

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Нижегородский государственный университет им. Н.И. Лобачевского

М.А. Кройц, О.А. Морозов

МЕТОДЫ РЕШЕНИЯ ОБРАТНЫХ ЗАДАЧ В ЦИФРОВОЙ ОБРАБОТКЕ СИГНАЛОВ

Практикум

Рекомендовано методической комиссией физического факультета
для студентов ННГУ, обучающихся по направлению подготовки
09.03.02 «Информационные системы и технологии»

Нижний Новгород
2020

УДК 004.942
ББК 22.3
К-82

Рецензент: к.ф.-м.н., с.н.с. **П.Е. Овчинников**

К-82 Кройц М.А., Морозов О.А. Методы решения обратных задач в цифровой обработке сигналов: Практикум – Нижний Новгород: Нижегородский госуниверситет, 2020. – 26 с.

В пособии рассмотрены подходы к решению ряда учебных задач, представляющих собой обратные задачи цифровой обработки сигналов, представлены постановка задачи, теоретические сведения, замечания по программной реализации и оформлению результатов.

Практикум предназначен для студентов физического факультета ННГУ, его целью является оказание методической помощи при изучении и практическом освоении методов решения обратных задач в области цифровой обработки сигналов и экспериментальных данных при проведении практических занятий по дисциплине «Информационные технологии».

Ответственный за выпуск:
председатель методической комиссии физического факультета ННГУ
к.ф.-м.н., доцент **А.А. Перов**

УДК 004.942
ББК 22.3

© Нижегородский государственный университет
им. Н.И. Лобачевского, 2020

Содержание

Введение	4
1 Деконволюция свёртки методом максимума энтропии	5
1.1 Цель учебной задачи	5
1.2 Теоретические сведения	5
1.3 Постановка задачи	6
1.4 Замечания по реализации	10
1.5 Дополнительные вопросы	10
1.6 Вывод функционала энтропии	11
1.7 Пример пользовательского интерфейса	12
2 Решение фазовой проблемы с помощью алгоритма Фиенупа	13
2.1 Цель учебной задачи	13
2.2 Теоретические сведения	13
2.3 Постановка задачи	14
2.4 Замечания по реализации	16
2.5 Дополнительные вопросы	17
2.6 Пример пользовательского интерфейса	18
3 Построение и визуализация базиса Каруена-Лоэва	19
3.1 Цель учебной задачи	19
3.2 Теоретические сведения	19
3.3 Постановка задачи	21
3.4 Замечания по реализации	23
3.5 Дополнительные вопросы	24
Список литературы	25

Введение

Методы цифровые обработки сигналов занимают важное место при решении большого класса задач в различных областях науки и техники и в настоящее время являются активно развивающимся направлением информационных технологий. В различных областях экспериментальной физики и в технике при изучении физических процессов часто возникает необходимость получить оценку исходного сигнала или оценить значения параметров модели физического процесса на основе зарегистрированных (наблюдаемых) данных, поступающих с различных датчиков. Большинство таких задач относится к классу обратных задач. В качестве примеров можно привести задачи восстановления сигналов и улучшения качества изображений в (радио)астрономии, компьютерной томографии, неразрушающем контроле материалов. Обратные задачи часто оказываются некорректными и требуют специальных методов решения для получения устойчивого решения (методов регуляризации и учета априорной информации). Понятие корректности задачи ввел французский математик и механик Ж. Адамар (1865-1963). Задача является корректно поставленной по Адамару, если выполняются следующие условия (условия корректности):

1. задача имеет решение при любых допустимых исходных данных (решение задачи существует);
2. исходным данным соответствует единственное решение (однозначность решения задачи);
3. решение задачи устойчиво к малым возмущения исходных данных (устойчивость решения).

Третье условие играет важную роль при решении многих некорректных обратных задач, где получить точное решение может оказаться практически невозможным, поскольку исходные данные физической или технической задачи всегда задаются с некоторой погрешностью. В этих условиях необходимо получить такое решение, в котором малые возмущения исходных данных не приводили бы к большим отклонениям в решении.

Практикум предназначен для студентов физического факультета ННГУ по направлению подготовки «Информационные системы и технологии». Целью данного пособия является оказание методической помощи при изучении и практическом освоении методов решения обратных задач в области цифровой обработки сигналов и экспериментальных данных при проведении практических занятий по дисциплине «Информационные технологии». В пособии рассмотрены подходы к решению ряда учебных задач, представляющих собой обратные задачи цифровой обработки сигналов, представлены постановка задачи, теоретические сведения, замечания по программной реализации и оформлению результатов компьютерного моделирования.

1 Деконволюция свёртки методом максимума энтропии

1.1 Цель учебной задачи

Задача деконволюции свёртки является одной из самых распространённых задач на практике цифровой обработки сигналов. Особенно когда имеем дело с линейными системами, в которых выходной сигнал вычисляется как линейная свёртка входного сигнала с импульсной характеристикой линейной системы. Такую задачу называют прямой или задачей распространения, когда с помощью линейной операции свёртки находят выходной сигнал, зная входной сигнал и импульсную характеристику линейной системы. Если же даны импульсная характеристика и выходной сигнал, и требуется восстановить входной сигнал, то такая задача называется обратной задачей. Такая задача относится к некорректным задачам и для её решения требуется применить какой-нибудь способ регуляризации или доопределить задачу. Целью данной вычислительной лабораторной работы является обращение циклической свёртки с помощью метода максимума энтропии в качестве регуляризирующего требования к решению.

1.2 Теоретические сведения

Критерий корректности по Адамару

Как известно из теории некорректных задач, критерий корректности по Адамару заключается в выполнении трёх условий:

1. Решение существует;
2. Решение единственное;
3. Решение устойчиво при любых допустимых значениях f из пространства Φ исходных данных, т.е. непрерывно зависит от исходных данных f .

