

Введение

Рассмотрим временной ряд, представленный на Рис. 1.

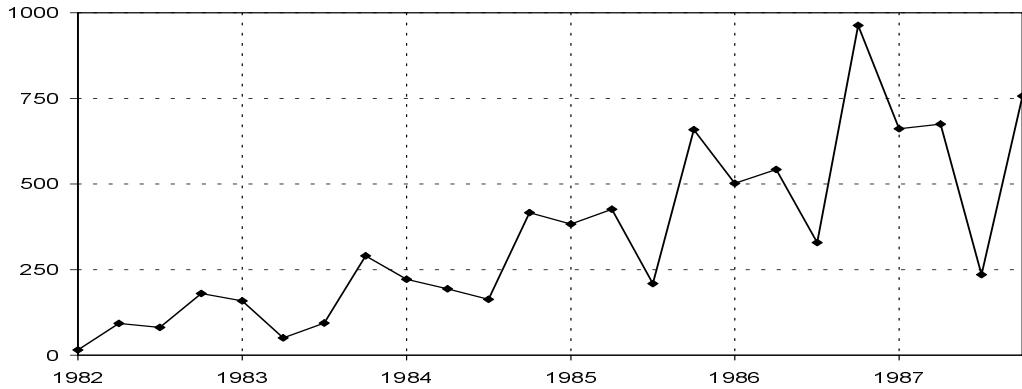


Рис. 1. Темпы роста заболеваемости СПИДом в США. Исходный ряд.

Этот ряд описывает ежеквартальные увеличения числа людей в США, заболевших СПИДом в течение 1982-1988 гг.

Очевидно, что рассматриваемый ряд нестационарен как по среднему, так и по дисперсии. Кроме того, он короткий (всего имеется 24 точки). Это исключает надежду на то, что стандартные методы типа Фурье-анализа в применении к исходному ряду могут привести к разумным выводам. Применение же метода "Гусеница" (без центрирования и нормирования, см. раздел I.1.2) позволяет провести разложение ряда на следующие информативные компоненты, см. Рис. 2: тренд – 1 ГК; годовая периодика – 3-4 ГК; полугодовая периодика – 2 ГК. Сумма последних двух периодик, также изображенная на Рис. 2, может быть проинтерпретирована как сезонная компонента.

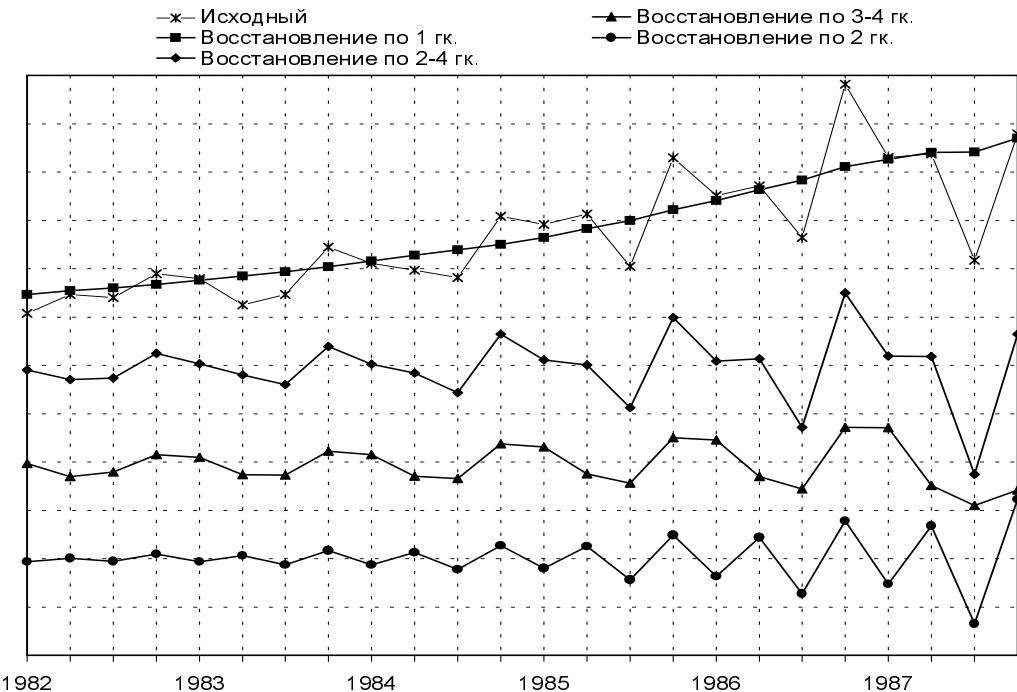


Рис. 2. Темпы роста заболеваемости СПИДом в США. Основные компоненты.

Приведенный пример является одним из простейших, где применение метода "Гусеница" помогает увидеть и выделить разного рода закономерности в поведении временных рядов, зачастую априори не очевидные. В книге приводится большое количество других, в большинстве существенно более сложных, примеров. Наряду с примерами использования "Гусеницы" и подробным описанием методологии, авторы уделяют большое внимание также теоретическому исследованию метода.

Отметим, что под временными рядом мы понимаем набор $\{f(i\Delta)\}_{i=1}^N$ значений, вообще говоря, произвольной функции одной переменной $f(t)$, $t > 0$, в равнотстоящих точках $i\Delta$ ($i = 1, \dots, N$). При этом переменная t не обязательно имеет смысл времени.

Во многих естественных науках сложилось представление о возможности описания природных процессов с помощью функций, состоящих из нескольких слагаемых

$$f(t) = f_T(t) + f_n(t) + f_r(t) + \varepsilon(t), \quad t \in [0, T],$$

где $f_T(t)$ – медленная нерегулярная составляющая, обычно называемая трендом, часто ее пытаются описывать алгебраическими полиномами невысоких порядков;

$f_n(t)$ – периодическая или сумма периодических составляющих; в зависимости от области приложений они называются сезонными, суточными и т.п. вариациями;

$f_r(t)$ – быстрые нерегулярные малые вариации, в которые обычно включают все, что не укладывается в формальную модель, иногда сюда включают и случайные шумы;

$\varepsilon(t)$ – чисто случайная составляющая, описываемая случайным процессом определенного типа.

Во многих частных случаях к настоящему времени созданы мощные теории с развитым аппаратом приложений и компьютерными реализациями в виде библиотек и пакетов программ. Так, для функций вида $f(t) = f_T(t) + \varepsilon(t)$ такой теорией является теория аппроксимации (при малых $\varepsilon(t)$) или метод наименьших квадратов математической статистики (при больших $\varepsilon(t)$), для функций вида $f(t) = f_n(t)$ хорошо работает теория гармонических рядов Фурье.

Однако, во многих ситуациях возникают достаточно большие сложности эффективного исследования функций. Примером может быть случай $f(t) = f_T(t) + f_n(t)$, где при отсутствии априорной информации о частотах компонент периодической составляющей не работают ни теория аппроксимации, ни теория рядов Фурье. В различных конкретных приложениях предложено много эвристических приемов, но они, как правило, плохо теоретически обоснованы. В настоящей книге делается попытка показать, что при использовании метода "Гусеница" часто оказывается возможным (приближенно) выделить отдельные составляющие исходного ряда.

В книге подробно описаны:

- методология (главы I, II, III),
- теоретическое обоснование (главы III, V),
- применения (главы I, IV, VII),
- программное обеспечение (глава VI).

Глава I является вводной. Она посвящена описанию базового алгоритма анализа одномерных временных рядов, реализующего метод "Гусеница", а также его возможностей с демонстрацией примеров применения. В этой главе формальное описание алгоритма сочетается с неформальным, включающим в себя очень важное для методов такого рода рассмотрение визуальных представлений результатов метода. Обсуждается выбор параметров, при этом учитывается практический опыт применения метода авторами. Примеры подобраны так, чтобы, с одной стороны, наглядно проиллюстрировать возможности "Гусеницы" и продемонстрировать как можно больше способов "разглядывания" результатов, которые могут помочь при использовании метода, а с другой стороны, обсудить возникающие при применении метода проблемы.

В Главе II описаны основные идеи, приводящие к методам типа "Гусеница", а также сами эти методы. Материал главы в значительной степени является обзорным, хотя, в силу обширности затрагиваемых тем, не претендует на полноту и строгость изложения. Разделы данной главы описывают стандартную процедуру анализа главных компонент (АГК), использование АГК для анализа стационарных временных рядов и в теории динамических систем, динамический анализ Фурье, вейвлет-анализ, а также два метода, по сути совпадающие с некоторыми вариантами "Гусеницы" – анализ сингулярного спектра (Broomhead и др.) и метод развертки (Бухштабер), см. точные ссылки в конце главы II.

В Главе III базовый алгоритм, подробно описанный в Главе I, распространен для решения задач обработки многомерных рядов и полей, а также прогнозирования одномерных и многомерных рядов. Развитие идеи метода "Гусеница", проведенное в данной главе, показывает богатое и до конца не использованное содержание этого метода. Изложение иллюстрируется большим количеством примеров.

В Главе IV описаны результаты подробного сравнительного анализа метода "Гусеница" и классических методов анализа одномерных временных рядов на трех стандартных наборах данных: объем пассажироперевозок на международных авиалиниях, заболеваемость СПИДом в США и среднемесячное число яиц на несушку в США. Результаты сравнительного анализа показывают, что для "хороших" данных метод "Гусеница" по крайней мере так же хорош, как и испытанные классические средства. При этом часто он позволяет достичь того же результата более естественным путем или обнаружить эффект, который не проявляют стандартные методы.

Глава V посвящена теории метода "Гусеница". Ключевым разделом данной главы является первый, в котором метод исследован на языке функционального анализа, точнее – теории операторов Гильберта-Шмидта. Показано, что на этом языке оказывается возможным не только описать единым образом сам метод, его различные модификации и обобщения, но и объяснить некоторые наблюдаемые на практике явления, возникающие при обработке реальных временных рядов. В частности, исследованы асимптотические свойства метода, существенно отличающиеся от его "допредельных" свойств.

В Главе VI описана реализующая базовый вариант метода "Гусеница" программа, которая находится на прилагаемой к книге дискете.

Глава VII содержит примеры, в которых метод "Гусеница" позволил получить содержательные с практической точки зрения результаты. Разделы этой главы включают в себя интерпретацию результатов, данную специалистами в

соответствующих областях науки.

В заключении сделаем несколько замечаний, относящихся к отечественной теории метода "Гусеница" (обзор зарубежных аналогов метода см. в Главе II).

Разработка метода явилась результатом стажировки по математической биологии в МГУ в 1961 г. под руководством А.Н. Колмогорова выпускника математико-механического факультета ЛГУ О.М. Калинина. В это время А.Н. Колмогоров проявил активный интерес к задаче выделения периодичностей в природных процессах. Отметим, в частности, работу 1962 г. Арато, Колмогорова и Синая, посвященную Эйлеровскому (10 месяцев) и Чандлеровскому (14 месяцев) периодам в движении полюсов Земли. Метод был сформулирован О.М. Калининым в 1969 г. и впервые изложен в дипломной работе М.Д. Белонина, а затем опубликован в 1971 г. в обзоре "Факторный анализ в нефтяной геологии" (точную ссылку см. в конце Главы II). В 1982 г. раздел этого обзора, посвященный методу "Гусеница", был включен в книгу М.Д. Белонина, В.А. Голубевой и Г.Т. Скуброва "Факторный анализ в геологии", М., "Недра". В дальнейшем этот метод подвергался анализу и развитию в самых различных направлениях. Первым опубликованным практическим применением метода можно считать диссертацию М.М. Кислицина, выполненную под руководством О.М. Калинина и В.М. Ахутина в 1977 г. (см. также статью М.М. Кислицина в настоящем сборнике).

Метод многократно реализовывался на разных типах ЭВМ (начиная с М-20, необходимо отметить значительную роль Т.П. Кистер, которая во многом способствовала пониманию особенностей программной реализации метода), однако не нашел широкого применения до появления ЭВМ серии СМ, допускающих диалог с пользователем. Особенностью метода (как это видно из других статей этого сборника) является существенное использование диалога и визуализации процесса вычислений. Первой интерактивной реализацией метода была программа на СМ-4, реализованная в 1989 г. О.В. Богдановой и В.Н. Солнцевым. В этой же программе впервые "Гусеница" была дополнена весьма существенным блоком восстановления. В 1991–1992 гг. она была перенесена на ПК студентками математико-механического факультета В.Б. Мартыновой и Л.Ю. Котовой под руководством Н.Ю. Баланиной. Появление этой программы послужило толчком к началу целого ряда теоретических, программистских и прикладных исследований, промежуточный итог которых представлен в настоящем сборнике. Отметим также статью В.М. Бухштабера, 1994 г., идеи которой оказали серьезное влияние на ряд ниже публикуемых исследований.

Авторы будут благодарны всем читателям, которые сообщат свои впечатления, замечания или пожелания по адресу:

198904, Санкт-Петербург, Старый Петергоф, Библиотечная пл. 2, Санкт-Петербургский гос. университет, НИИ математики и механики, НКЦ "ТРЕНД".

E-mail: com@trend.niimm.spb.su

Глава I Базовый алгоритм и его основные возможности

I.1 Метод "Гусеница": описание алгоритма

А.А. Жиглявский В.Н. Солнцев

В данном разделе описан базовый алгоритм. Для большей наглядности формальному описанию предшествует его подробное неформальное описание, сопровождающееся дискуссией.

1 Базовый алгоритм, развернутое описание

Рассмотрим временной ряд $\{x_i\}_{i=1}^N$, образованный последовательностью N равнотстоящих значений некоторой (возможно, случайной) функции $f(t)$:

$$x_i = f[i] = f((i-1)\Delta t), \quad \text{где } i = 1, 2, \dots, N.$$

Отметим, что существует несколько вариантов базового алгоритма (см. параграф 2), в данном параграфе мы опишем лишь один из наиболее часто используемых (корреляционный вариант).

Базовый алгоритм метода "Гусеница" можно разбить на четыре этапа.

Этап 1. (*Развертка одномерного ряда в многомерный*)

Выберем некоторое число $M < N$, называемое *длиной гусеницы*, и представим первые M значений последовательности f в качестве первой строки матрицы X . В качестве второй строки матрицы берем значения последовательности с x_2 по x_{M+1} . Последней строкой с номером $k = N - M + 1$ будут последние M элементов последовательности: x_k, x_{k+1}, \dots, x_N :

$$X = (x_{ij})_{i,j=1}^{k,M} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_M \\ x_2 & x_3 & x_4 & \dots & x_{M+1} \\ x_3 & x_4 & x_5 & \dots & x_{M+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k+1} & x_{k+2} & \dots & x_N \end{pmatrix}. \quad (1)$$

Эту матрицу, элементы которой равны $x_{ij} = x_{i+j-1}$, можно рассматривать как M -мерную выборку объема k или M -мерный временной ряд, которому соответствует M -мерная траектория – ломаная в M -мерном пространстве из $k-1$ звена. Отметим, что матрица X в (1) записана в традиционном для прикладной статистике виде "строка – индивид столбец – признак". При изложении теоретических аспектов часто будет удобнее сопоставлять индивиду столбец.

Далее по обычной схеме проводится анализ главных компонент (АГК).

