, вейвлет-преобразованием данных с исключением из вейвлетного спектра шумовых уровней разложения, и полосовыми частотными фильтрами при непосредственном выделении периодических сигналов по априорным данным о составе сигнала. Следует заметить, что понятие «шума» является относительным, т.к. шумовая составляющая регистрируемых данных также формируется определенными физическими процессами и может содержать определенную информацию об изучаемых объектах и/или условиях их регистрации, которая в дальнейшем (после выяснения природы и характера шумов или степени их корреляции с какими либо процессами/параметрами среды) может использоваться как для повышения точности интерпретации основной информации, так и для получения совершенно новой информации о среде. С этих позиций при очистке главных информационных сигналов от шумов можно считать целесообразным выделение шумов в отдельную составляющую частотно-координатного разложения сигнала.

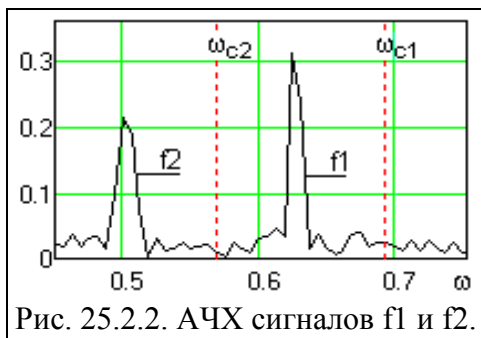


Рис. 25.2.2. АЧХ сигналов f1 и f2.

Если в спектре сигнала хорошо выражена или априорно известна самая высокая частота  $\omega_{\max}$  информационной составляющей, то прямой способ очистки сигнала – вырезание ВЧ-фильтром высокочастотных шумов из спектра сигнала от этой границы ( $\omega_{c1}$  модели сигнала, рис. 25.2.2) до частоты Найквиста  $\omega_N$  с переводом в координатную область в качестве первой функции IMF-1. При использовании для очистки OEMD, с учетом частоты сигнала f1 и частотной избирательности EMD (рис. 25.2.1) для той же операции требуется (3-4)-х кратный последовательный отсев шумов с суммированием результатов отсева в одну функцию IMF-1, если не требуется частотно-координатного анализа шумовых составляющих.

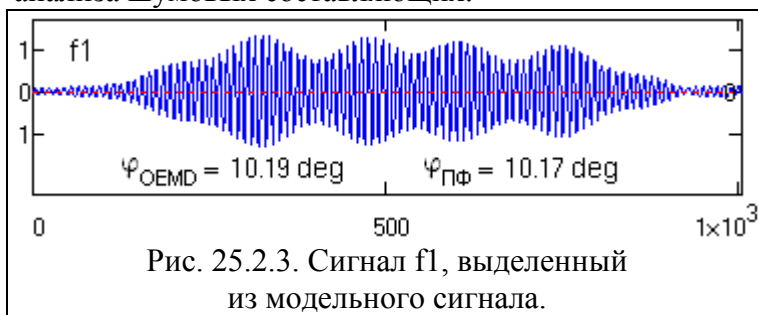


Рис. 25.2.3. Сигнал f1, выделенный из модельного сигнала.

Для сравнительной оценки качества очистки модельного сигнала от шумов при известных частотных границах локального сигнала f1 выполним выделение сигнала в интервале  $\omega_{c1}-\omega_{c2}$  полосовой фильтрацией (ПФ) и методом OEMD (в IMF-2) и вычислим параметр VDA между

выделенными сигналами и модельной функцией f1. Результаты операции приведены на

рис. 25.2.3. Качество очистки от шумов обоих методов практически одинаково.

Но амплитудные, частотные или фазовые изменения информационных частот на пространстве задания сигнала эквивалентны их модуляции и появлению у главных пиков несущих частот боковых частот, которые могут не выделяться на спектре, а их вырезание приведет к частичной потере информации. Так, для нашего модельного сигнала корректная установка среза высокочастотного фильтра на очистку от шумов соответствует (по априорно известному значению частоты модуляции сигнала  $f_1$ ) частоте  $\omega_{c1}$ , что совершенно не фиксируется на спектре сигнала (рис. 25.2.2).

**Ширина переходной зоны частотных фильтров.** Исследуем еще один параметр, непосредственно влияющий на процесс EMD – ширина переходной зоны (ПЗ) фильтра  $\Delta\omega$ .

Применительно к ВЧ-фильтрам для выделения шумов ширина ПЗ является отрицательным фактором. При расширении значения ПЗ влево от  $\omega_{c1}$  на рис. 25.2.2 в выделяемые шумы начинают попадать и исключаться из главной информации модуляционные частоты сигнала  $f_1$ . При использовании OEMD характер влияния ПЗ изменяется на прямо противоположный. При расширении ПЗ вправо от  $\omega_{c1}$  (см. рис. 25.2.4) часть шумов из зоны ПЗ начинает оставаться в интервале сигнала  $f_1$  и параметр VDA начинает увеличиваться, но все модуляционные частоты в сигнале  $f_1$  сохраняются. Этот фактор начинает играть положительную роль, если ширина боковой полосы сигнала  $f_1$  априорно неизвестна и при установке границы  $\omega_{c1}$  НЧ-фильтра может быть допущена ошибка.

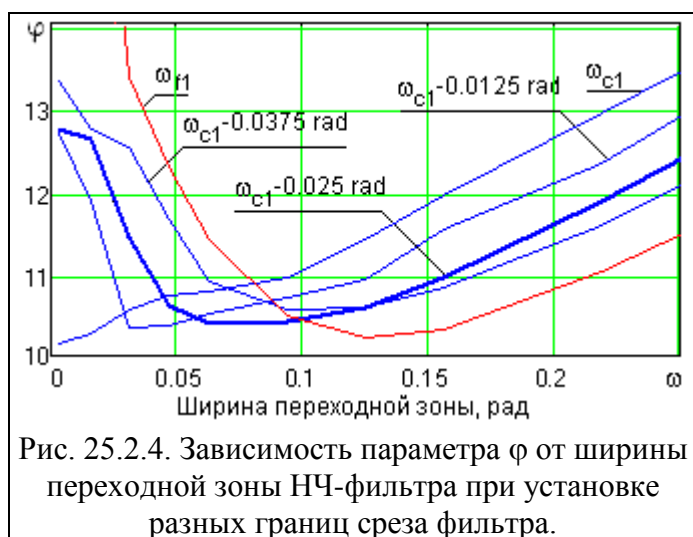


Рис. 25.2.4. Зависимость параметра  $\varphi$  от ширины переходной зоны НЧ-фильтра при установке разных границ среза фильтра.

На рис. 25.2.4 приведены зависимости параметра VDA для модельного сигнала  $f_1$  от ширины переходной зоны НЧ-фильтра (по уровням 0.9-0.1) при разных установках границы среза фильтра в интервале от  $\omega_{c1}$  – верхней границы модуляционных частот, до  $\omega_{f1}$  – несущей частоты сигнала  $f_1$  (сдвиг относительно  $\omega_{c1}$  показан в обозначениях кривых).

При установке границы среза НЧ-фильтра OEMD в пределах модуляционных частот выделяемого сигнала при ширине ПЗ, много меньшей ширины боковой полосы модуляционных частот, параметр VDA за счет потери части модуляционных частот возрастает по отношению к минимально возможному (при границе по  $\omega_{c1}$ ). Но при увеличении ширины ПЗ параметр VDA начинает уменьшаться (уменьшаются потери боковых частот) и достигает определенного минимума (при полном включении боковых частот в состав выделяемого сигнала), близкого к абсолютному минимуму, после чего начинает снова увеличиваться за счет включения в состав сигнала все большей доли шумовых частот. Для полного включения всех боковых частот в состав выделяемого сигнала значение ПЗ должно быть таким, чтобы коэффициент передачи НЧ-фильтра на границе боковых частот был не ниже 0.707 (ширина ПЗ порядка двух значений ширины верхних боковых частот сигнала,  $\Delta\omega \approx 2(\omega_{c1} - \omega_{f1})$ ). Широкие зоны минимумов VDA по зависимости от ширины переходных зон делают процесс OEMD устойчивым к ошибкам установки границ и ширины переходных зон.