Теорема свёртки

Свёртка обладает таким свойством, что в частотной области она переходит по теореме свёртки в простое умножение сигналов:

$$y = x \otimes h \leftrightarrow Y = X \cdot H. \quad (1)$$

Казалось бы, что очень простым способом обращения свёртки является решение уравнения свёртки в частотной области относительно искомого входного сигнала. Этот подход носит название «метод Фурье»:

$$Y = X \cdot H \Rightarrow X = \frac{Y}{H} \Rightarrow x(t) = \mathcal{F}^{-1} \left\{ \frac{Y}{H} \right\}. \quad (2)$$

Недостатком метода Фурье является неустойчивость. Если частотная характеристика линейной системы в каком-то месте очень мала (или вообще нулевая), то могут получиться очень большие значения, из-за ограниченной точности представления чисел с плавающей запятой в памяти ЭВМ ситуация ещё усугубляется. Задача будет неустойчива, если дробь вычисляется в числах с плавающей запятой и при этом H или Y стремятся к нулю.

Метод максимума энтропии (ММЭ)

Метод максимума энтропии (ММЭ) основан на принципе максимума энтропии, который играет важнейшую роль в данной задаче и в принципе в теории информации и обработки экспериментальных данных и гласит:

Если мы имеем дело с неполными данными, то, чтобы сделать выводы на их основе, мы должны опираться на такое распределение вероятностей, которое наилучшим образом согласовывалась бы с имеющимися данными и при этом имела максимальную информационную энтропию Шеннона.

1.3 Постановка задачи

В данной задаче, по сути, решаем не только обратную задачу, но и задачу прямого распространения, чтобы создать данные, на основе которых можно было бы решить обратную задачу. Входной сигнал представляет собой сумму гауссовых куполов, то есть определяется по следующей формуле:

$$s(t) = \sum_{n=1}^N A_n \cdot \exp \left(- \left(\frac{t - t_{0n}}{\sigma_n} \right)^2 \right). \quad (3)$$

Для решения данной задачи можно задать значение $N = 3$. Чтобы понять, как работает алгоритм и как сказывается вид входного сигнала на качестве восстановленного сигнала, предлагается построить сигнал примерно следующего вида:

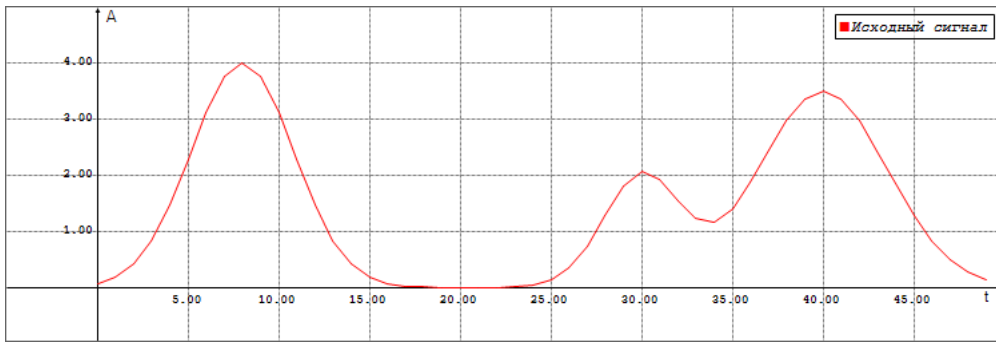


Рис. 1. Вид входного сигнала $x(t)$

Важно отметить, что данная задача представляет собой задачу многомерной оптимизации. Решение будет уточняться на каждой итерации, причём таких итераций в зависимости от требуемой точности может быть достаточно много. Поэтому стоит отойти от привычной длины сигналов порядка 1024 отсчётов и в этой задаче построить сигналы длиной не более 50 отсчётов.

На следующем шаге необходимо определить вид импульсной характеристики. Она будет представлять собой гауссов купол, симметрично расположенный на концах импульсной последовательности. Формула вычисления значений отсчётов имеет вид:

$$h(t) = \exp\left(-\frac{t^2}{\sigma^2}\right). \quad (4)$$

Амплитуду можно выбрать единичную, поскольку из-за нормировки величина амплитуды на входе сигналов и качестве восстановленного сигнала не скажется, а является лишь масштабирующим параметром. В итоге импульсная характеристика должна иметь примерно следующий вид:

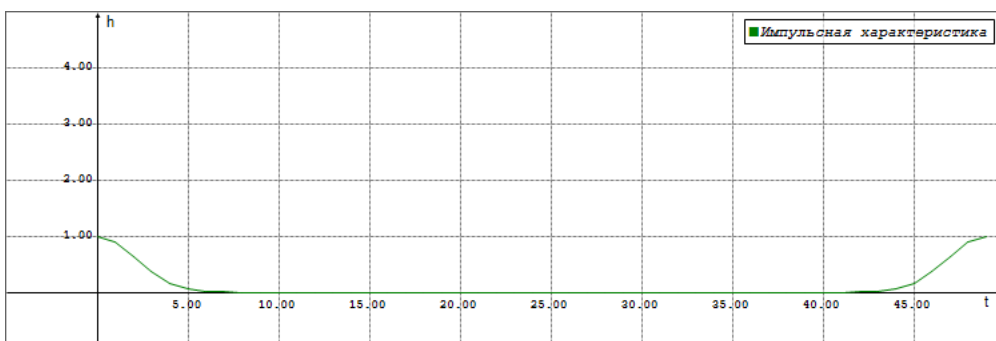


Рис. 2. Вид импульсной характеристики $h(t)$

В качестве заключительного шага подготовки к основной задаче необходимо с помощью циклической свёртки вычислить выходной сигнал линейной системы с полученной импульсной характеристикой. Для этого воспользуемся формулой:

$$y(t) = x(t) \otimes h(t) \Rightarrow y_i = \sum_k x_k \cdot h_{i-k}. \quad (5)$$

Таким образом, получим выходной сигнал, имеющий следующий вид:

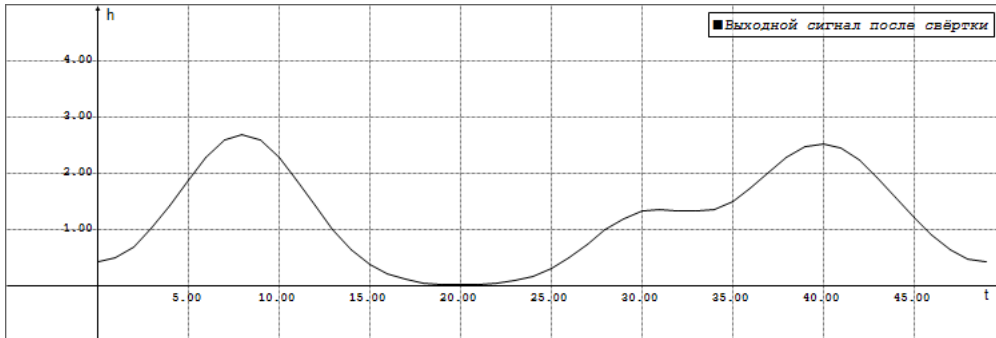


Рис. 3. Вид выходного характеристики $y(t)$

Дальше можно приступить к основной части задачи. Доопределим задачу и используем теор-етико-информационный подход. В качестве функционала-критерия (1) будем использовать функционал информационной энтропии Шеннона, в качестве дополнительного условия (2) будем использовать уравнение формирования, являющееся линейным ограничением. Функционал энтропии является выпуклым, поэтому после наложения линейных ограничений (а свёртка, участвующая в уравнении формирования, есть линейная операция) обязательно получим единственный минимум оптимизационной задачи, то есть единственное решение, причём дополнительных стабилизирующих функционалов, как в методе регуляризации Тихонова, нет необходимости, поскольку требование максимума энтропии само по себе уже является очень мощным условием. Таким образом, получается функционал следующего вида:

$$\Phi = - \underbrace{\sum_{i=1}^M x_i \ln x_i}_{(1)} + \underbrace{\sum_{i=1}^M \lambda_i \left(y_i - \sum_{k=1}^M x_k h_{i-k} \right)}_{(2)} \rightarrow \text{opt}. \quad (6)$$

где λ_i – неопределённые множители Лагранжа. На основе такого функционала выражаем входной сигнал и составим новый функционал среднеквадратичного отклонения, то есть сводим вариационную задачу к оптимизационной. Это широко применяемый подход. Для этого необходимо выразить входной сигнал в явном виде. Как известно из курса

вариационного исчисления, условием оптимальности функционала является равенство нулю его первой вариации, то есть всех частных производных первого порядка по всем переменным. После такого шага получим результат, в следующем виде:

$$x_i = \exp(\dots) = f(\vec{\lambda}). \quad (7)$$

Подставив это выражение в линейное ограничение, то есть в уравнение формирования, получим выходной сигнал как функцию от параметра регуляризации $\vec{\lambda}$, который можно сравнить с имеющимися отсчётами выходного сигнала, например в виде функционала среднеквадратичного отклонения

$$y_i = \sum_{k=1}^M f_k(\vec{\lambda}) h_{i-k}. \quad (8)$$

В итоге получим функционал среднеквадратичного отклонения вида

$$\Phi = \sum_{i=1}^M \left(\bar{y}_i - f_i(\vec{\lambda}) \right)^2. \quad (9)$$

Вектор неопределённых множителей Лагранжа $\vec{\lambda}$ находим методом многомерной оптимизации данного функционала, для чего предлагается использовать метод Хука-Дживса. На основе полученных множителей Лагранжа вычисляется восстановленный сигнал, который необходимо сравнить с исходным входным сигналом.

Следует отметить, что задача многомерной оптимизации при большой размерности является вычислительно сложной. Чтобы ускорить вычисления, предлагается при реализации оптимизируемой функции избегать использования тройных циклов, а вместо них ввести временный массив для расчётов, что позволяет реализовать оптимизируемую функцию через два двойных цикла.

Дополнительная задача

Исследуйте, что будет, если перед восстановлением накладывать на выходной сигнал белый гауссов шум с заданной дисперсией. При каком значении ρ_ω алгоритм перестанет надёжно работать?

1.4 Замечания по реализации

Процесс восстановления сигнала необходимо визуализировать как анимацию. Предусмотреть возможность сброса, останова и продолжения/запуска процесса восстановления. В данной задаче необходимо задавать через пользовательский интерфейс:

- Амплитуды, дисперсии и положения центров составляющих гауссовых куполов во входном сигнале;
- Дисперсию и амплитуду импульсной характеристики;
- Точность восстановления;
- Дисперсию белого шума, наложенного на выходной сигнал перед восстановлением.

Требуется вывести на экран:

- Графики исходного и восстановленного сигналов;
- Графики импульсной характеристики и сигнала после циклической свёртки;
- Среднеквадратичное отклонение между исходным и восстановленным сигналами на один отсчёт;
- Значение функционала среднеквадратичной ошибки.

1.5 Дополнительные вопросы

1. Почему данная задача является некорректной?

Ответ. Некорректность заключается в том, что линейная система действует как фильтр, то есть в выходном сигнале могут быть не все частоты, содержащиеся в исходном входном сигнале. Какие именно частоты были подавлены фильтром неизвестно, а значит решений существует бесконечное множество. С помощью ММЭ из этого множества выбирается единственное решение с максимальной информационной энтропией Шеннона.

2. Почему мы выбираем для оптимизации именно метод Хука-Дживса? В чём заключается его преимущество перед другими методами? Можно ли применить другие методы оптимизации?

Ответ. Отличительной особенностью метода Хука-Дживса является тот факт, что он в отличие от других методов осуществляет оптимизацию функционала без необходимости вычисления производных. Это важный аспект, поскольку дифференцирование само по себе является неустойчивой операцией. На самом деле: производная необязательно существует во всех интересующих нас точках.

3. Почему мы имеем право применить метод максимума энтропии? Какие условия должны быть выполнены, чтобы метод максимума энтропии был применимым?

Ответ. Метод максимума энтропии применим тогда, когда об искомом сигнале известно, что он обладает конечной энергией (то есть имеется возможность нормировки) и положительно определён, чтобы его можно было рассматривать как распределение вероятностей. Для сигнала в данной задаче выполняются указанные требования, поэтому имеем право воспользоваться методом МЭ. Также в задаче очевидно присутствуют неполные данные, поскольку некоторые частотные составляющие исходного сигнала могут быть подавленными в результате применения линейной системы, которая действует на сигнал как фильтр.