Этап 2. (*Анализ главных компонент: сингулярное разложение выборочной корреляционной матрицы*)

Сначала вычисляются средние арифметические значения и стандартные отклонения по столбцам матрицы X

$$\bar{x}_j = \frac{1}{k} \sum_{i=1}^k x_{i+j-1}; \quad s_j = \sqrt{\frac{1}{k} \sum_{i=1}^k (x_{i+j-1} - \bar{x}_j)^2}. \quad (2)$$

Обратим внимание, что эти величины имеют смысл скользящих среднего и стандарта с прямоугольным окном ширины k .

Обозначим через $X^* = (x_{ij}^*)_{i,j=1}^{k,M}$ матрицу, полученную из X в результате центрирования по столбцам и нормирования на стандарты s_j :

$$x_{ij}^* = (x_{ij} - \bar{x}_j)/s_j \quad i = 1, \dots, k, \quad j = 1, \dots, M. \quad (3)$$

Операции вычитания выборочных средних (центрирование) и деления на выборочные стандарты (нормирование) не являются обязательными. Одна из операций (нормирование) или обе могут отсутствовать. Кроме того, возможно производить центрирование элементов матрицы X как по столбцам, так и по строкам (двойное центрирование).

Для дальнейшей интерпретации процесса работы алгоритма и его результатов полезно представлять геометрический смысл различных шагов алгоритма в двух пространствах: M -мерном пространстве строк \mathcal{B}_M и в k -мерном пространстве столбцов \mathcal{B}_k матрицы X . Обычно матрица X в пространстве \mathcal{B}_M трактуется как множество из k M -мерных точек, а в пространстве \mathcal{B}_k – как набор из M k -мерных векторов. При этом геометрически операция центрирования в пространстве \mathcal{B}_M соответствует параллельному переносу начала координат в центр тяжести множества точек, а в пространстве \mathcal{B}_k – проектированию исходного множества векторов на подпространство, ортогональное вектору $\mathbf{1}_k$ состоящему из k единиц. Операция нормирования в пространстве \mathcal{B}_M соответствует изменению масштабов по всем осям координат так, чтобы величина рассеяния, характеризуемая величиной выборочной дисперсии, стала равной единице; в пространстве \mathcal{B}_k эта операция сводится к приведению всех M векторов к единичной длине.

Далее вычисляется матрица

$$R = \frac{1}{k} X^* (X^*)^T. \quad (4)$$

Если элементы X^* вычисляются по формуле (3), то матрица R является выборочной корреляционной матрицей с элементами

$$r_{ij} = \frac{1}{k} \sum_{l=1}^k \frac{1}{s_i s_j} (x_{i+l-1} - \bar{x}_i)(x_{j+l-1} - \bar{x}_j).$$

(В пространстве \mathcal{B}_k элементы выборочной корреляционной матрицы являются косинусами углов между векторами, полученными после операции центрирования.)

Следующий шаг, как обычно в АГК, состоит в вычислении собственных чисел и собственных векторов матрицы R , т.е. разложение ее

$$R = P \Lambda P^T, \quad (5)$$

где

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_M \end{pmatrix}$$

– диагональная матрица собственных чисел и

$$P = (p_1, p_2, \dots, p_M) = \begin{pmatrix} p_{11} & p_{21} & \dots & p_{M1} \\ p_{12} & p_{22} & \dots & p_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1M} & p_{2M} & \dots & p_{MM} \end{pmatrix}$$

– ортогональная матрица собственных векторов матрицы R . (При этом выполняются следующие соотношения: $P^T = P^{-1}$; $P^T P = P P^T = I_M$, $\Lambda = P^T R P$, $\sum_{i=1}^M \lambda_i = M$, $\prod_{i=1}^M \lambda_i = \det R$.)

Матрицы Λ и P совместно имеют множество интерпретаций, основанных на АГК. В частности, матрицу P можно рассматривать как матрицу перехода к главным компонентам

$$X^* P = Y = (y_1, y_2, \dots, y_M). \quad (6)$$

Если изучается выборка из случайной совокупности, то собственные числа матрицы R являются выборочными дисперсиями соответствующих главных компонент, а квадратные корни из них – выборочными стандартами. Выборочные стандарты допускают геометрическую интерпретацию: они пропорциональны длинам полуосей эллипсоида рассеяния, описываемого матрицей R . Графическое представление собственных чисел и некоторых функций от них в АГК традиционно используется для выявления структуры исследуемой совокупности и отбора и интерпретации главных компонент. В частности, из тождества

$$\sum_{j=1}^M \lambda_j = M$$

следует, что, при умножении всех собственных чисел на $100\% / M$ получаются доли дисперсии в процентах, которые по аналогии с факторным анализом используются при интерпретации отдельных главных компонент как доли общей информации, "объясняемые" каждой главной компонентой. Для интерпретации групп главных компонент часто используют также накопленные доли дисперсии.

Следует отметить, что, если верно часто формулируемое предположение о стационарности изучаемой последовательности x_1, \dots, x_N в том смысле, что любой отрезок ряда длины $s = \min(M, k)$ несет всю основную информацию о структуре изучаемого процесса, то корреляционная матрица R будет иметь почти диагональный вид, т.е. по диагоналям будут стоять достаточно близкие числа. В этом случае можно усреднить матрицу R по диагоналям и получить отрезок длины M оценки автокорреляционной функции $r(\tau)$:

$$r(\tau) = \frac{1}{M - \tau} \sum_{j=1}^{M-\tau} r_{j,j+\tau} \quad \tau = 1, \dots, M, \quad (5)$$

$$r(0) = 1.$$

График этой функции, так же как и графики средних и стандартов s_j , могут дать много полезной информации при интерпретации результатов обработки.

Этап 3. (*Отбор главных компонент*)

Если пронормировать значения главных компонент на выборочные стандарты (при $\lambda_M \neq 0$)

$$Y^* = Y\Lambda^{-1/2} = (y_1^*, y_2^*, \dots, y_M^*),$$

то легко увидеть, что главные компоненты оказываются ортонормированными:

$$Y^{*T}Y^* = I_M,$$

т.е. получается разложение исходного M -мерного процесса на естественные ортогональные компоненты. Каждый из векторов y_j во многих случаях может быть проинтерпретирован, так же как и соответствующий собственный вектор p_j . Дело в том, что вектор y_j можно рассматривать как результат проектирования исходной M -мерной нормированной и центрированной совокупности на направление, определяемое соответствующим собственным вектором p_j .

В то же время, преобразование

$$y_j = X^* p_j$$

очень близко к линейному преобразованию исходного процесса с помощью дискретного оператора свертки, т.е.

$$y_j[l] = \sum_{q=1}^M X_{lq}^* p_{jq} = \sum_{q=1}^M (x_{l+q-1} - \bar{x}_q) \frac{p_{jq}}{s_q} = \sum_{q=1}^M x_{l+q-1} \frac{p_{jq}}{s_q} - \sum_{q=1}^M \bar{x}_q \frac{p_{jq}}{s_q}.$$

Таким образом, процедура "Гусеница" порождает набор линейных фильтров, настроенных на составляющие исходного процесса. При этом собственные векторы матрицы R выступают в роли переходных функций соответствующих фильтров. При отсутствии нормировки и центрирования они в точности соответствуют переходным функциям линейных фильтров.

Визуальное и аналитическое изучение как собственных векторов, так и главных компонент, полученных в результате линейной фильтрации, может дать много интересной информации о структуре изучаемого процесса и свойствах составляющих его слагаемых. Чрезвычайно большую визуальную информацию дает изучение двумерных графиков, аналогичных фигурам Лиссажу, когда по осям x и y откладываются различные пары собственных векторов или главных компонент. Известно, что, если по осям откладывать значения синусоид одной и той же частоты, но с разными фазами, то на плоскости получается эллипс. Из ортогональности собственных векторов и главных компонент следует, что сдвиг фаз между такими парами обязательно будет равен $\pm\pi/2$ и эллипс переходит в окружность.

Как видно из многих примеров, приводимых в этой книге, во многих случаях гусеничные собственные функции образуют пары, аналогичные парам синус-косинус в разложении в ряд Фурье. При этом, как правило, и амплитуды их близки по величине. На двумерных диаграммах для этих пар мы имеем картины – фигуры Лиссажу, похожие на круговые спирали. Это позволяет построить очень простой алгоритм оценки частот этих функций. Пусть

$$\begin{cases} x(t) = a(t) \cos(\omega(t)t + \psi(t)) \\ y(t) = a(t) \sin(\omega(t)t + \psi(t)) \end{cases}$$

— пара таких функций с медленно меняющимися амплитудой, частотой и фазой, рассматриваемые как декартовы координаты такой спирали с параметром t ,

$$\begin{cases} x(t) = r(t) \cos(\omega(t)) \\ y(t) = r(t) \sin(\omega(t)), \end{cases}$$

где

$$r(t) = \sqrt{x^2(t) + y^2(t)} = a(t)\alpha(t) = \arctan \frac{y(t)}{x(t)} = \omega(t)t + \psi(t).$$

Возьмём две последовательности с интервалом τ , точки этой спирали $(r(t), \alpha(t))$ и $(r(t + \tau), \alpha(t + \tau))$, и рассмотрим приращение полярного угла

$$\begin{aligned} \Delta\alpha(t) &= \alpha(t + \tau) - \alpha(t) = \omega(t + \tau)(t + \tau) + \psi(t + \tau) - \omega(t)t + \psi(t) = \\ &= (\omega(t + \tau) - \omega(t))t + \omega(t + \tau)\tau + \psi(t + \tau) - \psi(t) = \\ &= (\omega'(t)\tau + o(\tau))t + (\omega(t) + \omega'(t)\tau + o(\tau))\tau + \psi'(t)\tau + o(\tau). \end{aligned}$$

Если $\omega(t)$ и $\psi(t)$ действительно достаточно медленно меняющиеся функции такие, что $\omega'(t)$ и $\psi'(t)$ имеют порядок $0(\tau^2)$, то

$$\Delta\alpha(t) \approx \omega(t)\tau = \frac{2\pi\tau}{T(t)}, \quad \omega(t) \approx \frac{\Delta\alpha(t)}{\tau} \text{ и } T(t) \approx \frac{2\pi\tau}{\Delta\alpha(t)}.$$

Таким образом, вычисляя приращения полярного угла, мы можем оценить мгновенную частоту $\omega(t)$ и мгновенный период $T(t)$ "гармонического колебания", соответствующего выбранной паре собственных векторов.

Следует отметить две чисто вычислительные проблемы. Первая связана со скачком, близким к $\pm 2\pi$, при вычислении значений $\alpha(t)$, возникающим из-за того, что значения $\arctan(\alpha(t))$ определяются лишь с точностью до 2π . Она легко решается алгоритмически. Вторая, более принципиальная, возникает из-за необходимости учёта частоты Найквиста, возникающей при равномерной дискретизации по времени. Принципиально невозможно оценить частоты, большие $\omega_N = \frac{\pi}{\tau}$. Вообще говоря, следует осторожно интерпретировать гусеничные гармонические компоненты с частотами, близкими к частоте Найквиста. Они визуально воспринимаются как биения, т.е. колебания с частотой ω_N модулированные синусоидой меньшей частоты.

Этап 4. (Восстановление одномерного ряда)

Следующим ключевым элементом метода "Гусеница" является процедура восстановления. Эта процедура основана на следующих достаточно простых соотношениях.

Из ортогональности матрицы P следует, что при умножении матрицы главных компонент Y на P^T восстанавливается матрица X^* , при этом получается разложение

$$X^* = YP^T = (y_1, y_2, \dots, y_M) \begin{pmatrix} p_1^T \\ p_2^T \\ \vdots \\ p_M^T \end{pmatrix} = \sum_{l=1}^M y_l p_l^T = \sum_{l=1}^M X_l^*$$

нормированной и центрированной матрицы X^* в сумму матриц X_l^* , каждая из которых порождена одним собственным вектором матрицы R . Далее производится денормировка X^* с помощью умножения этой матрицы на диагональную

матрицу S , состоящую из выборочных стандартов, и центрирование путем добавления к элементам каждого столбца соответствующих выборочных средних:

$$X = \bar{x} \mathbf{1}_k^T + X^* S = X_0^* + \sum_{l=1}^M X_l^* S = \sum_{l=0}^M X_l^* S.$$

В результате получается исходная матрица диагональной структуры в виде суммы $(M + 1)$ матриц. Переход к исходному ряду формально может быть осуществлен усреднением по побочным диагоналям. Обозначим через \mathcal{A} этот оператор усреднения

$$x = \mathcal{A}(X) = \sum_{l=0}^M \mathcal{A}(X_l^* S). \quad (7)$$

Таким образом, получается разложение исходного временного ряда на сумму $(M + 1)$ рядов. При этом различные слагаемые или группы слагаемых часто могут быть проинтерпретированы.

Отметим сразу одну очень важную особенность описанного метода – его интерактивность, то есть использование диалога исследователя и ЭВМ в процессе применения метода. Причем эта интерактивность не должна рассматриваться как следствие общей тенденции развития современного программного обеспечения персональных ЭВМ. Скорее, наоборот, эффективная реализация алгоритма стала возможной только благодаря возможностям современных ПК. Его интерактивность связана с типично статистическим свойством алгоритма – необходимостью интерпретации промежуточных результатов и управлением работой алгоритма в процессе многоэтапной процедуры обработки. Опыт многолетнего использования различных реализаций алгоритма на ЭВМ разных поколений показал, что чем больше промежуточных результатов удается "увидеть", тем более полным оказывается решение поставленной задачи.

2 Базовый алгоритм, формальное описание

Приведем формальное описание базового алгоритма метода "Гусеница" для анализа одномерного временного ряда $\{x_i\}_{i=1}^N$. Это описание в значительной степени резюмирует описание предыдущего параграфа, включая при этом разные варианты метода.

Алгоритм.

1. Разворотка одномерного ряда в многомерный. Выберем некоторое число $M < N$ (длину гусеницы), положим $k = N - M + 1$ и образуем матрицу $X = (x_{ij})_{i,j=1}^{k,M}$ с элементами $x_{ij} = x_{i+j-1}$, см. (1).

2. Вычисление матрицы X^*

По матрице X вычислим матрицу $X^* = (x_{ij}^*)_{i,j=1}^{k,M}$. Общая формула для элементов этой матрицы такова:

$$x_{ij}^* = (x_{ij} - \mu_j)/\delta_j, \quad i = 1, \dots, k, \quad j = 1, \dots, M. \quad (8)$$

$$\mu_j = \begin{cases} 0 & , \\ \bar{x}_j & , \end{cases} \quad \delta_j = \begin{cases} 1 & , \\ s_j & , \end{cases} \quad (9)$$

где

$$\bar{x}_j = \frac{1}{k} \sum_{i=1}^k x_{ij}, \quad s_j = \sqrt{\frac{1}{k} \sum_{i=1}^k (x_{ij} - \bar{x}_j)^2}.$$

3. Вычисление матрицы R по формуле (4).
4. Сингулярное разложение матрицы R в виде (5).
5. Вычисление главных компонент согласно (6).
6. Отбор главных компонент: выбираем некоторое количество r главных компонент с номерами i_1, \dots, i_r .
7. Восстановление многомерного ряда:

$$\tilde{\mathbf{X}}^* = \sum_{l=1}^r y_{i_l} p_{i_l}^T$$

8. Децентрирование и денормирование (если необходимо) матрицы $\tilde{\mathbf{X}}^*$, результат – матрица $\tilde{\mathbf{X}}$.