**OEMD при очистке от шумов.** Отсюда следует простая логика установки НЧ-фильтров OEMD для очистки от шумов. Граница среза фильтра устанавливается непосредственно после пика максимальной частоты очищаемого от шумов сигнала,

который обычно хорошо фиксируется по спектру сигнала или известен по априорным данным. Ширина переходной зоны устанавливается по априорным данным или по ожидаемым амплитудно-частотным нестационарностям сигнала.

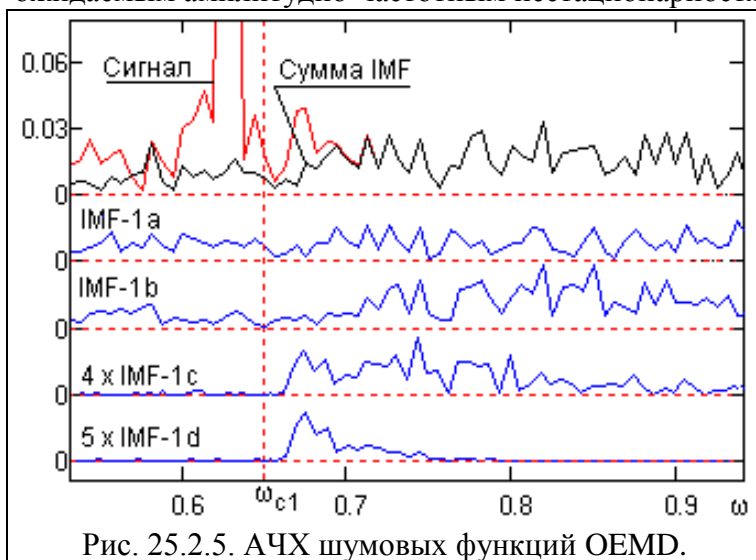


Рис. 25.2.5. АЧХ шумовых функций OEMD.

каких-либо шумовых частот в сигнале, а формируются интерференцией выделенных высокочастотных шумов. Это позволяет обнулить значения спектров в интервале  $0-\omega_{c1}$ , что дает дополнительное уменьшение VDA на (3-5)%.

При очистке от шумов OEMD имеет возможность отдельной установки частоты срезов и ширины переходных зон фильтров при отборе функций IMF-1 (т.е. фильтров с разными параметрами для отсева в IMF-1a, IMF-1b и др.), что позволяет повысить управляемость процессом очистки сигналов с особо сложной структурой боковых частот (в диалоговом режиме контроля за спектрами IMF-1).

### 25.3. ВЫДЕЛЕНИЕ ЛОКАЛЬНЫХ ЧАСТОТНЫХ СОСТАВЛЯЮЩИХ ИНФОРМАЦИИ

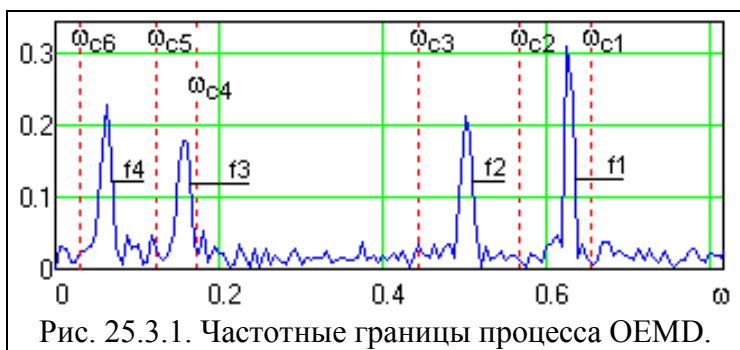


Рис. 25.3.1. Частотные границы процесса OEMD.

$f_k$  (рис. 25.1.1) и установленные частотные границы локализации выделения модовых функций IMF. Результат OEMD тестового сигнала приведен на рис. 25.3.2.

Эффект использования OEMD наглядно виден при сравнении рис. 25.3.2 и 25.1.3. Отметим некоторые особенности выполнения декомпозиции сигналов.

На рис. 25.2.5 приводятся модули спектров 4-х последовательно отсеянных OEMD функций IMF-1 ( $\omega_{c1}=0.65$  rad, ПЗ 0.075 rad), а также спектр суммы этих функций в сопоставлении со спектром сигнала в области локального сигнала  $f_1$  (функции IMF-1c и 1d в увеличенном масштабе).

На рисунке можно видеть, что в функциях IMF появляются низкочастотные составляющие спектров в интервале  $0-\omega_{c1}$ . Эти составляющие не отображают

#### Частотные параметры

**OEMD.** Аналогичная методика установки границ срезов НЧ-фильтров и ширины переходных зон OEMD может применяться при последовательном выделении локальных частотных групп по всему частотному диапазону сигнала. На рис. 25.3.1 приведен модуль спектра тестового сигнала



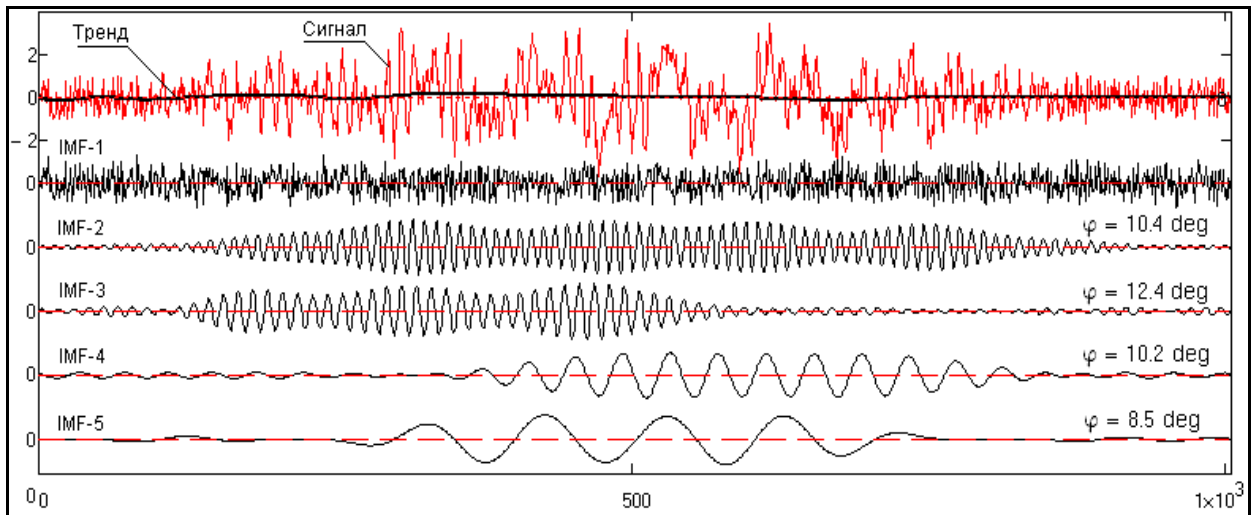


Рис. 25.3.2. Результаты OEMD модельного сигнала.

При выделении модулированных или нестационарных сигналов ( $f_1$  и  $f_2$ ) границы НЧ-фильтров целесообразно устанавливать за пределами боковых полос сигналов с малой шириной переходной зоны ( $\omega_{c3}$  и  $\omega_{c2}$ ). Если ширина боковых полос неизвестна, то можно отодвигать в область более низких частот левые границы выделения соответствующих сигналов ( $\omega_{c3}$  для  $f_2$ ,  $\omega_{c2}$  для  $f_1$ ) с расширением переходной зоны вплоть до несущей частоты сигнала.