1.6 Вывод функционала энтропии

$$\Phi = - \sum_{i=1}^M x_i \ln x_i + \sum_{i=1}^M \lambda_i \left(y_i - \sum_{k=1}^M x_k h_{i-k} \right);$$

$$\begin{aligned} \frac{\partial \Phi}{\partial x_i} &= - \ln x_i - 1 + \underbrace{\frac{\partial}{\partial x_i} \left[\sum_{i=1}^M \lambda_i y_i \right]}_0 - \frac{\partial}{\partial x_i} \left[\sum_{i=1}^M \lambda_i \sum_{k=1}^M x_k h_{i-k} \right] = \\ &= - \ln x_i - 1 - \left[\sum_{k=1}^M \lambda_k h_{k-i} \right] = 0; \end{aligned}$$

$$\ln x_i = -1 - \sum_{k=1}^M \lambda_k h_{k-i} \Rightarrow x_i = \exp \left(-1 - \sum_{k=1}^M \lambda_k h_{k-i} \right) = f(\vec{\lambda});$$

$$y_i(\vec{\lambda}) = \sum_{k=1}^M f_k(\vec{\lambda}) h_{i-k} = \sum_{i=1}^M \left[\exp \left(-1 - \sum_{k=1}^M \lambda_k h_{k-i} \right) h_{i-k} \right];$$

$$\Phi_\sigma = \sum_{k=1}^M \left(\bar{y}_i - y_i(\vec{\lambda}) \right)^2.$$

1.7 Пример пользовательского интерфейса

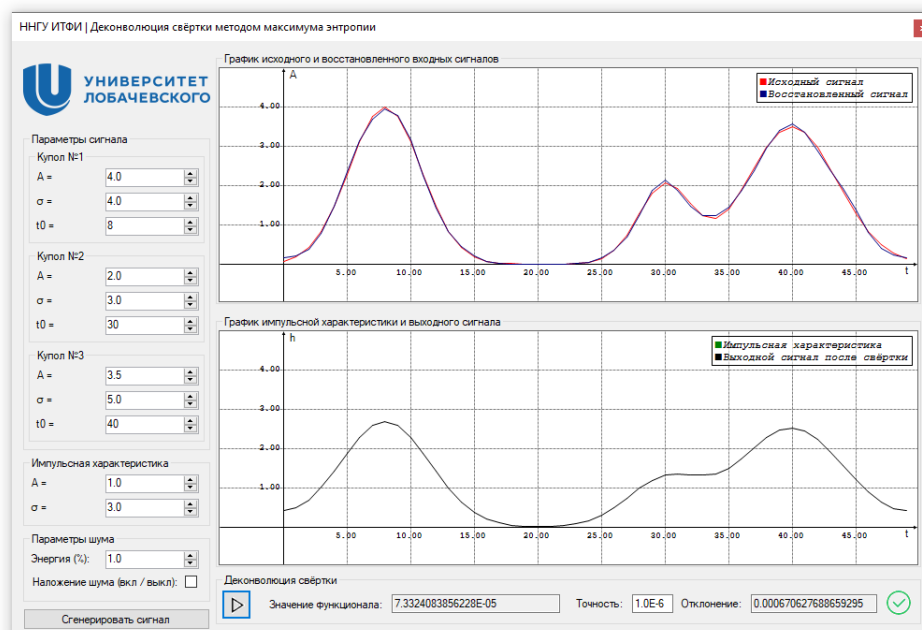


Рис. 4. Пример пользовательского интерфейса

На рисунке 4 приведен пример пользовательского интерфейса моделирующей программы. В верхнем окне программы представлены исходный и восстановленный сигналы после деконволюции, в нижнем окне представлены импульсная характеристика и исходный сигнал после циклической свёртки. Меню слева позволяет задавать параметры генерируемых сигналов. Нижнее меню служит для управления восстановительным процессом и анимацией, а также выводит текущее отклонение между исходным и восстановленным сигналами. По достижении заданной точности программа сообщает о завершении процесса деконволюции.

2 Решение фазовой проблемы с помощью алгоритма Фиенупа

2.1 Цель учебной задачи

Очень часто на практике известным является не сам спектр (который представляет собой комплексную величину), а лишь его модуль. Такая ситуация встречается, например, в рентгеновской кристаллографии и электронной микроскопии. Чтобы восстановить исходный сигнал, необходимо восстановить потерянную информацию о фазе. Чтобы можно было максимально просто учесть априорную информацию о сигнале, рекомендуется использовать итерационный алгоритм. Целью данной вычислительной лабораторной работы является реализация итерационного алгоритма Фиенупа для восстановления фазовой информации на примере сигнала, состоящего из гауссовых кривых.

2.2 Теоретические сведения

Алгоритм Фиенупа основан на использовании так называемых псевдодифференциальных операторов с использованием преобразования Фурье. При применении алгоритма Фиенупа можно наблюдать целый ряд интересных особенностей, которые можно объяснить с помощью свойств преобразования Фурье.

Свойства преобразования Фурье

1. **Линейность:**

$$a \cdot g(t) + b \cdot h(t) \xrightarrow{\mathcal{F}} a \cdot G(f) + b \cdot H(f).$$

2. **Сдвиг по времени:**

$$h(t - t_0) \xrightarrow{\mathcal{F}} H(f) \cdot \exp(-j2\pi f t_0).$$

3. **Сдвиг по частоте:**

$$h(t) \cdot \exp(j2\pi f_0 t) \xrightarrow{\mathcal{F}} H(f - f_0).$$

4. **Масштабирование:**

$$\frac{1}{|\alpha|} \cdot h\left(\frac{t}{\alpha}\right) \xrightarrow{\mathcal{F}} H(\alpha f).$$

5. **Теорема свёртки:**

$$g(t) \otimes h(t) \xrightarrow{\mathcal{F}} G(f) \cdot H(f),$$

$$g(t) \cdot h(t) \xrightarrow{\mathcal{F}} G(f) \otimes H(f).$$

2.3 Постановка задачи

Поскольку для применения алгоритма Фиенупа изначально должен быть известен модуль спектра, в качестве подготовительного шага предстоит получить модуль спектра некоторого известного сигнала, модуль спектра которого будет служить входными данными алгоритма. В этой задаче предлагается сгенерировать сигнал, состоящий из пяти гауссовых куполов. Такой сигнал определяется по формуле

$$s(t) = \sum_{n=1}^5 A_n \cdot \exp\left(-\left(\frac{t - t_{0n}}{\sigma_n}\right)^2\right). \quad (10)$$