9. Ганкелизация (восстановление одномерного ряда путем усреднения по побочным диагоналям $\tilde{\mathbf{X}}$):

$$\tilde{x}_s = \begin{cases} \frac{1}{s} \sum_{i=1}^s \tilde{x}_{s-i+1,i} & 1 \leq s \leq M, \\ \frac{1}{M} \sum_{i=1}^M \tilde{x}_{s-i+1,i} & M \leq s \leq k, \\ \frac{1}{N-s+1} \sum_{i=1}^{N-s+1} \tilde{x}_{k-i+1,i+s-k} & k \leq s \leq N. \end{cases} \quad (10)$$

Способ задания μ_j и δ_j в формулах (9) будет определять название метода:

$\mu_j = 0$ – "Гусеница" без центрирования;

$\mu_j = \bar{x}_j$ – "Гусеница" с центрированием;

$\delta_j = s_j$ – "Гусеница" с нормированием;

$\delta_j = 1$ – "Гусеница" без нормирования.

Так, "Гусеница" с центрированием и нормированием (корреляционная) определяется выбором $\delta_j = s_j$, $\mu_j = \bar{x}_j$. "Гусеницу" с центрированием, но без нормирования ($\delta_j = 1$, $\mu_j = \bar{x}_j$) будем называть ковариационной.

К четырем вариантам алгоритма, определяемым различными комбинациями формулы (9), можно прибавить еще один – "Гусеницу" с двойным центрированием в которой

$$x_{ij}^* = x_{ij} - (\tilde{x}_i + \bar{x}_j - \mu), \quad \tilde{x}_i = \frac{1}{M} \sum_{j=1}^M x_{ij}, \quad \mu = \frac{1}{kM} \sum_{i=1}^k \sum_{j=1}^M x_{ij}.$$

Обычно из контекста ясно, о каком из вариантов "Гусеницы" идет речь.

Заметим, что при программной реализации пункты приведенного выше алгоритма не воспроизводятся непосредственно. Например, этап построения матриц X и X^* отсутствует, а формулы преобразованы к виду, удобному для проведения ускоренных вычислений.

3 Выбор параметров при применении метода "Гусеница"

3.1 Выбор длины гусеницы

Как видно из описания метода, основным управляющим параметром метода является M – длина гусеницы.

При геометрической интерпретации этот параметр является размерностью пространства, в котором исследуется траектория многомерной ломаной линии, в которую переводится исходный временной ряд процедурой гусеница. Естественным условием является $M < N/2$, так как размерность множества k точек (вершин ломаной) в M -мерном пространстве не превосходит $\min(M, k-1)$. Этот подход тесно связан с аналитической интерпретацией метода гусеница как аппроксимацией решения уравнения в конечных разностях с l коэффициентами, см. Главу 2. Можно сказать, что l – это число степеней свободы функции $f(t)$, а следовательно, и соответствующего ей временного ряда. В процедуре "Гусеница" это выражается в том, что при $M > l$ у корреляционной матрицы окажется только l ненулевых собственных чисел. (Для определённости рассмотрим корреляционную "Гусеницу"). Реально исходные данные выражаются числами с конечным (и обычно небольшим) числом цифр, что соответствует наличию ошибок округления. Поэтому вместо $M - l$ нулевых собственных чисел получаются маленькие собственные числа, но в АГК отработаны приемы отсекания малых собственных чисел. Так что, если возникает ситуация, когда часть собственных чисел очень близка к нулю и при этом может быть отделена от остальных, то может быть выдвинута гипотеза о детерминированной структуре временного ряда. Однако, эта ситуация достаточно редкая и при исследовании природных процессов может быть постулирована только для относительно коротких (по сравнению с длиной гусеницы) временных рядов.

В общем случае выбор длины гусеницы существенно зависит от задачи, решаемой этим методом. Рассмотрим три наиболее типичных и исследованных разработчиками случая.

а). Если решается задача анализа исходного временного ряда, например, с целью отыскания скрытых периодичностей с неизвестными периодами, то брать очень короткую гусеницу рискованно, так как может оказаться $M < l$ (в таких ситуациях l обычно неизвестно и может быть достаточно большим). Здесь можно рекомендовать выполнить два прохода. Сначала делается только вычисление собственных чисел при максимальной длине гусеницы, которую допускает имеющаяся у исследователя программная реализация метода и оценивается l . Затем полное исследование ряда проводится при длине гусеницы M , равной или чуть большей оцененного l . Этот "запас" необходим, поскольку в методе гусеница в общем случае не выполняется правило АГК, что информативность главной компоненты пропорциональна величине соответствующего собственного числа. Бывают случаи, когда некоторые периодичности проявляются при весьма малых собственных числах.

б). Если решается более узкая задача, например, сглаживание исходного ряда или выделение периодичности с известным периодом, то работает другой механизм, связанный с "Гусеницей" – механизм фильтрации. Как уже отмечалось в параграфе 1, выделение некоторой главной компоненты в M -мерном представлении исходного временного ряда эквивалентно пропусканию ряда че-

рез фильтр, переходная функция которого совпадает с собственным вектором этой компоненты. Спектральная функция фильтра равна преобразованию Фурье этого собственного вектора. Ширина полосы пропускания зависит от формы переходной функции фильтра и определяется как видом собственного вектора, так и длиной интервала усреднения, т.е. длиной гусеницы M . Чем больше M , тем уже может быть сделана полоса фильтра. Выбор нескольких главных компонент эквивалентен параллельному соединению соответствующих фильтров, что позволяет управлять формой спектральной характеристики.

в). Еще один случай выбора M – при необходимости наилучшего выделения (или исключения) периодического, не обязательно гармонического, колебания с известной частотой и, следовательно, периодом. В этом случае бывает необходимо для лучшего решения задачи изменить длину временного ряда N (естественно, в сторону уменьшения). Из теории метода следует, что периодические колебания наилучшим образом выделяются, если M равно, а N кратно длине периода выделяемого колебания.

Напомним, однако, что метод чрезвычайно устойчив относительно изменения длины гусеницы, и, поэтому, когда речь идет о выборе M , то следует понимать, что проявляется резонансный эффект относительно длины гусеницы не столько в количественном, сколько в качественном смысле.

3.2 Отбор главных компонент

Следующим элементом методики проведения анализа методом "Гусеница", который не может быть выполнен априори, является отбор главных компонент, информативных в том или ином смысле. Напомним, что имеем дело с двумя параллельными представлениями исходного ряда:

- представлением ряда в виде суммы слагаемых (7), которые не обязаны быть ортогональными;
- представлением ряда в виде траектории в M -мерном пространстве, которая в ортонормальном базисе главных компонент (собственных векторов корреляционной матрицы) раскладывается на ортогональные в совокупности компоненты.

Для этих двух представлений имеется 4 набора интерпретируемых объектов:

1). Собственные числа корреляционной матрицы M -мерного представления исходного одномерного ряда. Их интерпретация во многом сходна с АГК.

2). Набор собственных векторов корреляционной матрицы. Поскольку их элементы упорядочены оператором формирования гусеницы (1), их можно изучать как функции времени (точнее, временные ряды длины M).

3). Набор главных компонент M -мерного представления. Они, как и соответствующие им собственные векторы, образуют ортогональную систему и также представимы как функции от номера элемента.

4). Всевозможные восстановленные по разным множествам главных компонент временные ряды, получаемые в результате последовательного применения двух обратных операторов (оператора перехода от главных компонент к исходному M -мерному представлению и оператора \mathcal{A} перехода от M -мерного к одномерному представлению временного ряда).

Следует разделить два крайних случая, где логика отбора несколько отличается:

– При $M \ll N$ мы имеем дело с длинными (с точки зрения гусеницы) рядами. В этом случае более естественным кажется интерпретация собственных векторов как переходных функций линейных фильтров, а соответствующих главных компонент как результатов действия этих фильтров. Здесь применима терминология, связанная с линейной фильтрацией такая, как "ширина полосы пропускания", фильтры "высоких и низких частот" и тому подобные термины спектрального подхода.

– При $M \simeq N/2$ метод "Гусеница" можно интерпретировать как метод аппроксимации исходного временного ряда рядами конечного ранга, см. Главу II. Здесь более уместен геометрический подход к интерпретации отдельных шагов и результатов применения метода. Наиболее простым является поиск гармонических компонент исследуемого процесса. Как уже отмечалось выше, каждому синусоидальному слагаемому ряда соответствуют две главные компоненты, имеющие вид отрезков синуса и косинуса одной и той же частоты. Их легко обнаруживать по двумерным графикам для пар собственных векторов корреляционной матрицы или пар соответствующих главных компонент. Кроме того, для таких пар собственные числа обычно оказываются достаточно близкими, поэтому оказывается возможным угадывать эти пары по графику собственных чисел или квадратных корней или логарифмов из них.

Глава II О методах типа "Гусеница". Основные идеи и способы реализации

Д.Л. Данилов В.Н. Солнцев

Авторам неизвестно об упоминании в литературе методов типа "Гусеница" в работах, опубликованных ранее 1971 г. По-видимому, впервые подобный метод упоминается в работе О.М. Калинина (см. сборник Белонина и др. [4], а также [3]), где предлагается алгоритм для обнаружения скрытых периодичностей в геологических данных, использующий технику главных компонент. В современной зарубежной литературе описан достаточно широкий класс методов, алгоритмически и идеально близких к методу "Гусеница". Большая часть этих методов известна как *Singular Spectrum Analysis* (SSA), в русском переводе *анализ сингулярного спектра* (АСС). Его теоретическими основами являются, с одной стороны, классический анализ главных компонент, а с другой — быстро развивающаяся в последнее время теория динамических систем. Кроме того, имеются определенные аналогии метода "Гусеница" с динамическим анализом Фурье и вейвлет-анализом. В данной главе сделана попытка проследить указанные истоки и аналогии.

1 Анализ главных компонент

Так как всё дальнейшее изложение в большой мере опирается на анализ главных компонент (АГК), опишем в данном разделе классическую схему АГК.

Анализ главных компонент как статистический метод восходит к работам Hotelling [38] и Rao [11]. Одним из первых систематических изложений прикладных аспектов метода главных компонент является статья Rao [54]. Опишем кратко теоретико-вероятностную схему, соответствующую разложению многомерной случайной величины на главные компоненты (ГК).

Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — некоторое вероятностное пространство. Рассмотрим в нем d -мерный случайный вектор $\bar{\xi}' = (\xi'_1, \dots, \xi'_d)^T : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^d, \mathcal{B}^d)$.

Будем предполагать, что дисперсии компонент вектора $\bar{\xi}'$ положительны: $\text{Var } \xi'_i = \sigma_i'^2 > 0, i = 1, \dots, d$, и обозначим через $D \bar{\xi}'$ диагональную матрицу из этих дисперсий: $D \bar{\xi}' = \text{diag}(\sigma_1'^2, \dots, \sigma_d'^2)$. Рассмотрим следующие варианты:

1. Ковариационные главные компоненты. Рассмотрим центрированный вектор $\bar{\xi}'$, который обозначим через $\bar{\xi}$:

$$\bar{\xi} = \bar{\xi}' - E \bar{\xi}'.$$

В этом случае анализ производится по \mathbf{C}_ξ — ковариационной матрице $\bar{\xi}$ (она же является ковариационной для $\bar{\xi}'$).

2. Нецентральные ковариационные главные компоненты. В этом случае анализ производится по $\bar{\mathbf{C}}_\xi$ — матрице вторых моментов вектора $\bar{\xi}$ (другой термин для $\bar{\mathbf{C}}_\xi$ — нецентральная ковариационная матрица).

3. Корреляционные главные компоненты. Рассмотрим центрированный и нормированный вектор $\bar{\xi}'$, который обозначим через $\bar{\xi}$:

$$\bar{\xi} = (\mathbf{D}\bar{\xi}')^{-1/2}(\bar{\xi}' - \mathbf{E}\bar{\xi}').$$

В этом случае анализ производится по \mathbf{R}_ξ — ковариационной матрице ξ (она же является корреляционной для $\bar{\xi}'$).

Рассмотрим подробнее случай корреляционных главных компонент. Пусть \mathbf{R}_ξ — ковариационная матрица вектора ξ (то есть корреляционная матрица вектора $\bar{\xi}'$). Обозначим через $r = \text{rank } \mathbf{R}_\xi$ ранг матрицы \mathbf{R}_ξ . Пусть $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d \geq 0$ — упорядоченная по убыванию последовательность собственных чисел матрицы \mathbf{R}_ξ , а $\{V_1, V_2, \dots, V_d\}$ — соответствующая этим собственным числам ортонормированная система собственных векторов. Очевидно, что $\lambda_i > 0$ при $i = 1, \dots, r$ и $\lambda_i = 0$ при $i = r + 1, \dots, d$. Обозначим

$$\mathbf{V} = (V_1 : \dots : V_d); \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_d).$$

По определению, для матриц \mathbf{R}_ξ , \mathbf{V} и Λ выполнено соотношение $\mathbf{R}_\xi \mathbf{V} = \mathbf{V} \Lambda$. Кроме того,

$$\mathbf{R}_\xi = \mathbf{V} \Lambda \mathbf{V}^T = \sum_{i=1}^d \lambda_i V_i V_i^T.$$

Будем называть вектор

$$\mathbf{U} = (U_1, \dots, U_d)^T = \mathbf{V}^T \bar{\xi} = (V_1^T \bar{\xi}, \dots, V_d^T \bar{\xi})$$

вектором ковариационных главных компонент вектора $\bar{\xi}$.

Имеют место следующие простейшие свойства главных компонент.

1. $\mathbf{E} \mathbf{U} = \mathbf{E} \mathbf{V}^T \bar{\xi} = \mathbf{V}^T \mathbf{E} \bar{\xi} = 0$.
2. $\text{Var} U_i = \text{Var} V_i^T \bar{\xi} = V_i^T \mathbf{R}_\xi V_i = V_i^T \sum_{j=1}^d \lambda_j V_j V_j^T V_i = \lambda_i V_i^T V_i V_i^T V_i = \lambda_i$.
3. $\mathbf{E} U_i U_j = \mathbf{E} V_i^T \bar{\xi} V_j^T \bar{\xi} = \mathbf{E} V_i^T \bar{\xi} \bar{\xi}^T V_j = V_i^T \mathbf{E} \bar{\xi} \bar{\xi}^T V_j = V_i^T \mathbf{R}_\xi V_j$
 $= V_i^T \sum_{k=1}^d \lambda_k V_k V_k^T V_j \lambda_j V_i^T V_j V_j^T V_i + \lambda_i V_i^T V_i V_i^T V_j = 0$,

так как $V_i^T V_j = 0$ при $i \neq j$.

Таким образом, главные компоненты являются некоррелированными случайными величинами с дисперсиями $\text{Var} U_i = \lambda_i$, $i = 1, \dots, d$.

4. Вектор \mathbf{U} зависит от выбора базиса в каждом из собственных подпространств, отвечающих собственным числам λ_i . В случае простых собственных чисел соответствующие главные компоненты определяются с точностью до знака. Для кратных собственных чисел произвол в выборе базиса гораздо шире.
5. При $r \leq i \leq d$ $\mathbf{E} U_i = 0$ и $\text{Var} U_i = \lambda_i = 0$, так что $U_i = 0$ почти всюду.

Обозначим

$$U_i^0 = \begin{cases} \frac{1}{\sqrt{\lambda_i}} U_i & i \leq r, \\ 0 & r < i \leq d. \end{cases}$$

Будем называть вектор $\mathbf{U}^0 = (U_1^0, \dots, U_d^0)$ вектором нормированных ковариационных главных компонент вектора $\bar{\xi}$.