Если в спектре сигнала встречаются достаточно протяженные «пустые» зоны (как например, интервал между сигналами  $f_3$  и  $f_2$ ), то целесообразно эти интервалы также отсеивать в отдельные IMF или выделять и суммировать с шумовым IMF-1. В противном случае они войдут в состав последующего выделяемого сигнала ( $f_3$ ).

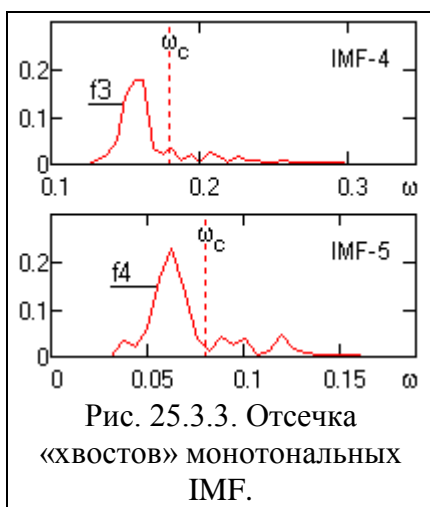


Рис. 25.3.3. Отсечка «хвостов» монотональных IMF.

На рис. 25.3.3 приведены модули спектров IMF-4 и IMF-5. По априорным данным функции  $f_3$  и  $f_4$  однотональные, следовательно, «хвосты» IMF - низкочастотные шумы сигнала. Для IMF-4 это остаточный шум из интервала  $\omega_{c4}-\omega_{c3}$ , не отсеянный в предыдущую «пустую» IMF вследствие его значительного размера, т.е. для полного исключения шума требовался, как минимум, двойной отсев (как при очистке от шумов). Для IMF-5 это шумы межчастотного интервала  $\omega_{f4}-\omega_{f3}$ . «Хвосты» шумов однотональных IMF всегда односторонние, в сторону высоких частот, что позволяет идентифицировать их на спектрах IMF и отсеивать с переводом а шумовую IMF-1.

На рис. 25.3.4 приведено семейство графиков эквивалентной частотной передаточной функции процесса OEMD при выделении всех информационных сигналов  $f_1$ - $f_4$  с разными реализациями наложенного шумового сигнала (8 реализаций).

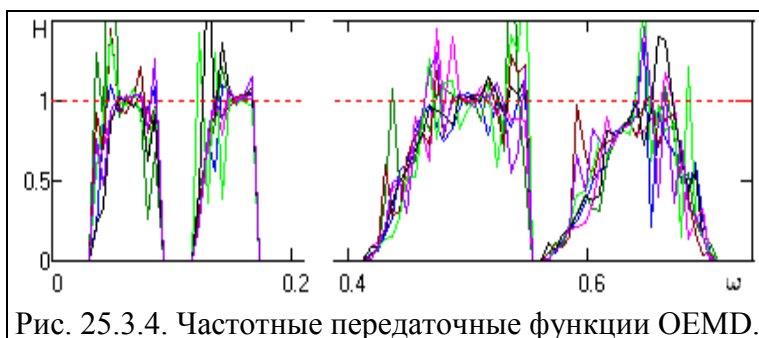


Рис. 25.3.4. Частотные передаточные функции OEMD.

**Передаточные функции OEMD** вычислялись по отношению суммы модулей спектров функций IMF-2, IMF-3, IMF-4 и IMF-5 (рис. 25.3.2) к модулю спектра реализации сигнала  $f_k$  (рис. 25.1.1). Все функции IMF вычислялись в однотипном режиме, описанном

выше, с одной установкой параметров OEMD для всех реализаций сигнала  $f_k$ .

Таблица 25.3.1.

	$\varphi$ , рад	$\overline{\mp\Delta\varphi}$ , рад
$\sum f_N$	15.92	0.72
f1	10.98	0.96
f2	12.13	0.93
f3	10.11	0.98
f4	9.14	0.90

В таблице 25.3.1 приведены данные статистической обработки результатов OEMD этих 8-ми реализаций по углу расхождения с модельными функциями f1-f4 и их суммой.

По этим данным можно сделать заключение, что OEMD сохраняет адаптивность к содержанию исходного сигнала, «изменяя» эквивалентную частотную передаточную характеристику в зависимости от конкретной реализации шумовых сигналов, при этом обеспечивается устойчивость выделения сигналов с погрешностью не более 10%.

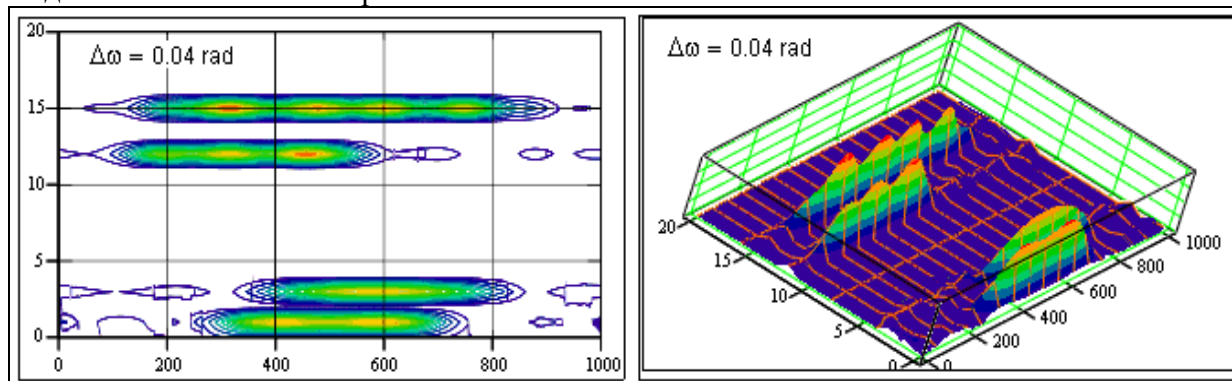


Рис. 25.3.5. Спектр мгновенных частот выделенного сигнала без IMF-1

На рис. 25.3.5 приведено частотно-временное распределение мгновенных частот преобразования (спектр Гильберта-Хуанга) данного модельного сигнала без шумовой IMF-1.

#### 25.4. ОЧИСТКА ОТ ШУМОВ ПРОИЗВОЛЬНЫХ НЕПЕРИОДИЧЕСКИХ СИГНАЛОВ

**Условия очистки.** Рассмотрим применение OEMD для решения типовой задачи геофизических измерений - очистку от шумов произвольных сигналов, представленных числовыми рядами в линейном варианте  $y(x) = f(x, p)$ , где  $x$  – координата среды (профиль съемки, ствол скважины),  $p$  – произвольный зарегистрированный или вычисленный физический параметр среды (интенсивность излучения, напряженность поля, плотность горных пород, и т.п.). В методах регистрации отклика среды при активном воздействии на нее периодическими излучателями (электромагнитные, акустические и т.п.) под  $f(x, p)$  понимается какой-либо физический параметр, выделенный или определенный обработкой из этого отклика.

Полезная информация в таких числовых рядах определяется геологическим строением среды и является преимущественно низкочастотной с гармониками более высоких частот в локальных областях аномалий и резких границ разнотипных геологических сред. Шумовой сигнал, как правило, нелинейный и нестационарный, спектр шумов перекрывается со спектром сигнала, а мощность шумов может быть соизмеримой с мощностью сигнала. В этих условиях задачей первичной обработки данных обычно является подавление шумов с сохранением линейной разрешающей способности интерпретации данных. Но в такой постановке задача относится к числу некорректных, а ее решение во многом имеет субъективный характер и не имеет каких-либо критериев оценки оптимальности для каждой конкретной реализации. Последнее прямо относится к методам выделения полезной информации из данных низкочастотными фильтрами, поскольку требует априорного задания среза фильтров.

Возможности метода OEMD для решения поставленной задачи, с учетом его адаптации под характер сигналов, требуют изучения. Однако задание какой-либо обобщенной математической модели сигнала под поставленные условия весьма

проблематично, и поэтому исследование выполним на реальных каротажных диаграммах.