Для получения лучшего результата рекомендуется брать достаточно узкие купола, не перекрывающиеся друг с другом. Таким образом, исходный сигнал должен иметь примерно следующий вид:

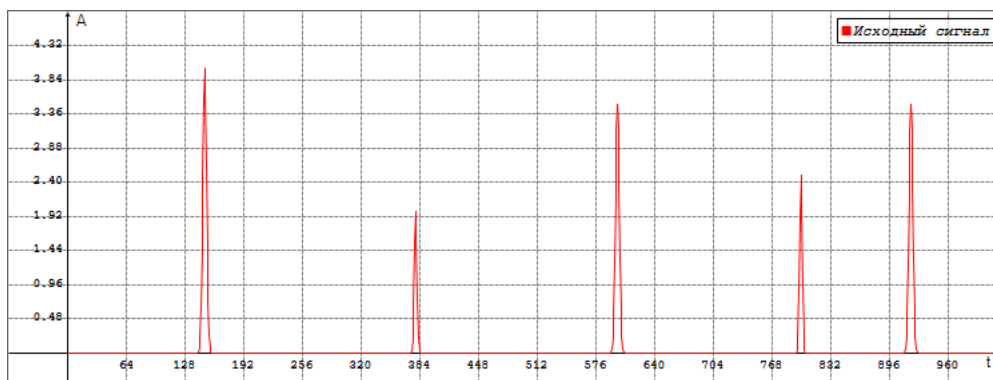


Рис. 5. Вид исходного сигнала $s(t)$

Вычислив модуль преобразования Фурье этого сигнала, получим модуль спектра, фазу которого в дальнейшем требуется восстановить с помощью алгоритма Фиенупа. На рис. 6 представлен характерный вид модуля спектра исходного сигнала:

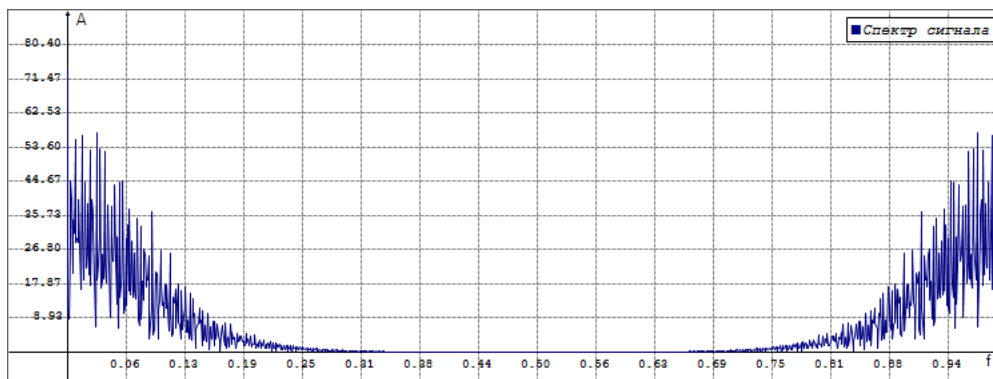


Рис. 6. Вид модуля спектра $|S(f)|$ исходного сигнала $s(t)$

После того, как был получен модуль спектр сигнала, предстоит реализовать непосредственно сам алгоритм Фиенупа. Итерационная схема алгоритма состоит из следующих шагов:

1. Инициализация:

$$|S| \rightarrow S_0(f) = |S| \cdot \exp(j\phi_0), \quad \phi_0 \in [0, 2\pi]. \quad (11)$$

Из известного модуля строится комплексная последовательность чисел, имеющих известный модуль и случайную фазу φ_0 с равномерным распределением в интервале $[0, 2\pi]$. Использование именно равномерно распределённой случайной фазы значительно повышает вероятность получить начальное приближение сигнала с нужными для реализации алгоритма восстановления свойствами.

2. Применение ограничений во временной области:

$$s_0(t) = \mathcal{F}^{-1}\{S_0(f)\}. \quad (12)$$

$$\tilde{s}_0(t) = \begin{cases} s_0(t), & s_0 \geq 0 \\ 0, & s_0 < 0 \end{cases}. \quad (13)$$

Поскольку априорно известно, что исходный сигнал представляет собой сумма гауссовых куполов и является положительно определённым, то можем восстановить сигнал по построенному комплексному спектру и во временной области применить ограничение на неотрицательность сигнала.

3. Замена модуля в частотной области на известный:

$$S_1(f) = \mathcal{F}\{\tilde{s}_0(t)\}, \quad (14)$$

$$\tilde{S}_1(f) = |S| \cdot \exp(j\phi). \quad (15)$$

Применение ограничения на неотрицательность во временной области влечёт за собой определённые изменения в спектральной области, в том числе модуля спектра. Из-за этого встаёт необходимость замены модуля на известный модуль спектра $|S|$. Можно заменить модуль комплексного числа только с помощью простых арифметических операций. При этом нет необходимости пользоваться тригонометрическими функциями для вычисления фазы.

4. Проверка сходимости итерационной схемы: Повторяются шаги (2) и (3) пока изменения сигнала не перестали превышать заданный порог среднеквадратичного отклонения на один отсчёт:

$$\| \tilde{s}^{(k)}(t) - \tilde{s}^{(k-1)}(t) \| < \varepsilon \sim 10^{-6}. \quad (16)$$

2.4 Замечания по реализации

Процесс восстановления сигнала необходимо визуализировать как анимацию. Предусмотреть возможность сброса, останова и продолжения/запуска процесса восстановления. В данной задаче необходимо задавать через пользовательский интерфейс:

- Амплитуды, дисперсии и положения центров составляющих гауссовых куполов во входном сигнале;
- Точность восстановления.

Требуется вывести на экран:

- Графики исходного и восстановленного сигналов;
- График модуля спектра исходного сигнала;
- Среднеквадратичное отклонение между сигналами на данной и предыдущей итерациях.

В результате выполнения алгоритма Фиенупа может получиться сигнал, зеркально отражённый относительно исходного сигнала или сдвинутый по времени. Чтобы иметь возможность совместить исходный и восстановленный сигналы, необходимо реализовать механизм циклического сдвига сигнала в ручном или (лучше) автоматическом режиме, а также предусмотреть возможность зеркального отражения сигнала.