Пусть Λ^{-1} — обобщенная обратная матрица к матрице Λ . Тогда, очевидно,

$$\mathbf{U}^0 = \Lambda^{-1/2} \mathbf{V}^T \bar{\xi}.$$

По вектору \mathbf{U} мы можем однозначно определить вектор $\bar{\xi}$:

$$\bar{\xi} = \mathbf{V} \mathbf{U} = V_1 U_1^T + \dots + V_d U_d^T,$$

а следовательно, и вектор $\bar{\xi}'$:

$$\bar{\xi}' = \mathbf{E} \bar{\xi}' + V_1 (\mathbf{D} \bar{\xi}')^{1/2} U_1^T + \dots + V_d (\mathbf{D} \bar{\xi}')^{1/2} U_d^T \quad (1)$$

В дальнейшем будем называть процесс определения вектора $\bar{\xi}'$ по набору главных компонент U_1, \dots, U_d , заданный формулой (1), *восстановлением* вектора $\bar{\xi}'$ по главным компонентам U_1, \dots, U_d . Аналогично можно определить восстановление по любому другому набору главных компонент. Вектор $W_i = V_i U_i^T$ называется вкладом i -ой главной компоненты в вектор $\bar{\xi}$, а вектор $W'_i = V_i (\mathbf{D} \bar{\xi}')^{1/2} U_i^T$ называется вкладом i -ой главной компоненты в вектор $\bar{\xi}'$. Так как векторы U_1, \dots, U_d некоррелированы, то и вклады различных компонент в вектора $\bar{\xi}$ и $\bar{\xi}'$ некоррелированы.

Рассмотренной теоретико-вероятностной схеме соответствует следующая конечно-выборочная конструкция.

Рассмотрим набор d -мерных векторов X'_1, \dots, X'_n . В дальнейшем этот набор мы будем трактовать как повторную выборку из некоторого генерально-го распределения. Построим следующее вероятностное пространство. Пусть $\Omega = \{X'_1, \dots, X'_n\}$ — дискретное пространство элементарных событий

$$\omega_i = X'_i, \quad i = 1, \dots, n.$$

Пусть $\mathcal{F} = 2^\Omega$ — множество всех подмножеств Ω и \mathbf{P} — такая вероятностная мера, что

$$\mathbf{P}(\{\omega_i\}) = \mathbf{P}(\{X'_i\}) = \frac{1}{n}, \quad i = 1, \dots, n.$$

Будем называть $(\Omega, \mathcal{F}, \mathbf{P})$ каноническим выборочным вероятностным пространством, соответствующим набору X'_1, \dots, X'_n . Пусть случайный вектор $\bar{\xi}' = (\xi'_1, \dots, \xi'_d)^T : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^d, \mathcal{B}^d)$ определяется равенством

$$\bar{\xi}'(\omega_i) = \omega_i = X'_i, \quad i = 1, \dots, n.$$

Будем называть этот вектор выборочным случайнм вектором, соответствующим набору X'_1, \dots, X'_n . Вектор $\bar{\xi}'$ имеет дискретное равномерное распределение:

$$\mathbf{P}\{\bar{\xi}' = X'_i\} = \frac{1}{n}, \quad i = 1, \dots, n.$$

Его математическое ожидание равно:

$$\mathbb{E} \bar{\xi}' = \frac{1}{n} \sum_{i=1}^n X'_i = \bar{\mu}.$$

Обозначим через X' матрицу со столбцами X'_1, \dots, X'_n . Тогда матрица X' полностью описывает распределение вектора $\bar{\xi}'$. Ковариационная матрица вектора $\bar{\xi}'$ в данной ситуации равна

$$\mathbf{R}_{\xi'} = \frac{1}{n} \sum_{i=1}^n X_i X_i^T, \quad X_i = X'_i - \bar{\mu},$$

и называется выборочной ковариационной матрицей.

2 АГК и анализ стационарных временных рядов

Взаимосвязь между анализом главных компонент и классическими методами анализа стационарных временных рядов подробно изложена в книге Бриллинджера [5]. Кратко рассмотрим эту взаимосвязь.

Пусть случайная величина Y представляет собой отрезок наблюдений некоторого действительного стационарного временного ряда $X(t)$, $t = -T \dots T$. В этом случае

$$Y = [X(-T), \dots, X(0), \dots, X(T)]^T.$$

Предположим, что $c_{XX}(u)$ есть автоковариационная функция ряда $X(t)$. Тогда ковариационная матрица Y имеет вид

$$\Sigma_{YY} = \begin{pmatrix} c_{XX}(0) & c_{XX}(1) & \dots & c_{XX}(2T) \\ c_{XX}(-1) & c_{XX}(0) & \dots & c_{XX}(2T-1) \\ \vdots & \vdots & \ddots & \vdots \\ c_{XX}(-2T) & c_{XX}(-2T+1) & \dots & c_{XX}(0) \end{pmatrix},$$

и главные компоненты случайной величины Y будут выражаться через собственные числа и собственные вектора этой матрицы. Согласно [5], §3.7, при выполнении условия $\sum_u |u| |c_{XX}(u)| < +\infty$ эти величины будут близки (при $T \rightarrow \infty$) к величинам

$$\lambda_s = \frac{1}{2T+1} \sum_{t=-2T}^{2T} c_{XX}(t) \exp(2\pi i t s),$$

$$V_s = \left(\frac{1}{\sqrt{2T+1}} \exp\left(\frac{-2\pi i t s}{2T+1}\right) \right)_{t=-T}^T, \quad s = -T \dots T,$$

соответственно. Поэтому главные компоненты вектора Y будут приближенно равны

$$U_s = \left(\frac{1}{\sqrt{2T+1}} \sum_{t=-2T}^{2T} \exp\left(\frac{-2\pi i t s}{2T+1}\right) X(t) \right), \quad s = -T \dots T.$$

Последнее выражение равно $d_x^T(2\pi s/(2T+1))$, где

$$d_x^T(\lambda) = \sum_{t=0}^{T-1} X(t) \exp(-i\lambda t)$$

— конечное преобразование Фурье отрезка стационарного временного ряда $X(t)$. Стандартная операция восстановления случайной величины Y по набору первых k главных компонент приводит к следующей аппроксимации отрезка ряда $X(t)$:

$$\hat{X}^{(k)}(t) = \frac{1}{2T+1} \sum_{s=1}^k d_x^T\left(\frac{2\pi s}{2T+1}\right) \exp\left(\frac{2\pi i ts}{2T+1}\right).$$

Отметим, что если, в некотором смысле,

$$\frac{1}{2\pi} \int_0^\lambda d_x^T(\lambda) dZ_x(\lambda) \xrightarrow{T \rightarrow \infty} Z_x(\lambda),$$

то

$$\hat{X}^{(k)}(t) \Rightarrow \int_0^{2\pi} \exp(it\lambda) dZ_x(\lambda),$$

что является представлением Крамера для $X(t)$. Кроме того, спектр матрицы Σ_{YY} аппроксимируется числами $2\pi f_{XX}(2\pi s/(2T+1))$, где $f_{XX}(\lambda)$ — спектр мощности ряда $X(t)$. Заметим, что упоминаемый ниже Анализ Сингулярного Спектра фактически является конечно-выборочной версией описанной выше генеральной схемы.

3 Главные компоненты в теории динамических систем

Подходы, так или иначе использующие преобразование исходного ряда в многомерную выборку, развиваются методами теории динамических систем. Так, об использовании матрицы "гусеничного" типа (см. раздел I.1) при анализе нелинейных сигналов впервые упоминается в работе [48], где описывается метод восстановления хаотических аттракторов по наблюденному временному ряду. Математическое описание этих методов в применении к нелинейным системам впервые опубликовал Takens (см. [60]), который доказал ряд теорем о представимости пространства состояний достаточно широкого класса динамических систем с помощью метода многомерной развертки наблюдаемого одномерного ряда. Заметим, что Takens фактически обобщил ранний результат Whitney [64]. В последующие годы метод применялся, в основном, специалистами в области нелинейной физики. Так, в обзоре Eckmann, Ruelle [25], содержится ряд результатов по изучению и экспериментальной оценке размерности траекторий динамических систем, энтропии и характеристических показателей этих систем, а также соотношений между этими величинами. Основным методом, используемым этими авторами при обработках результатов экспериментов, является метод развертки временных рядов. Авторы статьи предложили использовать

при анализе сигналов повторяющиеся конфигурации точек в пространстве развертки, что дает возможность не только аппроксимировать аттрактор, но и извлекать некоторую информацию о динамике системы. В работе Sauer и др. [57] излагаются теоретические основы техники развертки временных рядов. В частности, исследованы условия, необходимые для однозначной представимости пространства состояний исходной динамической системы при развертке её траекторий. Авторы определяют *состояние* детерминированной динамической системы как информацию, необходимую для определения всего будущего поведения системы. Следующая простая математическая модель полезна для понимания этого определения. Рассмотрим систему n обыкновенных дифференциальных уравнений с n неизвестными. Предположим также, что выполняются условия гладкости, такие что решение такой системы существует и единственно. Это означает, что через любую точку $\mathbf{a} = (x_1, \dots, x_n)$ в \mathbb{R}^n проходит единственная траектория, имеющая \mathbf{a} начальным условием. Поэтому эволюция системы полностью определяется начальным вектором \mathbf{a} . В этом случае *пространством состояний* системы является \mathbb{R}^n .

Предположим теперь, что наблюдаемый временной ряд представляет собой функцию состояния системы, т.е. для текущего состояния \mathbf{a} наблюдаемое (в данный момент) значение временного ряда есть $h(\mathbf{a})$. В этой связи важной проблемой является проблема представления, вообще говоря, абстрактного состояния системы \mathbf{a} в виде, доступном для работы с ним. Часто при этом вектор состояния \mathbf{a} замещается неким другим вектором, обозначим его для определенности \mathbf{b} . Пусть $F_t(\mathbf{a})$ – состояние системы спустя время t . Рассмотрим процедуру *многомерной развертки* временного ряда (см. [60, 57]).

Пусть A – компактное конечномерное пространство состояний системы. Оно может включать в себя периодические орбиты, хаотические аттракторы и др. Фиксируем вещественное $\tau > 0$. Для любого $\mathbf{a} \in A$ определим m -мерный вектор

$$\mathbf{b} = [h(\mathbf{a}), h(F_{-\tau}(\mathbf{a})), \dots, h(F_{-(m-1)\tau}(\mathbf{a}))]. \quad (2)$$

Этот вектор назовем *вектором развертки* временного ряда (также распространено название *вектор задержек*). Заметим, что \mathbf{b} есть наблюдаемый отрезок временного ряда, соответствующего функции измерения h , т.е.

$$\mathbf{b} = [x_t, x_{t-\tau}, \dots, x_{t-(m-1)\tau}], \quad (3)$$

где $x_t = h(\mathbf{a})$ в момент времени t (\mathbf{a} – соответствующее состояние системы). Известно, что при разумных условиях, наложенных на динамику системы F_t , отображение $D(\mathbf{a}) = \mathbf{b}$ является взаимно однозначным, если m больше чем удвоенная фрактальная размерность A , и функция h выбрана соответствующим образом. Последнее утверждение известно как теорема о фрактальном распространении вложения посредством координатной развертки¹ (см. [57]). Некоторые результаты относительно правильного выбора времени задержки τ содержатся в [46]. Любая структура гладкого многообразия на A также сохраняется при отображении D . В [57] показано, что если A – фрактальный аттрактор, то любая структура многообразия (например, неустойчивое многообразие) полностью воспроизводится в $D(A)$. Поэтому, например, положительные показатели Ляпунова на A могут быть оценены по множеству $D(A)$, даже если A – фрактальное множество, а не гладкое многообразие.

¹в оригинале “Fractal Delay Coordinate Embedding Prevalence Theorem”

Часто в качестве отображения D рассматривают не просто формулу (2), а более сложные отображения. В общем виде этот подход может быть описан как *фильтрованная развертка* временного ряда, т.е. замена (2) более общим выражением

$$\hat{\mathbf{b}} = M[h(\mathbf{a}), h(F_{-\tau}(\mathbf{a})), \dots, h(F_{-(w-1)\tau}(\mathbf{a}))]^T, \quad (4)$$

где $M - m \times w$ -матрица ранга m , $w \geq m$. В терминах исходного ряда можно записать

$$\hat{\mathbf{b}} = M[x_t, x_{t-\tau}, \dots, x_{t-(w-1)\tau}]^T, \quad (5)$$

т.е. вектор $\hat{\mathbf{b}}$, соответствующий состоянию \mathbf{a} , является линейной комбинацией некоторых векторов \mathbf{b} . Часто в качестве M используют также композицию прямого и обратного преобразований Фурье (см. [56]) и др.

Ряд алгоритмов для предсказания временных рядов, использующих технику развертки, описывается в работах Farmer и Sidorowich [26, 27], а также Casdagli [21]. В первой работе описывается техника коротко-временных прогнозов хаотических временных рядов, использующая локальную аппроксимацию конфигураций точек в пространстве развертки. Алгоритмы выделения сигнала, развивающие идеи Eckmann и Ruelle, описаны также в работах Sauer, см., например, [56].

Другим подходом к выделению полезного сигнала из зашумленного, в том числе с целью последующего прогноза, является подход, использующий регрессионные мотивы. При этом часто используется следующая идея. Исходное состояние \mathbf{a} динамической системы полностью определяет её будущее поведение, поэтому, если система вернулась в исходное состояние, то её эволюция полностью повторится. Однако на практике исходный ряд зашумлен и *наблюденный* временной ряд не воспроизводится точно. Возможно, однако, исследовать прошлые состояния системы с целью поиска близких, подразумевая, что эти близкие состояния не совпадают полностью лишь из-за присутствия шума. После нахождения набора таких состояний предсказание строится как регрессия будущего наблюдения на прошедшие (см., например, [56] и обзор [26]).

4 Метод АСС и его модификации

Использование разложения выборки развертки по собственным векторам ковариационной матрицы восходит к работам Broomhead [16, 17, 18]. Автор рассматривает применение т.н. *метода задержек* для преобразования (в оригинале — *развертки*) исходного временного ряда $(x_i)_{i=1}^N$ в *матрицу траекторий* X по правилу

$$X = M^{-1/2} \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_M^T \end{pmatrix}, \text{ где } x_i^T = (x_i, \dots, x_{i+n-1}), n = N - M + 1.$$

В дальнейшем Broomhead строит т.н. сингулярное разложение матрицы траекторий по системам собственных векторов *ковариационной* $\Xi = X^T X$, и *структурной* $\Theta = X X^T$ матриц развертки временного ряда:

$$X = S \Sigma C^T.$$

Здесь S и C — матрицы собственных векторов, соответствующих ненулевым собственным числам, структурной и ковариационной матрицы развертки соответственно, а $\Sigma = \text{diag}(\lambda_1, \dots, \lambda_n)$ — диагональная матрица собственных чисел Ξ , упорядоченных по убыванию их величин $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, т.е. :

$$\Xi C = C \Sigma^2, \quad \Theta S = S \Sigma_S^2, \quad \Sigma_S = \text{diag}(\lambda_1, \dots, \lambda_n, 0, \dots, 0).$$

В работе [16] впервые предложено производить восстановление траекторий наблюденных аттракторов путем проектирования их в подпространство нескольких первых собственных векторов, предположительно соответствующих свободному от шумов сигналу. Заметим, что вектора S фактически являются нормированными главными компонентами матрицы траекторий X , рассматриваемой как матрица многомерной выборки.