**Частотный диапазон информации.** Стандартный линейный интервал дискретизации при цифровой регистрации данных в скважинной геофизике составляет 10 см, при этом требования к пространственной разрешающей способности интерпретации данных по минимальной мощности пластов обычно не менее 50-100 см. Отсюда следует, что ожидаемая полезная информация является низкочастотной и занимает не более 5-10% главного частотного диапазона измерений. Естественно, что этот диапазон может быть и много меньше, если геологический разрез по скважине представлен пластами (более или менее однородными интервалами) с гораздо большей мощностью. Но в любом случае следует ожидать, что очистка от шумов потребует проведения трехкратного отсева шумов, т.е. формирование  $IMF-1 = IMF1a + IMF1b + IMF1c$ .

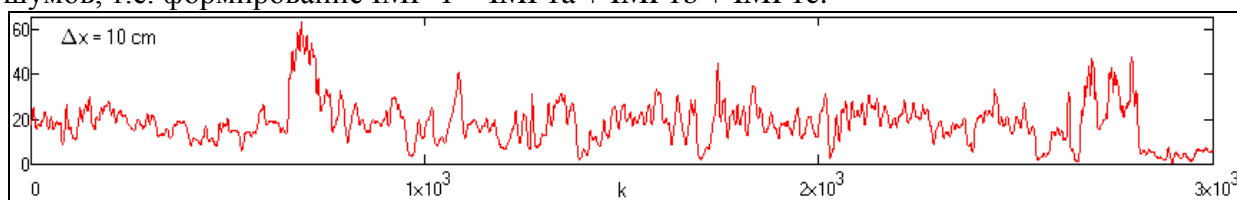


Рис. 25.4.1. Диаграмма нейтронного каротажа скважины

**Очистка от высокочастотных шумов.**

На рис. 25.4.1 приведен пример диаграммы нейтронного каротажа. Спектр сигнала приведен на рис. 25.4.2. Сделать какое-либо заключение о границе информационной части сигнала по данному спектру не представляется возможным. Для установки границы шумов была использована априорная информация о минимальной мощности однородных интервалов скважин порядка 0.7-1 м, что согласуется с визуальным контролем примерно однородной мощности шумов на спектре сигнала с порога порядка 0.2 рад.

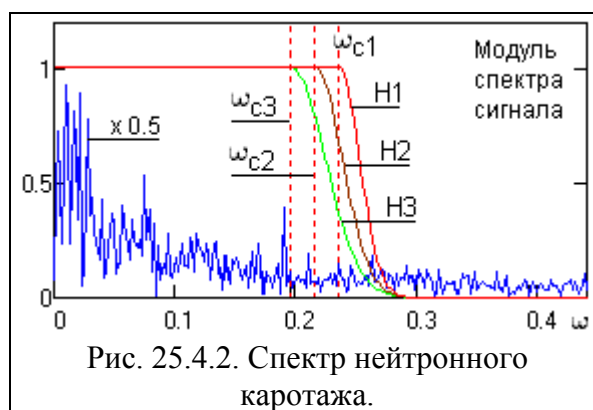


Рис. 25.4.2. Спектр нейтронного каротажа.

Задаем границы трех фильтров Н1-Н3 при  $\omega_{c1} > \omega_{c2} > \omega_{c3}$  с шагом между границами срезов  $\Delta\omega_c = 0.02$  рад (можно рекомендовать порядка (5-10)% ожидаемой ширины спектра информационной части сигнала). Ширина переходных зон фильтров установлена равной  $3\Delta\omega_c$  для Н1,  $4\Delta\omega_c$  для Н2 и  $5\Delta\omega_c$  для Н3, при этом граница полного подавления высокочастотных составляющих для всех трех фильтров примерно одна и та же. Входной и очищенный сигнал, а также выделенные шумы, приведены на рис. 25.4.3.



Рис. 25.4.3. Входной сигнал, выходной сигнал, и выделенные шумы нейтронного каротажа.

На спектре сигнала (рис. 25.4.2) довольно четко отмечается также группа низких частот в интервале 0-0.04 рад. На рис. 25.4.4 приведен результат очистки сигнала от шумов при сдвиге порогов всех фильтров на 0.15 рад влево. Естественно, что при этом



увеличилась степень сглаживания сигнала и уменьшилась линейная разрешающая способность. Но какой именно из очищенных сигналов, на рис. 25.4.3 или 25.4.4, принять для дальнейшей обработки и интерпретации, можно решить уже только на этапе сопоставления с другими геофизическими и геологическими данными по этому интервалу скважины.

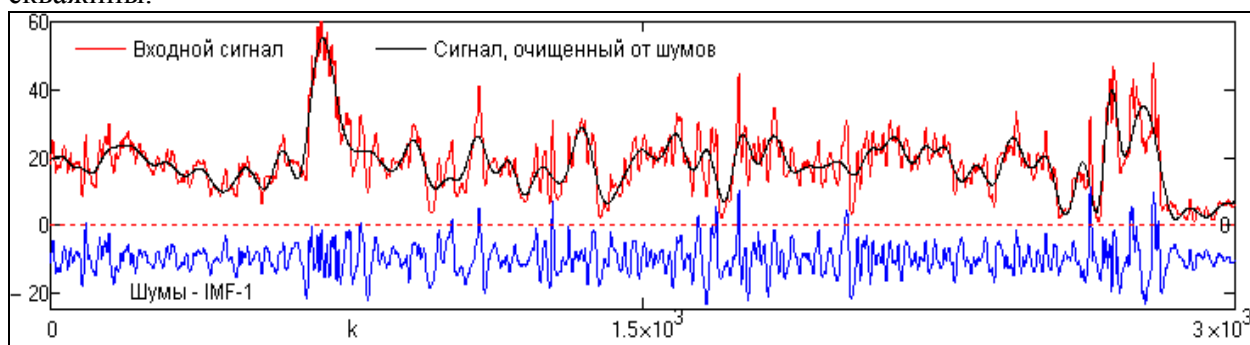


Рис. 25.4.4. Входной сигнал, выходной сигнал, и выделенные шумы нейтронного каротажа при сдвиге среза НЧ-фильтра в область низких частот.

На рис. 25.4.5. приведен еще один пример - диаграмма резистивиметрии (Mud Resistivity). Диаграмма центрирована для обработки.

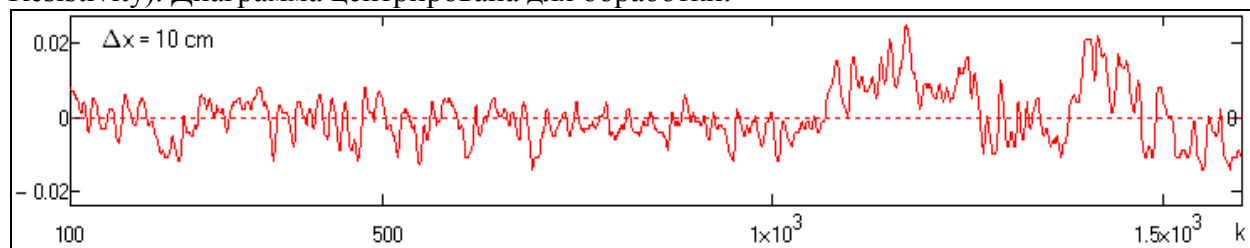


Рис. 25.4.5. Центрированная диаграмма резистивиметрии по стволу скважины.

Информационный сигнал резистивиметров обычно является еще более низкочастотным и занимает не более 3-5% главного диапазона на уровне мощных шумов по всему диапазону спектра. В этих условиях можно использовать фильтры отсева шумов с равными частотами срезов и переходных зон. Модули спектров сигнала и передаточных функций НЧ-фильтров приведены на рис. 25.4.6.



Рис. 25.4.6. АЧХ резистивиметрии

На рис. 25.4.7 и 25.4.8 приведены очищенные от шумов диаграммы с порогами отсева  $\omega_c = 0.066$  и  $0.1$  рад. Параллельно очистка сигнала проводилась низкочастотными фильтрами с теми же границами среза  $\omega_c$ .