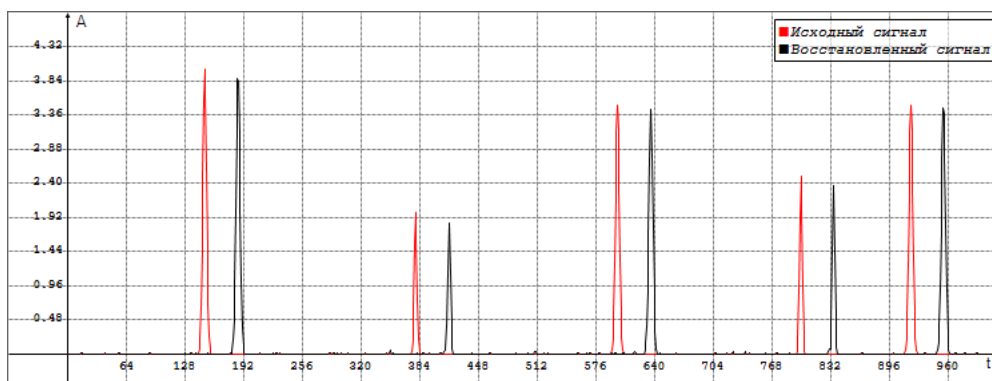


Рис. 7. Пример графика исходного сигнала и восстановленного сигнала, смещённого по времени

2.5 Дополнительные вопросы

1. Объясните, почему восстановленный сигнал может получиться смещённым по времени относительно исходного сигнала.

Ответ. Сдвиг во временной области, присутствующий всегда после завершения работы алгоритма Фиенупа, можно объяснить с помощью свойства (2) преобразования Фурье. Начальную инициализацию (см. шаг №1 итерационной схемы) можно понимать как умножение на комплексную экспоненту в частотной области, что согласно свойству (2) преобразования Фурье вносит постоянный сдвиг в сигнал во временной области.

2. Почему в некоторых случаях восстановленный сигнал является зеркальным отражением исходного сигнала?

Ответ. Зеркальное отражение сигнала есть частный случай свойства (4) преобразования Фурье, где $\alpha = -1$. Отсюда следует, что $h(-t) \xrightarrow{\mathcal{F}} H(-f)$, то есть нет разницы, происходит отражение во временной или в частотной области. Одному модулю спектра соответствуют сразу два комплексных спектра $H(f)$ и $H(-f)$, а значит и два сигнала $h(t)$ и $h(-t)$. Какой из сигналов получаем, определяется случайным образом.

2.6 Пример пользовательского интерфейса

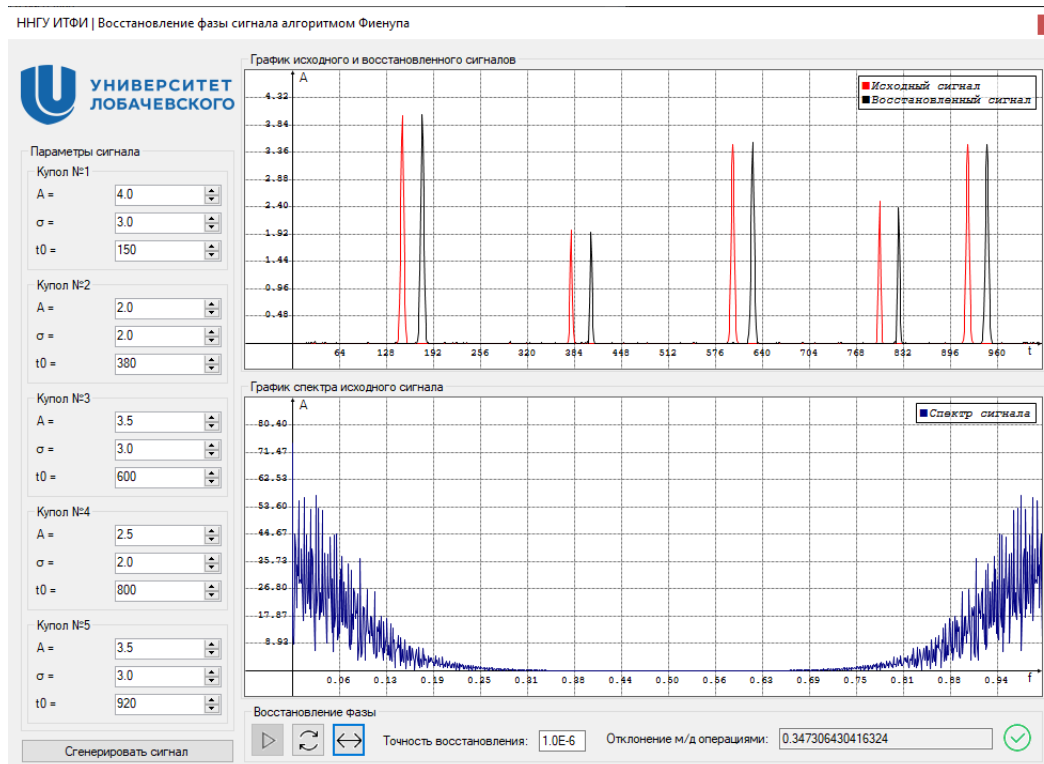


Рис. 8. Пример пользовательского интерфейса

На рисунке 8 приведен пример пользовательского интерфейса моделирующей программы. В верхнем окне программы представлены исходный и восстановленный сигналы, в нижнем окне представлен модуль спектра исходного сигнала, используемый в качестве входных данных для алгоритма восстановления. Меню слева позволяет задавать параметры генерируемого сигнала. Нижнее меню даёт пользователю возможность управлять восстановительным процессом и анимацией, а также позволяет запустить совмещение восстановленного сигнала с исходным. Там же выводится текущее отклонение между исходным и восстановленным сигналами. По достижении заданной точности программа сообщает о завершении процесса деконволюции.