Рассмотренные ранее два подхода к анализу эмпирических временных рядов (т.е. АГК и методы теории динамических систем) являются теоретическими основами Анализа Сингулярного Спектра (ACC) (Singular Spectrum Analysis). Развернутое описание ACC дается в работе [62]. (Упоминание об ACC можно найти в таких работах по теории обработки нелинейных сигналов, как Пристли [53] и Тонг [59].) Коснемся кратко основных положений метода.

ACC, как метод анализа данных, в течении многих лет использовался при изучении цифровых сигналов [45, 51]. Этот метод был представлен Colebrook [13] в океанографии и в нелинейной динамике Boomhead и King [16], и Fraedrich [28]. ACC основан на анализе главных компонент в векторном пространстве развертки временных рядов. Классический анализ главных компонент [52] используется при анализе многомерных временных рядов для нахождения главных осей последовательности M -мерных векторов ($X_{ij}, 1 \leq i \leq N, 1 \leq j \leq M$) и разложения её относительно ортонормированного базиса ($E^k, 1 \leq k \leq M$)

$$X_{ij} = \sum_{k=1}^M a_i^k E_j^k. \quad (6)$$

Коэффициенты разложения a_i^k называются *главными компонентами* (ГК), а базисные вектора E^k — *эмпирическими ортогональными функциями* (ЭОФ). Вектора E^k — собственные вектора кросс-ковариационной матрицы последовательности $\{X_i\}$, $X_i = \{X_{i1} \dots X_{iM}\}$, $i = 1, \dots, N$.

В случае так называемого *одномерного ACC*, если обозначить переменные скалярного ряда через $(x_i, 1 \leq i \leq N)$ то можно написать аналогичное разложение

$$x_{i+j} = \sum_{k=1}^M a_i^k E_j^k, \quad 0 \leq i \leq N - M, \quad 1 \leq j \leq M. \quad (7)$$

Аналогия с (6) получается преобразованием одномерного временного ряда x_i в многомерный $X_i = (x_{i+1}, x_{i+2}, \dots, x_{i+M})$. Помимо этого не существует формального отличия между двумя разложениями (6) и (7). Параметр M в разложении (7) называется *длиной окна* или *вложенной размерностью* и выбирается пользователем, в отличие от классического анализа главных компонент, где M — фиксированная размерность векторов данных.

Вектора E^k — собственные векторы матрицы T_x , тёплевой матрицы, содержащей в i -й строке, j -ом столбце значение кросс-ковариационной функции

вектора x в точке $i - j$. В обоих случаях разложение ковариационной матрицы по собственным векторам связано с сингулярным разложением прямоугольной матрицы траекторий; в случае АСС матрица траекторий содержит $N - M + 1$ векторов X_i в качестве столбцов. Сингулярное разложение — это класс алгоритмов большей общности в числовой линейной алгебре; необходимо отличать его от АСС, который является методологией для анализа временных рядов.

Для многомерного АСС [17, 44, 63], с наблюдаемыми L -мерными векторами данных $X_{l,i}$, $1 \leq l \leq L$, $1 \leq i \leq N$, разложение принимает вид

$$X_{l,i+j} = \sum_{k=1}^{LM} a_i^k E_{l,j}^k, \quad 1 \leq l \leq L, 1 \leq j \leq M. \quad (8)$$

Здесь вектор состояний, рассматриваемый в момент времени i , имеет вид

$$(X_{1,i+1}, X_{1,i+2}, \dots, X_{1,i+M}, X_{2,i+1}, \dots, X_{2,i+M}, \dots, X_{L,i+1}, \dots, X_{L,i+M}).$$

Величина M всё ещё имеет смысл длины окна. K -ый базисный вектор является собственным вектором блочной матрицы Тёплица T_x , содержащей кросс-ковариационные коэффициенты между различными компонентами l и со сдвигами от 0 до $M - 1$.

Формулы (6)-(8) являются следствиями общего биортогонального разложения Карунена-Лоэва [47]. Теорема разложения Карунена-Лоэва является краеугольным камнем АСС. Она, в свою очередь, основана на автоковариационной матрице наблюдаемого процесса x . Без потери общности можно предположить, что x имеет нулевое математическое ожидание. Автоковариационная матрица T_x обладает тёплицевой структурой:

$$T_x = \begin{pmatrix} c(0) & c(1) & \cdot & \cdot & \cdot & c(M-1) \\ c(1) & c(0) & c(1) & \cdot & \cdot & \cdot \\ \cdot & c(1) & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & c(1) \\ c(M-1) & \cdot & \cdot & \cdot & c(1) & c(0) \end{pmatrix}, \quad (9)$$

где $c(j)$, $0 \leq j \leq M - 1$, является ковариацией x для сдвига j .

Существуют различные способы оценивания T_x [41, 14, 40]. Среди наиболее часто используемых — оценка Юла-Уолкера

$$c(j) = \frac{1}{N} \sum_{i=1}^{N-j} x_i x_{i+j}, \quad (10)$$

а также оценка

$$c(j) = \frac{1}{N-j} \sum_{i=1}^{N-j} x_i x_{i+j}. \quad (11)$$

Алгоритм оценки авторегрессионных коэффициентов Бёрга (см. [20]), связанный с методом максимальной энтропии, также оценивает $c(j)$. Эта оценка опять является средним, как и в формулах (10-11), но с большими весами по отношению к середине ряда, чем к концам.

Оценка (10), используемая Боксом и Дженкинсоном [14], сильно смещена когда число данных мало; оценка (11) имеет большую дисперсию, но меньшее смещение [41]. Когда N большое, оценка Бёрга также может быть смещена при больших j , если длина периодов больше, чем N . Так, если в системе присутствуют очень низкие частоты или тренд, то оценка (11) смещена лишь незначительно, тогда как обе, (10) и оценка Бёрга, — сильно. Если такие частоты устраниТЬ, то оценка Бёрга и оценка (11) становятся эквивалентными.

Матрица Тёплица T_x симметрична и неотрицательна. Её собственные значения λ_k строго положительны, кроме тех случаев, когда ошибки измерений отсутствуют и данные поступают от динамической системы со строго квазипериодическим поведением; в последнем случае все собственные числа, за исключением конечного числа (равного удвоенному числу частот), нули. Они расположены в убывающем порядке, и собственные вектора E^k нормированы следующим образом

$$\sum_{k=1}^M E_j^k E_l^k = \delta_{jl}, \quad 1 \leq j \leq M, 1 \leq l \leq M, \quad (12)$$

Согласно формуле разложения ковариационной матрицы,

$$\sum_{k=1}^M \lambda_k E_j^k E_l^k = (T_x)_{jl} = c(j-l), \quad 1 \leq j \leq M, 1 \leq l \leq M. \quad (13)$$

K -ая ГК определяется как результат проектирования матрицы развертки исходного ряда на подпространство, порожденное k -ой ЭОФ:

$$a_i^k = \sum_{j=1}^M x_{i+j} E_j^k, \quad 0 \leq i \leq N - M. \quad (14)$$

Т.о., главные компоненты являются взвешенными средними процесса x и их размерность равна $N - M + 1$. Обозначив через B оператор обратного сдвига, а через Ψ_k — полином

$$\Psi_k(\zeta) = E_M^k + E_{M-1}^k \zeta + \dots + E_1^k \zeta^{M-1}, \quad (15)$$

получаем, что k -ая ГК может быть записана следующим образом

$$a_i^k = \Psi_k(B) x_{i+M}. \quad (16)$$

Главные компоненты являются фильтрованным вариантом исходного ряда. Тем не менее, они не позволяют единственным образом разложить сигнал в сумму различных компонент. Действительно, в разложении (7) каждый член зависит от индекса j , который меняется от 1 до M . Следовательно, есть M различных путей восстановления компонент сигнала, которые в общем случае не дают одного и того же результата. Другая проблема в использовании (7) для фильтрации, и, в особенности, для фильтрации в реальном времени и предсказания, это то, что получается ряд длины $N - M + 1$, а не длины N , как ожидалось. Мы покажем, как извлечь наилучшим путём ряд длины N , относящийся к данному множеству собственных элементов, которые мы назовём реконструированными компонентами (РК).

Рассмотрим подмножество \mathcal{A} собственных элементов, по которому производится восстановление. По аналогии с (7) построим ряд $y = R_{\mathcal{A}}x$ длины N такой, что величина

$$H_{\mathcal{A}}(y) = \sum_{i=0}^{N-M} \sum_{j=1}^M \left(y_{i+j} - \sum_{k \in \mathcal{A}} a_i^k E_j^k \right)^2 \quad (17)$$

минимальна. Иными словами, оптимальный ряд y является ближайшим, в среднеквадратичном смысле, к проекции расширенного ряда X на ЭОФ с индексами из \mathcal{A} . Решение этой минимизационной задачи описывается формулами

$$(R_{\mathcal{A}}x)_i = \frac{1}{M} \sum_{j=1}^M \sum_{k \in \mathcal{A}} a_{i-j}^k E_j^k, \text{ для } M \leq i \leq N - M + 1, \quad (18)$$

$$(R_{\mathcal{A}}x)_i = \frac{1}{i} \sum_{j=1}^i \sum_{k \in \mathcal{A}} a_{i-j}^k E_j^k \text{ для } M \leq i \leq M - 1, \quad (19)$$

$$(R_{\mathcal{A}}x)_i = \frac{1}{N - i + 1} \sum_{j=i-N+M}^M \sum_{k \in \mathcal{A}} a_{i-j}^k E_j^k \text{ для } N - M + 2 \leq i \leq N. \quad (20)$$

(Это не что иное как операция "гусеничного восстановления" (см. раздел I.1), состоящая в усреднении по побочным диагоналям.)

Если \mathcal{A} состоит из единственного индекса k , то ряд $R_{\mathcal{A}}$ называется k -й реконструированной компонентой и обозначается через x^k . РК обладают свойством аддитивности, т.е.

$$(R_{\mathcal{A}}x)_i = \sum_{k \in \mathcal{A}} x^k. \quad (21)$$

В частности, исходный ряд x может быть разложен в сумму его РК:

$$x = \sum_{k=1}^M x^k. \quad (22)$$

Заметим, что, несмотря на линейную форму записи, разложение ряда x в сумму x^k является существенно нелинейным, поскольку собственные векторы E^k зависят от x нелинейно. Главным преимуществом использования РК вместо ГК является возможность локализации точного момента времени какого либо события (например импульсного "всплеска") в исходном сигнале, особенно в случае когда длина такого "всплеска" в сигнале невелика.

5 АСС в непрерывном случае

В статье Vautard и Ghill [61] сформулирована модель АСС, относящаяся к анализу непрерывных процессов.

Рассмотрим случай непрерывного эргодического процесса $X(t)$ с нулевым средним на бесконечном интервале. Рассмотрим задачу нахождения собственных функций и собственных значений линейного интегрального оператора A , действующего на функции $\rho(t)$ из $L^2(-\tau, \tau)$ по формуле

$$A\rho(t) = \frac{1}{2\tau} \int_{-\tau}^{\tau} C_x(t-s)\rho(s)ds, \quad (23)$$

где $C_x(t)$ представляет собой автокорреляционную функцию процесса X . Собственные значения λ и функции $\rho(t)$ являются решениями задачи

$$A\rho = \lambda\rho. \quad (24)$$

Пространство $H = L^2(-\tau, \tau)$ играет роль пространства развертки и является гильбертовым со скалярным произведением

$$(f, g) = \frac{1}{2\tau} \int_{-\tau}^{\tau} f(t)g(t)dt. \quad (25)$$

Пусть $C_x(t)$ не обращается в 0 тождественно, и принадлежит $L^2(-\tau, \tau)$. Поскольку функция $C_x(t)$ симметрична относительно 0, то A — симметрический неотрицательный оператор, т.е. $(f, Ag) = (f, Ag)$ для любых $f, g \in H$ и $(f, Af) \geq 0$. В этом случае из общей теории операторов Гильберта-Шмидта (см., например, [55]) следует существование не более чем счетного набора собственных величин $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ и факт, что соответствующие собственные функции $\rho_k(t)$ образуют ортонормированный базис H . Эти функции также называются *эмпирическими ортогональными функциями* (ЭОФ). Каждый сегмент длины 2τ с центром в t исходного процесса X является, по предположению, функцией из H , которая может быть разложена по базису ЭОФ:

$$X(t+s) = \sum_{k=0}^{\infty} X_k(t)\rho_k(s). \quad (26)$$

где $|s| \leq \tau$. Проекционный коэффициент $X_k(t)$, определяемый как

$$X_k(t) = \frac{1}{2\tau} \int_{-\tau}^{\tau} X(t+s)\rho_k(s)ds, \quad (27)$$

представляет собой процесс с нулевым средним, называемый k -й главной компонентой (ГК).

Из теоремы Мерсера в применении к оператору Гильберта-Шмидта A следует, что

$$C_X(t-s) = \sum_{k=1}^{\infty} \lambda_k \rho_k(s)\rho_k(t), \quad (28)$$

для почти всех t и s в $(-\tau, \tau)$, непосредственным следствием чего является равенство

$$\sum_{k=1}^{\infty} \lambda_k = C_X(0). \quad (29)$$

Рассмотрим подробнее связь между АСС и спектральным анализом. Для этого продолжим $\rho(t)$ на все \mathbb{R} , используя тождество (23). Это легко делается поскольку C_x интегрируемо по любому интервалу длины 2τ . Более того, если все λ_k положительны, то знание ρ_k в интервале $(-\tau, \tau)$ допускает такое продолжение с использованием (23). Если теперь предположить, что C_x лежит в $L^2(\mathbb{R})$, то каждая собственная функция также интегрируема с квадратом на \mathbb{R} . Это предположение облегчает рассмотрение сходимости, но не является необходимым, если (23) понимается в смысле распределений. Преобразование Фурье (24) дает

$$\lambda_k \hat{\rho}_k(f) = \frac{1}{2\tau} P_x(f) \tilde{\rho}_k(f), \quad (30)$$

где $\hat{\rho}_k(f)$ обозначает преобразование Фурье ρ_k :

$$\hat{\rho}_k(f) = \int_{-\infty}^{\infty} \rho_k(t) e^{2i\pi ft} dt. \quad (31)$$

По теореме Хинчина-Винера $P_x(f)$, преобразование Фурье $C_x(t)$, является спектральной плотностью $X(t)$, часто называемой также спектром мощности X ,

$$\tilde{\rho}_k(f) = \frac{1}{2\tau} \int_{-\tau}^{\tau} \rho_k(t) e^{2i\pi ft} dt \quad (32)$$

есть усеченное преобразование Фурье для ρ_k .

Для малых τ условие ортогональности на собственные функции ρ_k с необходимостью ведет к искажению их приближенно-синусоидальной формы вблизи нуля. Более того, наблюдается, что АСС способен разделять ρ_k в непересекающиеся группы, каждая из которых соответствует отдельному пику в спектре мощности. Способность АСС разделять вклады от различных частотных областей может быть объяснена следующим образом. Пусть спектр X состоит из двух непересекающихся областей. Тогда X – сумма двух независимых процессов, каждый из которых обладает набором частот из соответствующей области X . При $\tau \rightarrow \infty$ усредняющее действие $\hat{\rho}_k$, ведущее к $\tilde{\rho}_k$, становится очень малым. Поэтому собственные функции для каждого отдельного процесса хорошо приближаются решением общей задачи на собственные значения. Следовательно, собственные функции разделяются на два различных и приблизительно ортогональных множества.