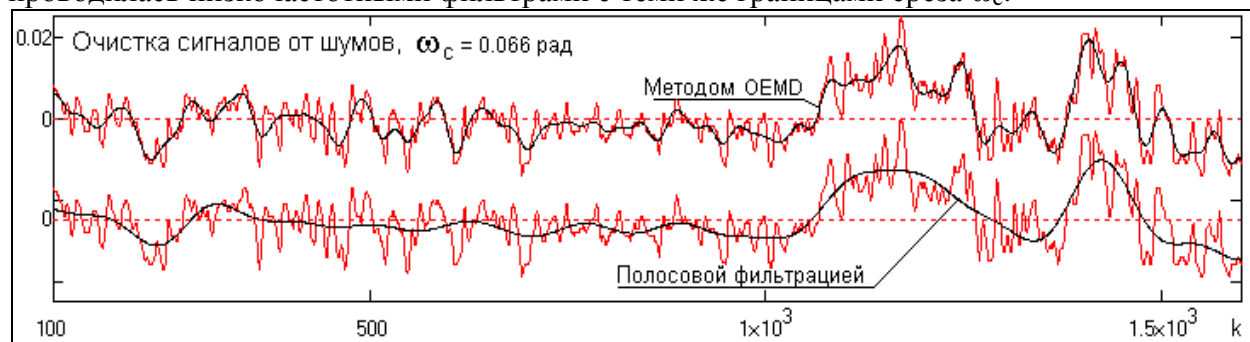


Рис. 25.4.7. Очистка диаграммы резистивиметрии от шумов, срез НЧ-фильтра 0.066 рад.

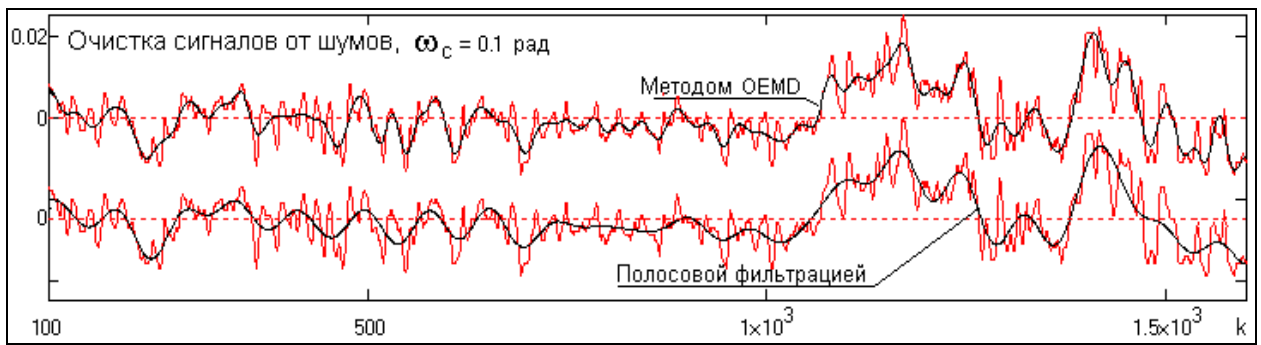


Рис. 25.4.8. Очистка диаграммы резистивиметрии от шумов, срез НЧ-фильтра 0.1 рад.

Как можно видеть по результатам сравнения диаграмм, метод OEMD обеспечивает более высокую разрешающую способность очищенного сигнала, причем устойчивость очистки, в отличие от частотной фильтрации, сохраняется при достаточно большом изменении порога отсева шумов и вариациях переходных зон.

Это определяется тем, что при отборе шумов процесс OEMD, протекающий в координатном пространстве, в большей степени учитывает динамику локальных неоднородностей распределения отсчетов (локальную статистику отсчетов), в отличие от частотной фильтрации. Свидетельством этого являются гистограммы выделенных «шумовых» сигналов, приведенные на рис. 25.4.9.

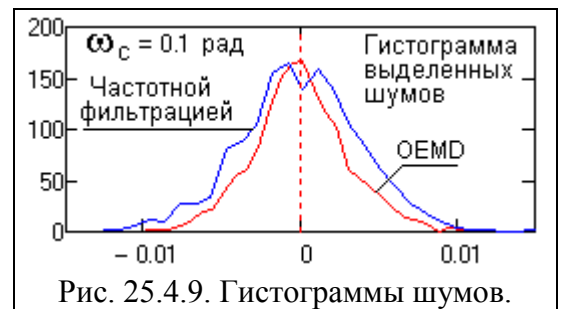


Рис. 25.4.9. Гистограммы шумов.

**Оптимальные границы отсева шумов.** Однако и визуальные, и априорные методы установки границ отсева шумов в определенной степени являются субъективными. Возможность установки оптимальных границ фильтров на основе объективных оценок динамики спектра сигнала рассмотрим на реальном и достаточно типовом сигнале - каротажной диаграмме ПС (Spontaneous Potential)  $f(k)$ ,  $k = 0 \dots K$ ,  $\Delta k = 10$  см, приведенной на рис. 25.4.10.

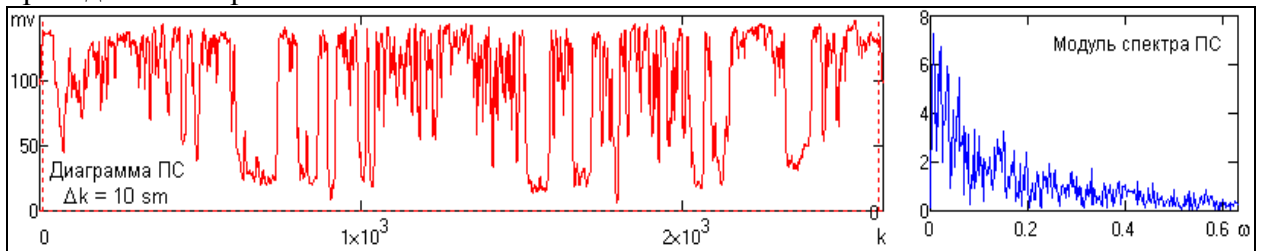
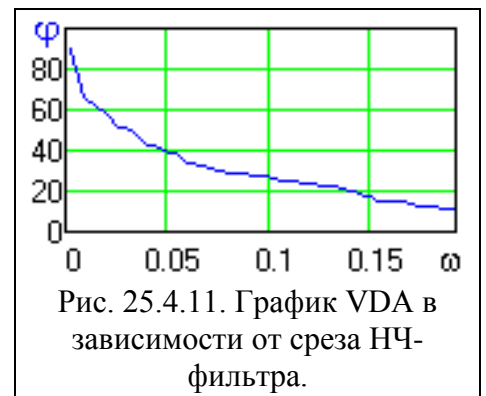


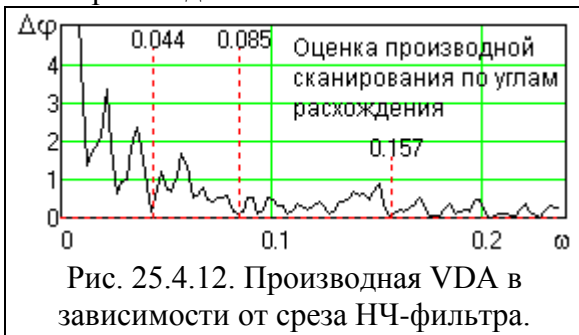
Рис. 25.4.10. Каротажная диаграмма ПС и модуль ее спектра.

Как видно на графике спектра ПС в низкочастотной части главного частотного диапазона, какого-либо визуального критерия установки частоты среза  $\omega_c$  низкочастотного фильтра для очистки сигнала от шумов в данном случае не имеется. Оценку регулярности и значимости информации, выделяемой НЧ-фильтром, можно выполнить по углу расхождения векторов (VDA) отфильтрованной информации и исходного сигнала  $f_k$  методом «Последовательного Расширения окна Фильтра» (ПРФ). Метод можно применять как в координатном, так и в частотном представлении сигнала. ПРФ по спектру является более быстрым.