3 Построение и визуализация базиса Каруена-Лоэва

3.1 Цель учебной задачи

Часто при решении задач обработки данных требуется разложить эти данные на систему функций. Набор (система) таких функций называется базисом. Встаёт актуальный вопрос о выборе таких функций для разложения. На примере преобразования Фурье (разложение на гармонические функции) можно понять, что есть удачные и не очень удачные разложения. Например, появление боковых лепестков при Фурье-разложении ограниченной прямоугольным окном выборки является прямым следствием неудачного разложения прямоугольного окна на гармонические функции. Кроме того, разложение в ряд Тейлора также будет неудачным, а ряд из нечётных степеней будет медленно сходящимся. Целью данной вычислительной лабораторной работы является ознакомление с базисом Каруена-Лоэва, оптимальным базисом для разложения сигналов произвольного вида. Попутно требуется ознакомиться с сингулярным разложением матриц, которое является широко применяемым инструментом в цифровой обработке сигналов. С его помощью решаются такие задачи как решение систем линейных уравнений методом наименьших квадратов, построение диаграмм направленности антенн, нахождение собственных чисел и векторов и многие другие.

3.2 Теоретические сведения

Сингулярное разложение матриц

Теорема о сингулярном разложении (Singular Value Decomposition, SVD) гласит, что для любых матриц A произвольного вида можно выполнить разложение (факторизацию) следующего вида:

$$A_{m \times n} = U_{m \times m} \times \Sigma_{m \times n} \times V^T_{n \times n}. \quad (17)$$

где U и V – унитарные матрицы, содержащие левые и правые собственные векторы, соответственно, а Σ – диагональная матрица, содержащая сингулярные числа, которые в случае квадратной матрицы совпадают с квадратным корнем из модуля квадрата собственных чисел этой матрицы, т.е. знак собственного числа не определяется.

Псевдообратная матрица Мура-Пенроуза

Квадратная матрица обратима тогда и только тогда, когда она невырождена, то есть её определитель не равен нулю. Для неквадратных и вырожденных матриц не существует обратных матриц. Однако существует обобщение этого понятия. Одним из таких обобщений является так называемая псевдообратная матрица Мура-Пенроуза, которая может быть получена из сингулярного разложения матрицы и обладает свойствами, похожими на свойства обратной матрицы. Имея сингулярное разложение матрицы, можно вычислить псевдообратную матрицу Мура-Пенроуза A^+ по формуле:

$$A^+ = V\Sigma^+U^T, \quad (18)$$

где U , Σ и V – те же матрицы, образующие сингулярное разложение матрицы A , для которой требуется найти псевдообратную матрицу. Σ^+ есть псевдообратная матрица сингулярных чисел, которая получается обращением ненулевых компонентов, то есть возведением сингулярных чисел в степень -1 :

$$\sigma_{ij}^+ = \begin{cases} 0, & \sigma_{ij} = 0 \\ \frac{1}{\sigma_{ij}}, & \sigma_{ij} \neq 0. \end{cases} \quad (19)$$

Псевдообратная матрица Мура-Пенроуза обладает следующими свойствами:

1. $AA^+A = A$;
2. $A^+AA^+ = A^+$;
3. $(AA^+)^T = AA^+$ и $(A^+A)^T = A^+A$;

Последнее свойство говорит о том, что произведение матрицы на её псевдообратную вне зависимости от порядка перемножения является симметричной матрицей. Более того, в комплексном случае она будет являться эрмитовой матрицей.

Важно отметить, что в общем случае произведение исходной матрицы на псевдообратную, в отличие от произведения матрицы на её обратную матрицу, не равно единичной матрице.

Более подробно можно ознакомиться с алгоритмом сингулярного разложения и его свойствами в литературе, например *Дж. Форсайт, М. Малькольм, К. Моулер «Машинные методы математических вычислений»*.

Решение наименьших квадратов

Если имеется система линейных уравнений вида $A\vec{x} = \vec{b}$, где A – матрица коэффициентов системы, \vec{x} – вектор неизвестных, а \vec{b} – вектор свободных членов или правой части, то решение, найденное по формуле

$$\vec{x} = A^+\vec{b}, \quad (20)$$

будет решением наименьших квадратов для данной системы линейных уравнений.

Базис Каруэна-Лоева

Теорема Каруэна-Лоева существует в разных формулировках и даёт определение базиса Каруэна-Лоева: Система координат $\{\Theta_i\}_{i=1,2,\dots}$ является оптимальной в том смысле, что информационная энтропия Шеннона будет минимальна по всем возможным системам тогда и только тогда, когда эта система является системой координат (базисом) Каруэна-Лоева:

$$H(\{\Theta_i\}) = \min_{\{\Psi_i\}} H(\{\Psi_i\}). \quad (21)$$

Поскольку энтропия Шеннона является мерой неопределённости, в этом случае требуется именно *минимум* энтропии, то есть минимальная неопределённость. В этом и заключается оптимальность базиса Каруэна-Лоева – минимальная неопределённость приводит к тому, что требуется минимальное количество базисных функции для максимально точного представления сигнала в этом базисе.

3.3 Постановка задачи

Данная задача состоит из двух частей, которые служат цели ознакомления с практическими применениями алгоритма сингулярного разложения (SVD). Он же будет использоваться в дальнейшем для нахождения собственных чисел и векторов автокорреляционной матрицы сигналов, для которых требуется найти базис Каруэна-Лоева.

Применение алгоритма сингулярного разложения

Целью данной подзадачи является ознакомление с алгоритмом сингулярного разложения на примере решения системы линейных уравнений с помощью псевдообратной матрицы Мура-Пенроуза.

Для этого предлагается сначала сгенерировать совместную систему линейных уравнений с случайными коэффициентами, состоящая из M уравнений с N неизвестными. Коэффициенты должны быть в интервале

$[-1, 1]$. Для простоты будем считать систему совместной, если её определитель не близок к нулю. Чтобы удобнее было вычислить определитель, можно ограничиться системами размера 3×3 . Полученную систему линейных уравнений необходимо решить с помощью обратной матрицы.

Дальше, вычислив сингулярное разложение матрицы коэффициентов, легко можно найти псевдообратную матрицу, которая служит инструментом нахождения решения системы методом наименьших квадратов. Необходимо убедиться в том, что найденный вектор \vec{x} решений действительно удовлетворяет всем уравнениям, например через вычисление невязок системы:

$$E = \frac{1}{M} \sum_{i=1}^M \left(b_i - \sum_{j=1}^N a_{ij} x_j \right)^2, \quad (22)$$

где a_{ij} – коэффициенты уравнения из матрицы коэффициентов. В качестве частного случая предлагается рассматривать матрицу следующего вида:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}. \quad (23)$$

Определитель этой матрицы равен нулю. Для того, чтобы получить из этой матрицы плохо обусловленную систему, нужно добавить к её элементам небольшие случайные числа $\delta \sim 0.01$. Можно убедиться в неустойчивости вычисления обратной матрицы, сравнивая решения, полученные через обратную матрицу и через псевдообратную матрицу Мура-Пенроуза.