На практике эффективное разделение происходит даже для малых окон и более сложного спектра. Это свойство разделения прямо следует из ортогональности собственных функций, т.е.

$$(\rho_i, \rho_j) = \frac{1}{2\tau} \int_{-\tau}^{\tau} \rho_i(t) \rho_j(t) dt = \delta_{ij}. \quad (33)$$

Ортогональность в (33) может иметь место в случаях, когда собственные функции находятся в противофазе, либо принадлежат различным частотным областям.

6 Ряды конечного ранга

Идея нахождения размерности подпространства, содержащего развертку временного ряда, упомянутая в [16, 17], получает развитие в работе Бухштабера [19], где рассматривается задача построения модели (аппроксимации) для временного ряда $f = (f_1, \dots, f_N)$ с $f_i = f((i-1)\Delta t)$, а функция $f(t)$ имеет вид

$$f(t) = \sum_{k=1}^K a_k(t) e^{\lambda_k t} \sin(\omega_k t + \phi_k). \quad (34)$$

В качестве подхода к решению этой задачи также используется метод многомерной развертки временного ряда. В указанной работе под *n-мерной разверткой* временного ряда $f = (f_1, \dots, f_N)$ подразумевается кусочно-линейная кривая \mathbf{X}_f , соединяющая последовательно вектора $X_j \in \mathbb{R}^n$, $j = 2, \dots, p = N - n + 1$, определяемые сегментами длины n исходного временного ряда $X'_j = (f_j, \dots, f_{j+n-1})$. Изучая геометрию развертки временных рядов, автор работы [19] вводит следующее основное определение: *n-ранг временного ряда f не превышает r* ($r < n$), если существует *r*-мерная гиперплоскость $L \subset \mathbb{R}^n$, такая что $\mathbf{X}_f \subset L$. В дальнейшем доказывается, что для временного ряда $f = (f_1, \dots, f_N)$, ранг которого не превосходит r , справедливы следующие утверждения.

1. Пусть $N_1 > N + n$. Ряд $f = (f_1, \dots, f_N)$ может быть дополнен до ряда $\tilde{f} = (f_1, \dots, f_N, \tilde{f}_{N+1}, \dots, \tilde{f}_{N_1})$, чей ранг не превышает $r < n$, если и только если существуют такие константы c_0, \dots, c_l для некоторого $l \leq r$, что

$$f_m = \sum_{s=1}^l c_{l-s+1} f_{m-s}$$

для всех $m = l + 1, \dots, N$.

2. Пусть $f = (f_1, \dots, f_N)$ — ряд, где $f_i = f((i-1)\Delta t)$ для функции f вида (34). *n*-Ранг f не превосходит r для всех $n > r$, N и Δt тогда и только тогда, когда $f(t)$ является решением обыкновенного линейного дифференциального уравнения

$$\sum_{q=0}^r b_q \frac{d^q}{dt^q} f(t) = c, \quad (35)$$

с постоянными коэффициентами b_0, \dots, b_r, c .

Пусть теперь L есть *r*-мерная гиперплоскость в \mathbb{R}^n . Для заданного ряда f обозначим через $\mathbf{X}_f(L)$ ортогональную проекцию развертки \mathbf{X}_f на L . Ясно, что $\mathbf{X}_f(L)$ является кусочно-линейной кривой в \mathbb{R}^n с узлами $(X_1(L), \dots, X_p(L))$. Обозначим

$$\rho_n(f, L) = \|\mathbf{X}_f - \mathbf{X}_f(L)\|^2 = \frac{1}{p} \sum_{q=1}^p \|X_q - X_q(L)\|^2,$$

где $\|\cdot\|$ — стандартная Евклидова метрика в \mathbb{R}^n . Рассмотрим множество $G'(r, n)$ всех *r*-мерных гиперплоскостей в \mathbb{R}^n . Используя *n*-мерную развертку \mathbf{X}_f , получаем функцию

$$\rho_n(f) : G'(r, n) \rightarrow \mathbb{R}^l, \quad \rho_n(f)(L) = \rho_n(f, L).$$

Обозначим

$$r_n(f) = \inf_{L \in G'(r,n)} \rho_n(f, L).$$

В указанной работе для вычисления $r_n(f)$ и описания набора $\mathcal{G}_n(f, r)$ всех r -мерных гиперплоскостей, на которых функция $\rho_n(f)$ принимает значение в точности $r_n(f)$, фактически используется метод главных компонент. Описан алгоритм, который по произвольной кусочно-линейной кривой $\mathbf{X}_f(L)$ строит ряд $f(L, r)$, и доказано, что для временного ряда $f_*(r) = f(L_*, r)$, $L_* \in \mathcal{G}_n(f, r)$ справедливо следующее свойство: если $g(t)$ - решение уравнения (35) с заменой f на g , то

$$\|f - g\|_\mu^2 \geq \|f - f_*(r)\|_\mu^2,$$

где μ — метрика в пространстве временных рядов длины N , индуцированная стандартной метрикой $\|\cdot\|$ в пространстве развертки. Т.о., $\|f - f_*(r)\|_\mu^2 = \delta_r$ дает нижнюю границу качества аппроксимации ряда f в классе функций вида (34), удовлетворяющих уравнению (35) (это достигается без явной оценки параметров модели (34)).

7 О многомерных модификациях АСС

Упомянутые ранее методы предназначены для обработки исключительно одномерных временных рядов. Обобщению техники развертки временных рядов на многомерный случай посвящена работа Ваутарда [62]. Описанный в этой работе алгоритм в основном соответствует одной из схем многомерной "Гусеницы" (см. раздел III.2). Однако, следует подчеркнуть некоторое отличие упомянутых методов от "Гусеницы". В то время как в "Гусенице" используется разложение выборки по собственным векторам "гусеничной" ковариационной либо корреляционной матрицы, АСС методы, особенно в последнее время, используют разложение по собственным векторам выборочной автоковариационной матрицы исходного ряда. Считается, что такая модификация улучшает чувствительность метода при анализе квазипериодических сигналов. "Гусеничный" алгоритм идейно ближе к оригинальным работам Broomhead и др. [16]. Данное замечание относится как к одномерным, так и к многомерным модификациям метода.

8 Динамический анализ Фурье

Исторически, ряды Фурье – инструмент, позволяющий увидеть структуру определённого класса функций и состоящий в разложении исходной (случайной) функции в ряд по фиксированной системе базисных функций. Разложение по $\sin - \cos$ жестко порождает свойство строгой периодичности. В этом смысле метод статичен; также он задает взаимно-однозначное соответствие: одномерный ряд значений функции времени преобразуется в одномерный же ряд значений функции частоты. Большие проблемы возникают с системами, свойства которых (спектр) изменяются во времени. Количество задач такого типа постоянно

увеличивается: медицина (электро-, кардио- и энцефалография), геофизика, радиофизика, экономика (см., например, [50])

Были и продолжаются попытки построить так называемый "динамический анализ Фурье" ("динамический спектральный анализ", "спектрально-временное картирование" [37, 58] и т.п.). Простейший вариант такой процедуры состоит в следующем. Весь интервал наблюдения изучаемой функции $f(t)$, $t \in [0, T]$, разбивается на N равных частей длины $t_N = \frac{T}{N}$, и затем на каждом из подинтервалов строится спектр Фурье $\varphi_i(\omega)$. Получается как-бы двумерное спектрально-временное представление процесса $f(t)$. Однако проблемы, возникающие здесь, аналогичны проблемам при построении классических гистограмм в статистике. Изменение N и смещение границ интервалов могут весьма существенно изменить полученную картину. При этом, если для детерминированной функции $f(t)$ ещё можно говорить о какой-то параметризации представления, то для случайной функции (реализации случайного процесса) ситуация только ухудшается. Это связано с тем, что преобразование Фурье является взаимно-однозначным преобразованием. Вся случайность, которая содержится в гладкой функции, переходит в спектр. Можно, как это обычно делается, для уменьшения случайности перейти к спектральной плотности и использовать спектральные окна. Здесь, однако, возникает проблема с интерпретацией спектров в случае детерминированной функции $f(t)$. А ведь во многих случаях трудно заранее определить долю случайности в исследуемом процессе.

Вспоминая, что в действительности мы всегда имеем дело не с непрерывными функциями, а с дискретными, т.е. с временными рядами, мы можем использовать процедуру, похожую на "Гусеницу". Пожертвовав трудоёмкостью, можно строить спектры не по последовательным интервалам длины t_N , а чаще, смещающая такой интервал на одну или несколько точек, т.е. с шагом дискретизации или кратным ему. Понятно, что спектры по перекрывающимся отрезкам функции будут, как правило, не слишком отличаться друг от друга, и мы получаем как бы "динамические спектры". Понять реальную природу этих спектров помогает спектральная теория линейной фильтрации. Рассмотрим вид получаемой k -ой гармоники "динамического спектра" (для простоты возьмём только косинусную компоненту)

$$c_k(\tau) = \int_{\tau}^{\tau+t_N} f(t) \cos \left[\frac{2\pi k}{t_N}(t - \tau) \right] dt, \quad \tau \in [0, T - \tau].$$

Очевидно, мы имеем свёртку исходной функции с волновым пакетом с гармоническим заполнением частоты $\frac{2\pi k}{t_N} = \omega_k$ и прямоугольной огибающей длины t_N . Эту операцию можно трактовать как фильтрацию исходной функции линейным фильтром с частотной характеристикой

$$F(\omega) = \int_0^{t_N} \exp\{i\omega t\} \cos \left(\frac{2\pi k}{t_N} t \right) dt = \int_0^{t_N} \exp\{-i\omega t\} \operatorname{Re} \exp\{i\frac{2\pi k}{t_N} t\} dt =$$

$$\operatorname{Re} \int_0^{t_N} \exp\{-i \left(\omega - \frac{2\pi k}{t_N} \right) t\} dt = \left| \frac{\cos(\omega - 2\pi k/t_N)}{(\omega - 2\pi k/t_N)} \right|$$

Фактически, вместо оценки коэффициента при косинусной составляющей частоты ω_k мы получили результат линейной фильтрации исходной функции $f(t)$

фильтром, имеющим достаточно широкую неравномерную полосу пропускания и лишь максимум на частоте ω_k . При этом для разных гармоник частотные характеристики фильтров имеют разную ширину, так что говорить о динамических спектрах можно лишь с очень большой натяжкой. Можно несколько улучшить огибающую волновых пакетов, "сгладив" прямоугольник, используя в качестве "окна", скажем, гауссовскую плотность [29]. Это, как правило, приводит к сужению полосы фильтров. Рассмотрим огибающую с фиксированным параметром масштаба σ^2 и переменным параметром сдвига b . Вид k -й гармоники спектра Фурье в этом случае

$$c_k(b, \sigma^2) = \int f(t) \exp\{i\omega_k t\} \exp\left\{-\frac{(t-b)^2}{2\sigma^2}\right\} dt.$$

Частотная характеристика такого фильтра существенно лучше:

$$\begin{aligned} F_k(\omega; b, \sigma^2) &= \int \exp\{i\omega t\} \exp\left\{i\omega_k t - \frac{(t-b)^2}{2\sigma^2}\right\} dt = \\ &= \sqrt{(2\pi\sigma^2)} \exp\left\{-\frac{(\omega - \omega_k)^2\sigma^4}{2}\right\}. \end{aligned}$$

Отсюда довольно естественно перейти к идеи вейвлет-анализа, который, однако, был сформулирован Гроссманом и Морле только почти через 40 лет [36].

9 Вейвлет-анализ и "Гусеница"

Вейвлет-преобразование одномерного сигнала состоит в его разложении по базису, сконструированному из обладающей определёнными свойствами солитоноподобной функции (вейвлета) посредством масштабных изменений и переносов. Каждая из функций этого базиса характеризует как определённую пространственную (временную) частоту, так и её локализацию в физическом пространстве (времени).

Следуя [2, 65] напомним элементы теории рядов Фурье и сопоставим их с основами вейвлет-анализа. Ряды Фурье определяются для функций, являющихся элементами $L^2(0, 2\pi)$ — кусочно непрерывных, 2π -периодических функций с

$$\int_0^{2\pi} |f(t)|^2 dt < \infty.$$

Любая такая функция допускает единственное представление

$$f(t) = \sum_{-\infty}^{\infty} c_k \exp(ikt),$$

где

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(t) \exp(-ikt) dt.$$

Можно сказать, что

$$\omega_k(t) = \exp(ikt), \quad k = \dots, -1, 0, 1, \dots,$$

является ортонормированным базисом $L^2(0, 2\pi)$, построенным с помощью масштабного преобразования единственной функции $\omega(t) = \exp(it)$ таким образом, что $\omega_k(t) = \omega(kt)$.

Получаемые в результате разложения Фурье коэффициенты c_k поддаются достаточно простой физической интерпретации, причём простота ни в коем случае не умаляет важности последующих выводов о характере исследуемого сигнала. Итак, каждая 2π -периодическая квадратично интегрируемая функция может быть получена суперпозицией масштабных преобразований базисной функции $\omega(t) = \exp(it) = \cos t + i \sin t$, т.е. является композицией синусоидальных волн с частотами, кратными базисной.

Если же мы априори знаем, что исследуемая функция не является строго 2π -периодической, то необходимо перейти к пространству $L^2(R)$. Но в этом случае очевидно, что синусоидальные волны $\omega_k(t)$ не принадлежат $L^2(R)$. Функции, образующие пространство $L^2(R)$, должны стремиться к нулю на $\pm\infty$ и для практических целей чем быстрее, тем лучше. Ортогональный базис этого пространства поэтому имеет более сложную структуру. Одних масштабных преобразований некоторой базисной функции для покрытия всей оси $R(-\infty, \infty)$ недостаточно. Оказывается возможным построить ортонормированный базис с помощью двухпараметрического семейства базисных функций, образованных масштабными и сдвиговыми преобразованиями одной базисной функции $\psi(t)$. В частности, можно выбрать функцию $\psi(t)$ (и не единственным образом) так, чтобы получилась ортонормированная система посредством масштабных преобразований $(1/2^j)$ и сдвигов $(k/2^j)$, где $k = \dots - 1, 0, 1, \dots$, $j = 1, 2, \dots$. Так, получается семейство функций вейвлет-базиса

$$\psi_{jk}(t) = 2^{j/2} \psi(2^j t - k).$$

Обратим внимание, что в случае вейвлет-преобразования правильнее говорить о параметрах сдвига и масштаба, а не о частотах. Для того, чтобы масштабный параметр мог интерпретироваться как частота, необходимо, чтобы исходная функция $\psi(t)$ обладала необходимыми спектральными свойствами, т.е. чтобы её преобразование Фурье можно было рассматривать как "спектральную линию".