Просканируем низкочастотную часть спектра сигнала НЧ-фильтром с последовательным сдвигом границы среза фильтра с интервалом  $\Delta\omega$  ( $\Delta\omega_{\min} = 2\pi/(K+1)$ ), начиная с первого отсчета спектра. VDA выделенной части спектра вычисляем с полным спектром сигнала, или (для повышения чувствительности) со спектром сигнала с обнуленной высокочастотной шумовой частью. Полученный график углов расхождения приведен на рис. 25.4.11.

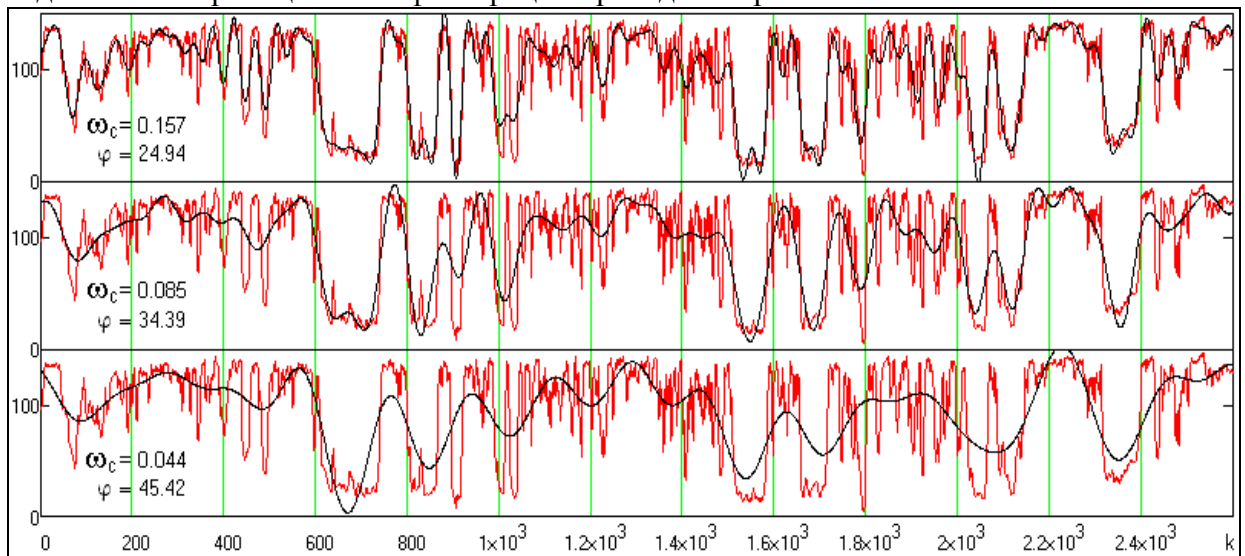


Интегральный характер вычисления VDA резко снижает влияние статистических шумов на результаты вычислений. Угол  $\varphi$  максимален для первого положения фильтра при  $\omega_c = \Delta\omega$  и постепенно уменьшается по мере увеличения  $\omega_c$  (в пределе стремится к 0). Но в силу неравномерности спектра это уменьшение также неравномерно и замедляется после пересечения срезом фильтра наиболее значимых регулярных гармоник спектра. Положение этих замедлений может быть зафиксировано по локальным минимумам производной VDA.



На рис. 25.4.12 приведен график оценки производной VDA (рис. 25.4.11). После прохождения границей среза фильтра всех регулярных гармоник информационной части сигнала при постоянной мощности статистических шумов по спектру сигнала градиент уменьшения углов расхождения также стремится к постоянному статистически флуктуирующему значению и кривая оценки производной выполаживается.

По графику на рис. 25.4.12 могут быть отмечены две возможно оптимальных границы низкочастотной фильтрации  $\omega_c = 0.085$  и  $0.157$  рад, и одна граница с явной утратой части полезной низкочастотной информации. Естественно, что какая-то часть информации будет теряться и при границе  $\omega_c = 0.085$ , но ее значимость может быть оценена только интерпретатором по критерию степени сглаживания кривой при сохранении линейной разрешающей способности выделения характерных информационных точек. Общий характер степени сглаживания диаграммы с выделенными границами НЧ-фильтрации приведен на рис. 25.4.13.



**Устойчивость отсева шумов.** Метод OEMD позволяет производить более гибкую

настройку очистки от шумов с визуальным контролем результатов очистки. Для рассматриваемых данных оказалось достаточным применить трехкратный отсев шумов в функцию IMF-1. Пример нескольких вариантов очистки с параметрами настройки приведен на рис. 25.4.14.

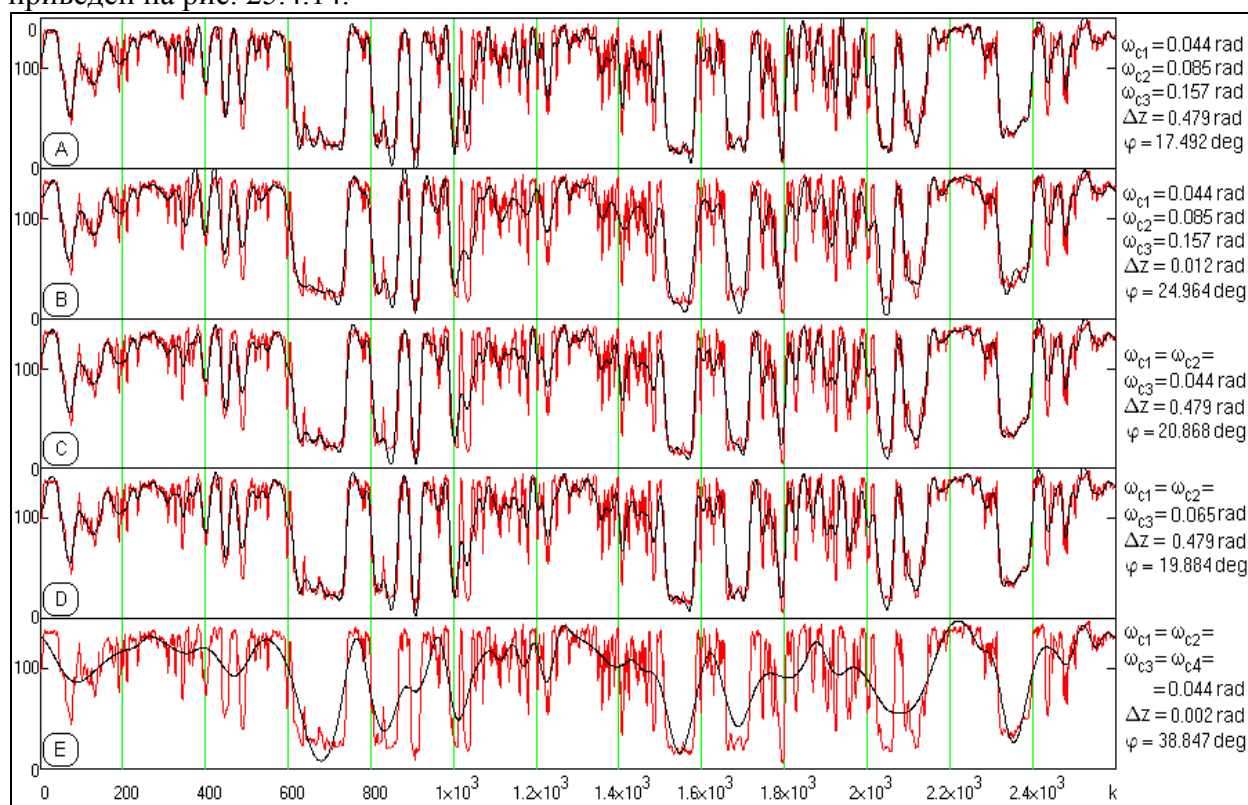


Рис. 25.4.14. Очистка диаграммы ПС от шумов при разных настройках параметров НЧ-фильтра.