Построение базиса Каруена-Лозва

В рамках данной подзадачи требуется сгенерировать сигналы трёх видов длиной $N = 500 \dots 1000$ отсчётов. Число сигнальных составляющих $M = 1 \dots 3$.

1. Полигармонический сигнал:

$$x_1(t) = \sum_{k=1}^M A_k \cdot \sin(2\pi f_k t + \varphi_k). \quad (24)$$

2. Гауссов сигнал:

$$x_2(t) = \sum_{k=1}^M A_k \cdot \exp \left(- \left(\frac{t - t_{0k}}{\sigma_k} \right)^2 \right). \quad (25)$$

3. Экспоненциально-затухающий полигармонический сигнал:

$$x_3(t) = \sum_{k=1}^M \exp(-\alpha_k t) \cdot \sin(2\pi f_k t + \varphi_k). \quad (26)$$

Для этих сигналов дальше строится автокорреляционная матрица (АКМ) с помощью формулы оценки автокорреляционной последовательности (АКП) для стационарных в широком смысле сигналов. Как известно, для таких сигналов автокорреляция зависит только от разности временных индексов, а не от самих индексов:

$$r_{xx}[m] = \frac{1}{N - |m|} \sum_{k=1}^{N-|m|} x[k] \cdot x[k + m], \quad (27)$$

где m – разность временных индексов или смещение АКП. Таким образом, компоненты АКМ вычисляются по формуле:

$$R_{ij} = r_{xx}[i - j]. \quad (28)$$

Автокорреляционная матрица должна быть порядка $P = 50 \dots 100$. Вычислив сингулярное разложение АКМ, мы получаем её собственные значения и унитарные матрицы U и V , содержащие в качестве столбцов левые и правые собственные векторы, соответствующие этим собственным значениям. Осталось графически изобразить собственные значения (располагать их по порядку убывания). Понятно, что если упорядочить собственные числа, то также нужно аналогично упорядочить собственные векторы, которые играют роль базисных функций. Поскольку мы рассматриваем дискретный случай, их компоненты играют роль отсчётов базисных функций Каруена-Лоэва. Их также необходимо изобразить графически.

3.4 Замечания по реализации

В данной задаче необходимо задавать через пользовательский интерфейс:

- Параметры генерируемой системой линейных уравнений M и N ;
- Тип, параметры и количество сигнальных составляющих для построения базиса Каруена-Лоэва;
- Порядок автокорреляционной матрицы;
- Номер собственного вектора (он же содержит отсчёты собственной функции).

Требуется вывести на экран:

- Данные о системе линейных уравнений и её решении, особенно невязки;
- Графики исходного сигнала;
- График собственных чисел по порядку их убывания;
- График выбранной собственной функции.

3.5 Дополнительные вопросы

Докажите самостоятельно, что для матрицы A полного ранга псевдообратная матрица A^+ совпадает с обратной матрицы A^{-1} .

Ответ. Требуется доказать, что $AA^+ = A^+A = I$, где A – квадратная матрица полного ранга, A^+ – её псевдообратная матрица Мура-Пенроуза, а I – единичная матрица. Для этого подставляем вместо матриц A и A^+ их определения через сингулярное разложение. Рассмотрим первый случай:

$$AA^+ = I,$$

$$U\Sigma V^T V\Sigma^+ U^T = I.$$

В силу свойства унитарности матрицы V получается, что $V^T V = V V^T = I$, то есть выражение упрощается и мы имеем:

$$U\Sigma\Sigma^+ U^T = I.$$

Зная, что псевдообратная матрицы сингулярных чисел состоит из одних нулей за исключением главной диагонали, на которой стоят обращения собственных чисел матрицы A , можно понять, что произведение $\Sigma\Sigma^+ = \Sigma^+\Sigma = I$. Это верно только тогда, когда матрица A полного ранга, поскольку лишь в этом случае все собственные (и тем самым и сингулярные) числа отличны от нуля. Остаётся:

$$UU^T = I.$$

Это равенство заведомо верно, опять же в силу свойства унитарности матрицы U . Таким образом, мы доказали, что в случае матрицы полного ранга её псевдообратная матрица совпадает с обратной. Аналогично доказывается второй случай $A^+A = I$.

Список литературы

- [1] Тихонов А.Н., Арсенин В.Я., Методы решения некорректных задач. – М.: Наука, 1974. – 286 с.
- [2] Тихонов А.Н., Гончарский А.В., Степанов В.В., Ягола А.Г., Численные методы решения некорректных задач. – М.: Наука, 1990. – 232 с.
- [3] Форсайт Дж., Малькольм М., Моулер К., Машинные методы математических вычислений. – М.: Мир, 1980. – 280 с.
- [4] Макс Ж., Методы и техника обработки сигналов при физических измерениях. Том 1. – М.: Мир, 1983.
- [5] Василенко Г.И., Теория восстановления сигналов. – М.: Сов. Радио, 1979. – 272 с.
- [6] Гончарский А.В., Черепашук А.М., Ягола А.Г., Некорректные задачи астрофизики. – М.: Наука, 1985. – 352 с.
- [7] Компьютеры в оптических исследованиях. / Под ред. Б.Фридена. – М.: Мир, 1983. – 488 с.
- [8] Бэйтс Р., Мак-Доннелл М., Восстановление и реконструкция изображений. – М.: Мир, 1989. – 336 с.
- [9] Сумин М.И., Некорректные задачи и методы их решения. Материалы к лекциям для студентов старших курсов: Учебное пособие. – Нижний Новгород: Изд-во Нижегородского госуниверситета, 2009. – 289 с.

Михаэль Александр **Кройц**
Олег Александрович **Морозов**

Методы решения обратных задач в цифровой обработке сигналов

Практикум

Федеральное государственное автономное образовательное учреждение
высшего образования
«Нижегородский государственный университет им. Н.И. Лобачевского»
603950, Нижний Новгород, пр. Гагарина 23