Таким образом, вейвлет-преобразование исходной функции $f(t)$ имеет вид

$$f(t) = \sum_{j,k} c_{jk} \psi_{jk}(t),$$

где $c_{jk} = c_\psi \int_{-\infty}^{\infty} f(t) \psi_{jk}^*(t) dt$, звёздочка обозначает комплексное сопряжение. Можно записать непрерывный аналог этого представления, аналогично интегральному Фурье-преобразованию, как функцию двух переменных: масштабного коэффициента и параметра сдвига b :

$$[W_\psi f](a, b) = \int_{-\infty}^{\infty} f(t) \psi_{ab}^*(t) dt = |a|^{-1/2} \int_{-\infty}^{\infty} f(t) \psi^*((t-b)/a) dt.$$

Тогда выполняется соотношение

$$c_{jk} = [W_\psi f] \left(\frac{1}{2^j}, \frac{k}{2^j} \right).$$

Обратим внимание на интерпретацию двухпараметрического набора коэффициентов c_{jk} . При фиксированном k этот набор можно рассматривать аналогично коэффициентам Фурье как амплитуды соответствующих вейвлет-гармоник в момент времени, определяемый параметром k . При фиксированном j набор фактически оказывается дискретной свёрткой исходной функции $f(t)$ и j -го базисного вейвлета, т.е. может рассматриваться как результат фильтрации функции $f(t)$ линейным фильтром с функцией отклика, равной j -ому базисному вейвлету. Если спектр Фурье этой базисной функции локализован около частоты ω_j , то можно говорить об узкополосной фильтрации и в этом случае весь набор коэффициентов c_{jk} представляет собой дискретное спектрально–временное представление функции $f(t)$.

Сделаем ещё несколько замечаний относительно частотно-временного представления сигналов. Многие реальные процессы в природе, обществе и технике являются сложными регуляторными системами, измеряемые характеристики которых могут быть представлены функциями, являющимися суммами периодических компонент с эволюционирующими во времени частотой и амплитудой. При этом характеристики более частотных компонент, как правило, меняются более быстро, чем низкочастотных. Поэтому для получения высокочастотной информации с хорошей точностью важно извлекать её из относительно малых временных интервалов, а не из всего сигнала; и наоборот – низкочастотную спектральную информацию извлекать из относительно широких временных интервалов сигнала. Фурье-преобразование, даже в его модификациях типа "динамического Фурье-преобразования", плохо работает в таких ситуациях, в то время как вейвлет-преобразование автоматически обладает подвижным частотно-временным окном, узким на малых масштабах и широким на больших.

Динамическое Фурье-преобразование также может быть записано как функция двух переменных: частоты ω и временного сдвига b :

$$F(\omega, b) = \int f(t)z(t-b) \exp(i\omega t) dt,$$

где $z(t)$ — временное окно, например, прямоугольной или гауссовой формы.

Таким образом, $F(\omega, b)$ есть разложение сигнала $f(t)$ по семейству функций $z(t-b) \exp(i\omega t)$, образованному из единственной функции $z(t)$ с помощью переносов b во времени и переносов ω по частоте, в то время как результат вейвлет-преобразования $\omega(a, b)$ есть разложение сигнала по семейству $\psi((t-b)/a)$, образованному из единственной функции $\psi(t)$ с помощью переносов b во времени и растяжений a также во времени. Вейвлет-преобразование представляет собой как бы непрерывный блок оконных преобразований Фурье с различными окнами для каждой частоты.

Базисные функции оконного преобразования Фурье имеют одно и то же разрешение по времени и частоте для всех точек плоскости преобразования, в то время как базисные функции вейвлет-преобразования имеют уменьшающееся с масштабом a разрешение по времени и увеличивающееся с масштабом a разрешение по частоте. Это свойство вейвлет-преобразования даёт ему большое преимущество при анализе сигналов, так как быстрые вариации сигналов (высокочастотные характеристики) хорошо локализованы, а для выявления медленно меняющихся характеристик достаточно хорошего низкочастотного разрешения. Вейвлет-преобразование, обладающее подвижным частотно-временным окном,

одинаково хорошо выявляет и низкочастотные, и высокочастотные характеристики сигналов.

Гусеничное преобразование можно точно также представить в двухпараметрической форме, если рассматривать матрицу главных компонент ковариационной матрицы совместно с вектором собственных чисел этой матрицы. Роль набора базисных функций выполняет набор собственных векторов (ортонормированный).

Отметим следующие отличия "Гусеницы" от вейвлет-преобразования:

1. Набор базисных функций порождается самой исследуемой функцией $f(t)$.
2. Имеются параметр (длина гусеницы) и варианты метода (одно или двухкратное центрирование и нормирование).
3. Возможность управляемого восстановления сигнала по интерпретируемым компонентам, в отличие от практически однозначных однократных Фурье- и вейвлет-преобразований.
4. Важное отличие "Гусеницы" от вейвлета состоит также в способности "Гусеницы" осуществлять многовариантный прогноз, согласованный с восстановлением. Причём это свойство "заложено" в структуре метода и даже в его предпосылках, в то время как вейвлет из-за априорной локальности набора базисных функций практически не приспособлен к построению прогноза.
5. Параметры ω в Фурье и a в вейвлете в численных реализациях дискретны и фиксированы, что затрудняет точную оценку положения частотных пиков, в то время как гусеница позволяет при наличии гармонических компонент в исследуемой функции $f(t)$ по практически всегда появляющимся парам собственных векторов или соответствующих главных компонент максимально точно оценить не только частоту и амплитуду, но и фазу этой составляющей. Более того, можно даже получить графики мгновенной частоты, амплитуды и фазы этой компоненты.
6. Вейвлет-преобразование для реальных временных рядов имеет чётко выраженный граничный эффект по параметру сдвига, чего нет у "Гусеницы".

"Гусеница" разделяет процедуру двумерного разложения на два этапа. На первом этапе сигнал переводится в двумерное представление с практически равноправными координатами, на втором производится сингулярное разложение этого двумерного представления по двум системам ортогональных функций. Интерпретация набора этих функций не является жёстко фиксированной, как в вейвлет-анализе, а определяется параметром разложения (длиной гусеницы) и особенностями самого исходного сигнала. В некоторых случаях (в длинных сигналах, $M \ll N$), если отсутствуют или исключены медленные тренды, получается разложение, достаточно схожее с вейвлет-преобразованием, а для строгого периодических сигналов и при соответствующем выборе параметров N и M "Гусеница" фактически осуществляет разложение в ряд Фурье.

Каждая отдельная гусеничная собственная функция, рассматриваемая как линейный фильтр, обладает при этом не очень "хорошей" в смысле Фурье характеристикой. Но это только подтверждает наше представление о том, что "Гусеница" даёт более широкое, чем спектральное, представление исследуемых сигналов.

Отметим также возможность взаимодействия "Гусеницы" с динамическим (оконным) Фурье и вейвлет-преобразованиями. Можно исследовать сечения $F(\omega, b)$ при фиксированном ω или $\omega(a, b)$ при фиксированном a как функции сдвига b с помощью гусеницы.

10 Практические применения

Методы типа "Гусеница" широко использовались при обработке экспериментальных данных самого различного характера. Одними из первых публикаций на эту тему являются статьи Кислицина [6, 7, 8]. Большая часть публикаций, так или иначе связанных с применением Анализа Сингулярного Спектра, выполнена специалистами отделения Изучения Атмосферы Калифорнийского университета. Общее число работ превышает два десятка. Отметим лишь следующие: работу Fraedrich [28] об оценке размерности наблюдаемых атмосферных атTRACTоров, цикл статей Ghil и др.[32, 33, 34] о выделении динамики из атмосферных и температурных данных. Интересными представляются также работы Kerrenne и Ghil [42, 43], где впервые приводятся обработки многомерных временных рядов возникающих при изучении некоторых атмосферных явлений. Из отечественных работ интерес представляют статьи Антоновского и др. [12, 1], где используется метод, алгоритмически наиболее близкий к "Гусенице". Одними из последних прикладных работ по физике атмосферы являются статьи [66], [67]. В числе последних отечественных прикладных работ, использующих методы типа "Гусеница", следует отметить статьи Пичугина [9, 10]. Отдельно можно упомянуть недавние работы [35], [49], где делается попытка анализа практических свойств АСС, особенно выявления его сильных и слабых сторон. Описание некоторых видов программного обеспечения, реализующего АСС, может быть найдено в [24] и [62].

Литература

- [1] Антоновский М.Я., Бухштабер В.М., Векслер Л.С. Применение многомерного статистического анализа для обнаружения структурных изменений во временных рядах данных экологических наблюдений. Проблемы экологического мониторинга и моделирования экосистем. Т. XV. — СПб: Гидрометеоиздат. 1993. — с. 193–213.
- [2] Астафьева Н.М. Вейвлет-анализ: основы теории и примеры применения. УФН, т. 166, 11, 1966. — с. 1145–1170.
- [3] Белонин М.Д., Голубева В.А., Скублов Г.Т., (1982), *Факторный анализ в геологии*. М., Недра.

- [4] Белонин М.Д., Татаринов И.В., Калинин О.М., Шиманский В.К., Бескровная О.В., Гранский В.В., Похитонова Т.Е., (1971), *Факторный анализ в нефтяной геологии*. М., ВИЭМС.
- [5] Бриллинджер, Д., (1980), *Временные ряды. Обработка данных и теория*, Москва, Мир.
- [6] Кислицин, М.М., (1978), Биотехнические системы в авиационной эргономике. *Вопросы кибернетики*. **51**, М.
- [7] Кислицин, М.М., (1978), Исследование функционального состояния пилотов в процессе летного труда методом многомерной статистики. *Вопросы кибернетики*. **51**, М.
- [8] Кислицин, М.М., (1978), Многомерная статистика временных рядов наблюдений в авиационной эргономике. *Вопросы кибернетики*. **51**, М.
- [9] Пичугин, Ю.А., (1996), Учет сезонных эффектов в задачах прогноза и контроля данных о приземной температуре воздуха. *Метеорология и гидрология*, 4, 52-64.
- [10] Пичугин, Ю.А., (1996), Использование ковариационной и корреляционной матриц при расчете главных компонентов в задаче учета сезонных эффектов при прогнозе и контроле данных приземной температуры воздуха. *Метеорология и гидрология*, 8, 17-26.
- [11] Rao, C.P., (1968), *Линейные статистические методы и их применения*. М., Наука.
- [12] Antonovskii, M.Ya., Buchstaber, V.M., Veksler, L.S., Application of multivariate statistical analysis for the detection of structural changes in the series of monitoring data. *Working paper WP-91-31*, IIASA, Laxenburg, 1991.
- [13] Colebrook, J.M., (1978), Continuous plankton records - zooplankton and environment, northeast Atlantic and North Sea, 1948-1975., *Oceanol. Acta* **1**, 9-23.
- [14] Box, G.E.P., Jenkins, G.M., (1970), *Time Series Analysis. Forecasting and Control*, Holden Day, San Francisco.
- [15] Brockwell, P., Davis, R., (1987), *Time Series: Theory and Methods*. Springer-Verlag, New York.
- [16] Broomhead, D.S., King, G.P., (1986), Extracting qualitative dynamics from experimental data. *Physica D*, **20**, 217-236 (1986).
- [17] Broomhead, D.S., King, G.P., (1986), On the qualitative analysis of experimental dynamical systems, in *Nonlinear Phenomena and Chaos*, Ed. S. Sarkar, Adam Hilger, Bristol, 113-144.
- [18] Broomhead, D.S., Jones, R., King, G.P., (1987) Topological dimension and local coordinates from time series data. *J. Phys. A*, **20**, L563-L569.

- [19] Buchstaber, V.M., (1994), Time Series Analysis and Grassmannians, *Amer. Math. Soc. Transl.*, (2), **162**.
- [20] Burg, J.P., (1967), Maximum entropy spectral analysis, статья, представлена на 37 ежегодной конференции Soc. Explor. Geophys., Oklahoma, а также в *Modern Spectrum Analysis*, ed. D.G. Childers, IEEE Press, N.Y., (1978), 34.
- [21] Casdagli, M., (1989), Nonlinear prediction of chaotic time series. *Physica D*, **35**, 335-356.
- [22] Danilov, D.L., (1996), Principal Components in Time Series Forecast. *1996 Proceedings of the Statistical Computing Section of the American Statistical Association*, 156-160.
- [23] Danilov, D., Solnsev, V.N., Zhigljavsky, A., (1996) Analysis and forecast of time series on the base of the principal components. *Abstracts of Second International Conference on Computing in Economics and Finance Geneva*, Switzerland, 26-28 June 1996.
- [24] Dettinger, M.D., Ghil, M., Strong, C.M., Weibel, W., and Yiou, P., (1995), Software expedites singular-spectrum analysis of noisy time series, *Eos Trans. AGU* **76(2)** 12, 14, 21.
- [25] Eckmann, J.-P., Ruelle, D., (1985), Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, **57** 617-656.
- [26] Farmer, J.D., Sidorowich, J., (1987), Predicting Chaotic Time Series, *Phys. Rev. Lett.*, **59**, 8, 845-848.
- [27] Farmer, J.D., Sidorowich, J., (1988), Exploiting chaos to predict the future and reduce noise, in: *Evolution, Learning and Cognition*, Ed. Y.C. Lee World Scientific, Singapore.
- [28] Fraedrich, K., (1986), Estimating the dimension of weather and climate attractors, *J. Atmos. Sci.*, **43**, 419-432.
- [29] Gabor, D., (1946), *Theory of communication*. J.IEE, 93. – pp. 429–457.
- [30] Golub, G., Van Loan, C., (1989), *Matrix Computations*. Second edition, The Johns Hopkins University Press.
- [31] Geman, S., Bienenstock, E., Doursat, R., (1992), Neural networks and the bias/variance dilemma. *Neural Comp.*, **4**, 1-58.
- [32] Ghil, M., Mo, K., (1991), Intraseasonal oscillations in the global atmosphere-Part I: Northern Hemisphere and Tropics, *J. Atmos. Sci.*, **48**, 752-779.
- [33] Ghil, M., Mo, K., (1991), Intraseasonal oscillations in the global atmosphere-Part I: Southern Hemisphere, *J. Atmos. Sci.*, **48**, 780-790.
- [34] Ghil, M., Vautard, R., (1991), Interdecadal oscillations and the warming trend in global temperature time series, *Nature*, **350**, 324-327.

- [35] Ghil, M., Yiou, P., (1996), Spectral methods: What they can and cannot do for climatic time series, in *Decadal Climate Variability: Dynamics and Predictability*, Eds. D. Anderson and J. Willebrand, Elsevier, Amsterdam.
- [36] Grossman, A., Morlet, J., (1984), Decomposition of Hardy functions into square integrable wavelets of constant shape, *SIAM J. Math.*, **1**, 723–736.
- [37] Haberl, R., Jilge, G., Pulter, R., Steibeck G., (1989), Spectral mapping of the electrocardiogram with Fourier transform for identification of patients with sustained ventricular tachycardia and coronary artery disease // *Europ. Heart J.*, Vol. 10, 316–322.
- [38] Hotelling, H., (1933), Analysis of a complex statistical variables into principal components, *J. Educ. Psych.*, **24**, 417-441, 498-520.
- [39] Huebner, U., Abraham, N.B., Weiss, C.O., (1989), Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared NH₃ laser. *Phys. Rev. A*, **40** 6354.
- [40] Jenkins, G.M., Watts, D.G., *Spectral Analysis and its Applications*, Holden Day, San Francisco.
- [41] Kendall, M.G., Stuart, A., (1968), *The Advanced Theory of Statistics*, Griffin, London.
- [42] Keppenne, C.L., Ghil, M., (1992), Adaptive spectral analysis and prediction of the Southern Oscillation Index, *J. Geophys. Res.*, **97**, 20449-20554.
- [43] Keppenne, C.L., Ghil, M., (1993), Adaptive filtering and prediction of noisy multivariate signals: an application to atmospheric angular momentum, *Intl. J. Bifurcation and Chaos*, **3**, 625-634.
- [44] Kimoto, M., Ghill, M., Mo., K.-C., (1991), Spatial structure of the extratropical 40-day oscillation, b *Proc. 8th Conf. on Atmospheric and Oceanic Waves and Stability (American Meteorological Society, Boston)*, 17.
- [45] Kumaresan, R., Tufts, D.W., (1980), Data-adaptive principal component signal processing, b *IEEE Proc. Conf. on Decision and Control*, (Albuquerque, 1980), 949.
- [46] Liebert, W., Schuster, H.G., (1988), Proper Choice of the Time Delay for the Analysis of Chaotic Time Series, *Phys. Lett. A*, **142**, 107-111.
- [47] Loeve, M., *Probability theory*, 3rd ed., Van Nostrand, Princeton.
- [48] Packard, N., Crutchfield, J., Farmer, D., Shaw, R., (1980), Geometry from a time series, *Physical Review Letters*, **45**, 712.
- [49] Palus, M., Dvorak, I., (1992), Singular-value decomposition in attractor reconstruction: Pitfalls and precautions, *Physica D*, **55**, 221-234.
- [50] Parker, B., (1979), Fourier analysis of electrocardiograms // *PACE*, Vol. 2, 245–248.