С учетом малой критичности OEMD к границам среза фильтров относительно информационной части сигнала, срезы фильтров  $\omega_c$  были установлены по графику на рис. 25.4.12 с нарастанием от первого локального минимума (0.044 рад) до третьего (0.157 рад) с шириной переходной зоны  $\Delta z = 0.479$  рад, одинаковой для всех трех фильтров (кривая А на рис. 25.4.14). При сравнении с первой кривой на рис. 25.4.13 наглядно видна ее более высокая детальность выделения разнородных локальных объектов. Изменением ширины переходной зоны можно изменять степень сглаживания кривой (кривая В). Адаптивность метода позволяет ему достигать практически таких же результатов при одной установке границы среза для всех фильтров (кривая С), при этом допускаются существенные вариации значения границы среза как в большую (кривая D), так и в меньшую стороны.

При включении четвертого фильтра отсева шумов в IMF-1 степень сглаживания сигнала существенно увеличивается (кривая E), но с этого уровня она также может регулироваться изменением границ фильтров и ширины их переходных зон.

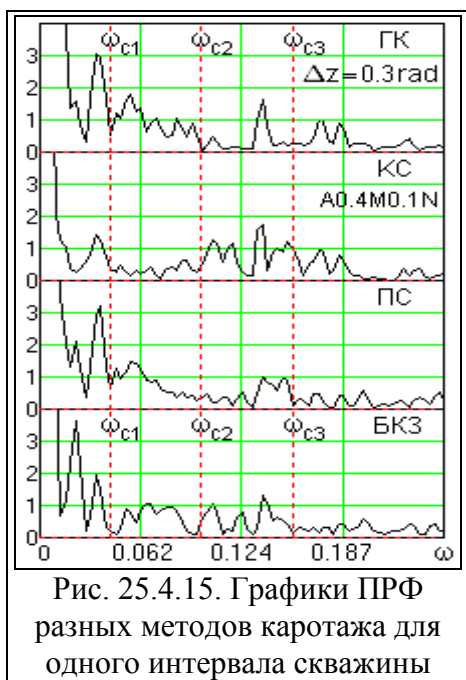


Рис. 25.4.15. Графики ПРФ разных методов каротажа для одного интервала скважины

В целом по результатам сравнения диаграмм можно сделать вывод, что метод OEMD обеспечивает более высокую разрешающую способность очищенного сигнала, причем устойчивость очистки, в отличие от частотной фильтрации, сохраняется при достаточно большом изменении порога отсева шумов и вариациях значений переходных зон.

Частотный диапазон информационной части каротажных диаграмм на изучаемых интервалах скважин определяется геологическим разрезом по стволу скважины и практически одинаков для всех методов каротажных исследований, направленных на детализацию геологического разреза. Это позволяет оптимальные частотные параметры фильтров для управления процессом EMD устанавливать по какому-либо методу с наиболее выразительным характером графика ПРФ и применять эти параметры (учитывая высокую степень адаптации процесса OEMD к характеру сигналов) для обработки других методов каротажа по данному интервалу скважины, а также других

интервалов и скважин с аналогичным геологическим строением. Пример графиков ПРФ по каротажным диаграммам одного интервала скважины комплекса ГК, КС, ПС и БКЗ (1 зонд) приведен на рис. 25.4.15, а на рис. 25.4.16 приведены результаты очистки от шумов этого комплекса диаграмм с параметрами фильтрации, установленными по диаграмме ГК.

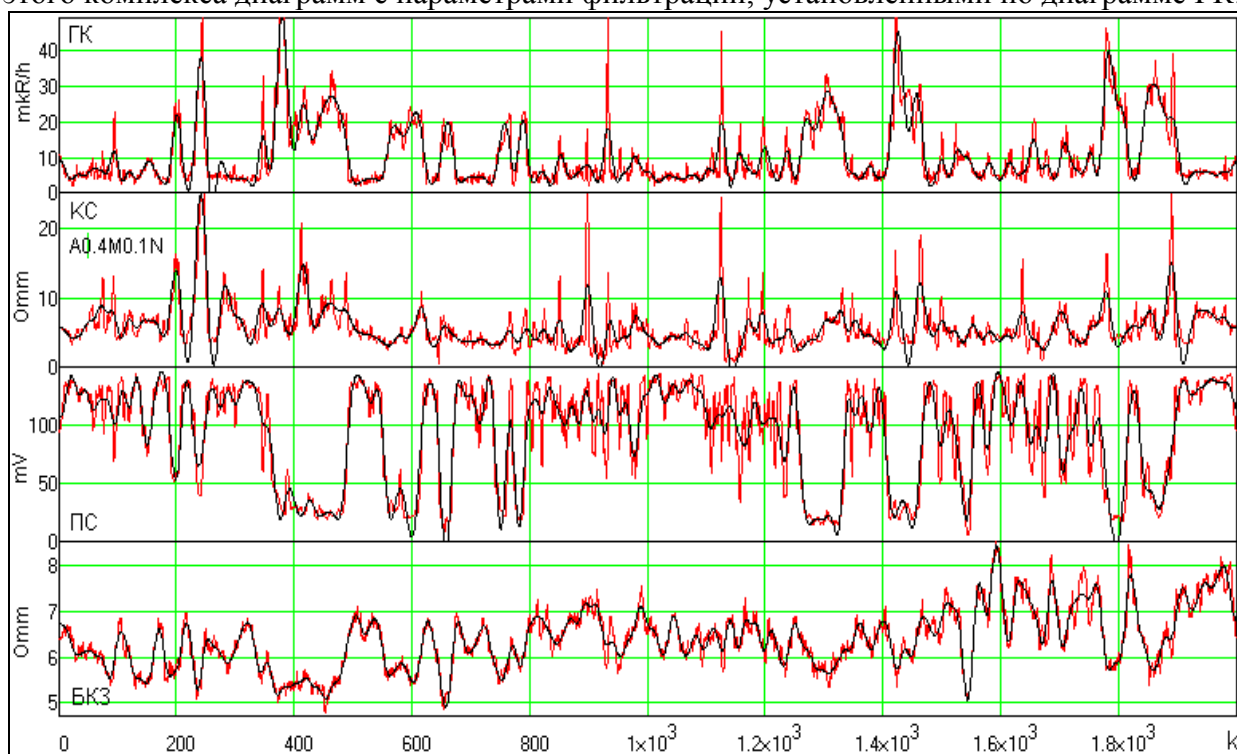


Рис. 25.4.16. Результаты OEMD разных методов каротажа при однотипной установке параметров НЧ-фильтра

Все вычисления OEMD выполнены программой, разработанной в среде Mathcad [3]. С дополнительными материалами по преобразованию Гильберта-Хуанга и эмпирической модовой декомпозицией сигналов можно познакомиться в работах [2].

### Заключение

Изложенная методика свободна как от физической сущности обрабатываемых данных, так и статистики распределения шумовых сигналов и дестабилизирующих

флюктуаций, и может эффективно использоваться для очистки от шумов геофизических данных практически всех геофизических методов и детального инструментального анализа информации в активных методах геофизики.

## ЛИТЕРАТУРА

### Основная

1. Цифровая обработка сигналов. Автор: Ричард Лайонс. Переводчик А. Бритов. Издательство: Бином-Пресс. ISBN 978-5-9518-0446-4, 0-13-108989-7; 2012 г.
2. Цифровая обработка сигналов. Моделирование в Simulink. Автор: А. И. Солонина. Издательство: БХВ-Петербург. ISBN 978-5-9775-0686-1; 2012 г.
3. Устройства генерирования, формирования, приема и обработки сигналов. Автор: О. В. Головин. Издательство: Горячая Линия - Телеком. ISBN 978-5-9912-0196-4; 2012 г.
4. Цифровая обработка сигналов. Автор: Jesse Russell. ISBN 978-5-5129-0330-8; 2012 г.
5. Сергиенко А.Б. Цифровая обработка сигналов: Учебник для вузов. – СПб.: Питер, 2003. – 608 с.
6. Солонина А.И. и др. Основы цифровой обработки сигналов. Учебное пособие. – СПб.: БХВ Петербург, 2005. – 768 с.
7. Давыдов А.В. Цифровая обработка сигналов. Тематические лекции: Учебное пособие в электронной форме. – Екатеринбург, УГГУ, ИГиГ, каф. ГИН. – <http://www.prodav.narod.ru/dsp/index.html>.

### Дополнительная

1. Антонию А. Цифровые фильтры: анализ и проектирование. – М.: Радио и связь, 1983. – 320 с.
2. Баскаков С.И. Радиотехнические цепи и сигналы: Учебник для вузов. - М.: Высшая школа, 1988.- 448 с.
3. Бат М. Спектральный анализ в геофизике. - М.: Недра, 1980.- 535 с.
4. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. – М.: Мир, 1989. – 540 с.
5. Блейхут Р. Быстрые алгоритмы цифровой обработки сигналов. – М.: Мир, 1989. – 448 с.
6. Вычислительная математика и техника в разведочной геофизике: Справочник геофизика. – М.: Недра, 1990. – 498 с.
7. Гольденберг Л.М. и др. Цифровая обработка сигналов: Справочник. - М.: Радио и связь, 1985.- 312 с.
8. Гольденберг Л.М. и др. Цифровая обработка сигналов: Учебное пособие для вузов. - М.: Радио и связь, 1990.- 256 с.
9. Гутников В.С. Фильтрация измерительных сигналов. – Л.: Энергоатомиздат, 1990. – 192 с.
10. Даджион Д., Мерсеро Р. Цифровая обработка многомерных сигналов. – М.: Мир, 1988. – 488 с.
11. Дмитриев В.И. Прикладная теория информации: Учебник для студентов вузов. - М.: Высшая школа, 1989.- 325 с.
12. Канасевич Э.Р. Анализ временных последовательностей в геофизике. - М.: Недра, 1985.- 300 с.
13. Клаербоут Д.Ф. Теоретические основы обработки геофизической информации с приложением к разведке нефти. – М.: Недра, 1981. – 304 с.
14. Кулханек О. Введение в цифровую фильтрацию в геофизике. – М.: Недра, 1981. – 198 с.
15. Купер Дж., Макгиллем А. Вероятностные методы анализа сигналов и систем. – М.: Мир, 1989. – 376 с.
16. Макс Ж. Методы и техника обработки сигналов при физических измерениях: В 2-х томах. - М.: Мир, 1983.

17. Марпл-мл. С.Л. Цифровой спектральный анализ и его приложения. – М.: Мир, 1990. – 584 с.
18. Никитин А.А. Теоретические основы обработки геофизической информации: Учебник для вузов. - М.: Недра, 1986.- 342 с.
19. Оппенгейм А.В., Шафер Р.В. Цифровая обработка сигналов. – М.: Связь, 1979. – 416 с.
20. Отнес Р., Эноксон Л. Прикладной анализ временных рядов. – М.: Мир, 1982. – 428 с.
21. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов. – М.: Мир, 1978. – 848 с.
22. Рапопорт М.Б. Вычислительная техника в полевой геофизике: Учебник для вузов. - М.: Недра, 1993.- 350 с.
23. Сиберт У.М. Цепи, сигналы, системы. – М.: Мир, 1988. – 336 с.
24. Хемминг Р.В. Цифровые фильтры. – М.: Недра, 1987. – 221 с.
25. Лосев А.К. Линейные радиотехнические цепи: Учебник для вузов. - М.: Высшая школа, 1971. - 560 с.
26. Васильев Д.В. Радиотехнические цепи и сигналы: Учебное пособие для вузов. - М.: Радио и связь, 1982. - 528 с.
27. Зиновьев А.Л., Филиппов Л.И. Введение в теорию сигналов и цепей: Учебное пособие для вузов. - М.: Высшая школа, 1975. - 264 с.
28. Севостьянов Б.А. Курс теории вероятностей и математической статистики. - М.: Наука, 1982. - 256 с.
29. Коваленко И.Н., Филиппова А.А. Теория вероятностей и математическая статистика: Учебное пособие. - М.: Высшая школа, 1982. - 256 с.
30. Худсон Д., Статистика для физиков. - М.: Мир, 1970. - 296 с.
31. Гурский Е.И. Теория вероятностей с элементами математической статистики: Учебное пособие для вузов. - М.: Высшая школа, 1971.- 328 с.
32. Вероятностные методы в вычислительной технике: Учеб. пособие для вузов / А.В.Крайников и др. - М.: Высшая школа, 1986. - 312 с.
33. Игнатов В.А. Теория информации и передачи сигналов. - М.: Советское радио, 1979.
34. Дьяконов В., Абраменкова И. MATLAB. Обработка сигналов и изображений. Специальный справочник. – СПб.: Питер, 2002, 608 с.
35. Астафьева Н.М. Вейвлет-анализ: Основы теории и примеры применения. / Успехи физических наук, 1996, т.166, № 11, стр. 1145-1170.
36. Дремин И.Л. и др. Вейвлеты и их использование. / Успехи физических наук, 2001, т.171, № 5, стр. 465-501.
37. Петухов А.П. Введение в теорию базисов всплесков. – СПб.: Изд. СПбГТУ, 1999, 132 с.
38. Адаптивные фильтры. /Под ред. К.Ф.Н.Коузэна и П.М.Гранта. – М.: Мир, 1988, 392 с.
39. Дьяконов В.П. Вейвлеты. От теории к практике. – М.: СОЛОН-Р, 2002. – 448 с.
40. Корн Г., Корн Е. Справочник по математике для научных работников и инженеров. – М.: Наука, 1984.
41. Овечкина Е.В. (НТИ УГТУ-УПИ), Поршневу С.В. (УГТУ-УПИ). Разработка методов оптимальной аппроксимации эмпирических зависимостей. (Статья в электронном журнале).
42. Новиков Л.В. Основы вейвлет-анализа сигналов. Учебное пособие. СПб, ИАНП РАН, 1999
43. Айфичер Э., Джервис Б. Цифровая обработка сигналов. Практический подход. / М., "Вильямс", 2004, 992 с.
44. Большаков И.А., Ракошиц В.С. Прикладная теория случайных потоков, М.: Сов.



радио, 1978,- 248с.

45. Солонина А.И. и др. Основы цифровой обработки сигналов. Учебное пособие. – СПб.: БХВ Петербург, 2005. – 768 с. URL: <http://lord-n.narod.ru/download/books/walla/dsp/Solonin.Osnovu.DSP.rar>

46. Хуанг Т.С. и др. Быстрые алгоритмы в цифровой обработке изображений. – М.: Радио и связь, 1984. – 224 с.

47. Сойфер В.А. Компьютерная обработка изображений. Часть 2. Методы и алгоритмы. – Соросовский образовательный журнал №3, 1996.

48. Апальков И.В., Хрящев В.В. Удаление шума из изображений на основе нелинейных алгоритмов с использованием ранговой статистики. - Ярославский государственный университет, 2007.

49. Андреев А.Л. Автоматизированные телевизионные системы наблюдения. Часть II. Арифметико -логические основы и алгоритмы. Учебное пособие. - СПб: СПб, ГУИТМО, 2005. – 88с.

50. Розенфельд Л.Г. и др. Возможности постобработки диагностических КТ и МРТ-изображений на персональном компьютере. - Украинський медичний часопис – № 6 (56) – XI/XII, 2006.

51. Лукин А. Введение в цифровую обработку сигналов (Математические основы).- М.: МГУ, Лаборатория компьютерной графики и мультимедиа, 2002. – <http://pv.bstu.ru/dsp/dspcourse.pdf>, <http://dsp-book.narod.ru/dspcourse.djvu>, <http://geogin.narod.ru/arhiv/dsp/dsp4.pdf>.