- [51] Pike, E.R., McWhirter, J.G., Bertero, M., de Mol, C., (1984), *IEEE Proc.* **131**, 660.
- [52] Preisendorfer, R.W., (1988), Principal Components Analysis in Meteorology and Oceanography, ed. C.D. Mobley, Elsevier, Amsterdam.
- [53] Priestley, M., (1988), *Nonlinear and Nonstationary Time Series Analysis*, Academic Press.
- [54] Rao, C.R., (1964), The use and interpretation of principal component analysis in applied research. *SANKHYA The indian journal of statistics*, **A**, **26**(4).
- [55] Riesz, F., Sz.-Nagy, B., (1955), *Functional Analysis*, F.Ungar, N.Y.
- [56] Sauer, T., (1993), Time Series Prediction Using Delay Coordinate Embedding. in: *Time Series Prediction: Forecasting the Future and Understanding the Past*. Eds. A.S.Weigend, N.A.Gershenfeld, Addison-Wesley.
- [57] Sauer, T., Yorke, J.A., Casdagli, M., (1991), Embedology. *J. Stat. Phys.*, **65**, 579-616.
- [58] Smetnev, A.S., Kulambaev, B.B., Akasheva, D.Y., (1992), Late ventricular potentials comparative value of time domain analysis and spectro-temporal mapping. XXIII international congress in electrocardiology // Book of abstracts, 104.
- [59] Tong, H., (1990), *Nonlinear Time Series*, Clarendon Press, Oxford.
- [60] Takens, F., (1981), Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, **898**, Springer-Verlag.
- [61] Vautard, R., Ghil, M., (1989), Singular-spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D*, **35**, 395-424.
- [62] Vautard, R., Yiou, P., and Ghil, M., (1992), Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, *Physica D*, **58**, 95-126.
- [63] Vautard, R., Plaut, G., (1992), Low-frequency oscillations and weather regimes in the Northern Hemisphere, in *Proc. ECMWF Workshop on Predictability (European Centre for Medium Range Weather Forecasts)*. Reading.
- [64] Whitney, H., (1936), Differentiable manifolds, *Ann. Math.*, **37**, 645.
- [65] Wavelet Analysis and Its Applications, (1992), Vol. 1: An Introduction in Wavelets, Vol. 2: Wavelets: A Tutorial in Theory and Applications. San Diego: Academ. Press Inc.
- [66] Yiou, P., Ghil, M., Jouzel, J., Paillard, D., Vautard, R., (1994), Nonlinear variability of the climatic system, from singular and power spectra of Late Quaternary records, *Clim. Dyn.*, **9**, 371-389.
- [67] Yiou, P., Jouzel, J., Johnsen, J., Rognvaldsson O.E., (1995), Rapid oscillations in Vostok and GRIP ice cores, *Geophys. Res. Lett.*, **22(16)**, 2179-2182.

Глава III Развитие базового алгоритма

III.1 Метод “Гусеница” для прогнозирования временных рядов

Д.Л. Данилов

1 Введение

Целью настоящей статьи является развитие подхода к прогнозированию временных рядов при помощи метода “Гусеница”. Длительное время метод использовался в основном для обнаружения скрытых (т.е. трудно выявляемых другими методами) периодичностей. С появлением алгоритма реконструкции ряда появилась возможность применять метод для сглаживания временных рядов. При этом реализовалось естественное желание аналитика выделить из изучаемого временного ряда его информативную компоненту и отбросить шумы. Следующим шагом в этом направлении должно было стать появление возможности провести экстраполяцию (продолжение) информативной части изучаемого ряда. Необычность и сложность этой задачи состоит в том, что в самом методе “Гусеница” не содержится никакой явной информации о дальнейшем поведении ряда. Существенное продвижение здесь может быть сделано при рассмотрении алгоритма “Гусеница” с точки зрения регрессионного анализа. В разделе 3 приводится алгоритм, позволяющий прогнозировать (экстраполировать) функции дискретного аргумента из достаточно широкого класса. Найдены условия, при которых по определенному числу значений такой функции можно получать любые последующие. Применение алгоритма проиллюстрировано на численных примерах (раздел 4). Другой вариант сходного по свойствам алгоритма прогнозирования можно найти в разделе 5. В разделе 6 алгоритмы прогнозирования детерминированных функций обобщены на случай произвольных детерминированных рядов, а в разделе 11 — и на случайные ряды. Здесь прогнозируемое значение рассматривается как статистическая оценка тренда случайного ряда. Для такого алгоритма также приведены некоторые его модификации (раздел 13), описаны свойства и большое число примеров (раздел 12). Связь между предлагаемым алгоритмом прогнозирования и некоторыми классическими регрессионными задачами рассмотрена в разделе 8.

2 Схема метода “Гусеница” и основные определения.

Для начала введем несколько определений и опишем основные понятия. Не ставя целью воспроизвести здесь все тонкости схемы “Гусеница”, авторы тем не менее считают возможным повторить некоторые детали метода, важные для понимания описываемого алгоритма прогноза. Из методологических соображений алгоритмы прогноза будут излагаться в рамках схемы нецентральных главных компонент.

Пусть $(f_i)_{i=1}^N$ — числовая последовательность. Задав параметр τ , $1 < \tau < N$, преобразуем ее в многомерную выборку, т.е. в последовательность векторов $X^{(1)}, \dots, X^{(n)}$, $n = N - \tau + 1$, $X^{(k)} = (x_i^{(k)})_{i=1}^\tau$ следующим образом

$$\mathbf{X} = (X^{(1)}, \dots, X^{(n)}) = \begin{pmatrix} f_1 & f_2 & \dots & f_n \\ f_2 & f_3 & \dots & f_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ f_\tau & f_{\tau+1} & \dots & f_N \end{pmatrix}. \quad (1)$$

Определение 2.1 Матрицу (1) будем называть нецентрированной матрицей наблюдений, соответствующей последовательности f_i .

Рассмотрим матрицу вторых моментов полученной многомерной выборки

$$C = \frac{1}{n} \mathbf{X} \mathbf{X}^T. \quad (2)$$

Применение хорошо известного метода главных компонент позволяет приводить эту матрицу к диагональному виду, с одновременным нахождением соответствующего базиса, который составлен из собственных векторов этой матрицы.

Определение 2.2 Обозначим через

$$\mathbf{V}_X = (v^{(1)}, v^{(2)}, \dots, v^{(\tau)}) = \begin{pmatrix} v_1^{(1)} & v_1^{(2)} & \dots & v_1^{(\tau)} \\ v_2^{(1)} & v_2^{(2)} & \dots & v_2^{(\tau)} \\ \vdots & \vdots & \ddots & \vdots \\ v_\tau^{(1)} & v_\tau^{(2)} & \dots & v_\tau^{(\tau)} \end{pmatrix}$$

матрицу собственных векторов матрицы вторых моментов. При этом будем предполагать, что собственные векторы упорядочены по убыванию соответствующих собственных чисел, т.е. $\lambda_1 > \lambda_2 > \dots > \lambda_\tau \geq 0$.

Обычные для анализа главных компонент операции получения нецентральных главных компонент:

$$\mathbf{U}_X = \mathbf{V}_X^T \mathbf{X} = (U_1 : \dots : U_\tau)^T$$

и восстановления исходной выборки по первым r главным компонентам:

$$\tilde{\mathbf{X}} = (v^{(1)}, v^{(2)}, \dots, v^{(r)}) \begin{pmatrix} U_1^T \\ \vdots \\ U_r^T \end{pmatrix} = \sum_{i=1}^r v^{(i)} U_i^T$$

применимы без ограничений. После восстановления матрицы $\tilde{\mathbf{X}}$ исходная последовательность восстанавливается усреднением по побочным диагоналям $\tilde{\mathbf{X}}$:

$$\tilde{f}_s = \begin{cases} \frac{1}{s} \sum_{i=1}^s \tilde{x}_{i,s-i+1} & 1 \leq s \leq \tau, \\ \frac{1}{\tau} \sum_{i=1}^\tau \tilde{x}_{i,s-i+1} & \tau \leq s \leq n, \\ \frac{1}{N-s+1} \sum_{i=1}^{N-s+1} \tilde{x}_{i+s-n,n-i+1} & n \leq s \leq N. \end{cases} \quad (3)$$

С геометрической точки зрения операция получения главных компонент есть изображение исходной выборки в базисе, составленном из выбранных собственных векторов, а восстановления – проектирование исходной выборки на гиперплоскость, порожденную выбранным набором собственных векторов матрицы вторых моментов. Процедуру восстановления исходной последовательности (3) иногда называют ганкелизацией матрицы $\tilde{\mathbf{X}}$.

В этом и последующем разделах нами будет уделено особое внимание функциям дискретного аргумента, порождающим многомерные выборки, лежащие в гиперплоскостях, размерности меньшей чем τ . В работе [15] показано, что к таким функциям относятся дискретизации решений обыкновенных однородных дифференциальных уравнений с постоянными коэффициентами т.е. $(f_i)_{i=1}^N$, где $f_i = f((i-1)\Delta t)$, $i = 1, \dots, N$, и

$$f(t) = \sum_{k=1}^m a_k(t) e^{\lambda_k t} \sin(\omega_k t + \varphi_k) \quad (4)$$

(здесь $a_k(t)$ — полиномы, а значения параметров λ_k , ω_k и φ_k произвольны). Отметим, что этот класс функций дискретного аргумента совпадает с множеством решений конечно-разностных уравнений конечного порядка с постоянными коэффициентами.

Замечание 2.1 Везде далее под конечно-разностным уравнением (к.р.у.) будет подразумеваться однородное конечно-разностное уравнение с постоянными коэффициентами, т.е. рекуррентное соотношение вида:

$$f_{r+k} + a_1 f_{r+k-1} + \dots + a_{r-1} f_{k+1} + a_r f_k = 0, \quad (5)$$

где a_i — вещественные постоянные и $a_r \neq 0$. Порядок записанного выше к.р.у. равен r . Рассмотрим числовой ряд $(f_i)_{i=1}^N$ являющийся решением конечно-разностного уравнения (5). Далее мы везде будем считать, что соответствующее к.р.у. *неприводимо*, т.е. невозможно записать уравнение меньшего порядка, решением которого являлся бы рассматриваемый ряд $(f_i)_{i=1}^N$. Более того, рассматривая некоторое решение f_i уравнения (5), мы всегда, если не оговорено обратное, будем полагать, что не существует к.р.у. меньшего порядка, такого, что f_i являлось бы его решением.

Введем следующие определения:

Определение 2.3 Если многомерная выборка, порожденная числовым рядом $(f_i)_{i=1}^N$, содержится в r -мерной гиперплоскости, и размерность r — одна и та же для любого τ ($r < \tau$), то ряд $(f_i)_{i=1}^N$ будет называться рядом ранга r . Заметим, что к.р.у. r -го порядка порождает ряд ранга r .

Определение 2.4 Числовой ряд $(f_i)_{i=1}^{N+1}$ называется продолжением ряда $(f_i)_{i=1}^N$, если порождаемая им при гусеничной обработке выборка лежит в той же гиперплоскости, что и у исходного ряда.

3 Детерминированный подход. Классическая схема.

Опишем идею предлагаемого алгоритма экстраполяции функции дискретного аргумента.

Пусть имеется числовой ряд $(f_i)_{i=1}^N$ ранга r . Рассмотрим базис гиперплоскости $v^{(1)}, \dots, v^{(r)}$, содержащей многомерную выборку, порожденную этим рядом. Не ограничивая общности можно считать, что в качестве этого базиса выбрана система собственных векторов, соответствующих отличным от нуля собственным числам матрицы вторых моментов многомерной выборки. Рассмотрим уравнение этой гиперплоскости, заданное в параметрической форме:

$$x(\bar{t}) = \sum_{i=1}^r t_i v^{(i)}, \quad (6)$$

при этом $\bar{t} = (t_1, \dots, t_r)$ — значение параметризации, соответствующее точке $x(\bar{t})$. Каждой точке многомерной выборки соответствует некоторое значение вектора параметризации \bar{t} , т.е.

$$X^{(1)} = x(\bar{t}^{(1)}), \dots, X^{(n)} = x(\bar{t}^{(n)}).$$

Пусть числовой ряд $(f_i)_{i=1}^{N+1}$ — продолжение ряда $(f_i)_{i=1}^N$, и пусть система базисных векторов $v^{(i)}$, вычисленных по ряду $(f_i)_{i=1}^N$ фиксирована. В этом случае последней $(n+1)$ -й точке выборки также соответствует некоторое значение вектора параметризации $\bar{t}^{(n+1)}$. Если окажется возможным восстановить значение $\bar{t}^{(n+1)}$ только по набору f_1, \dots, f_N , то все компоненты точки $X^{(n+1)}$ легко могут быть вычислены как:

$$X^{(n+1)} = \sum_{i=1}^r \bar{t}_i^{(n+1)} v^{(i)},$$

откуда нетрудно перейти к исходному ряду.

Указанная идея лежит в основе следующего алгоритма экстраполяции функции дискретного аргумента.

Алгоритм 3.1 Пусть имеется набор последовательных значений функции дискретного аргумента, образующий ряд $(f_i)_{i=1}^N$ ранга r , и количество элементов ряда N достаточно велико. (Заметим, что для прогнозирования ряда ранга r по гусенице длины τ достаточно взять $N \geq r + \tau$.)

1. В соответствии с (1) сформируем многомерную матрицу наблюдений.
2. Проведем построение матрицы вторых моментов (2).
3. Найдем базис $v^{(1)}, \dots, v^{(r)}$, соответствующий отличным от нуля собственным числам матрицы вторых моментов .
4. Запишем систему линейных алгебраических уравнений следующего вида:

$$\begin{cases} \sum_{j=1}^r h_j v_1^{(j)} &= f_{N-\tau+2} \\ &\dots \\ \sum_{j=1}^r h_j v_{\tau-1}^{(j)} &= f_N . \end{cases} \quad (7)$$

5. Исследуем систему (7) на совместность.

5.1 Если система (7) несовместна, то исходный ряд не допускает продолжения.

5.2 Если система (7) имеет решение, то исходный ряд имеет продолжение, которое записывается как:

$$x^* = \sum_{j=1}^r h_j^* v_\tau^{(j)},$$

где h_1^*, \dots, h_r^* — решение системы (7).

Замечание 3.1 Случай бесконечного множества решений системы (7) не реализуется ни для какого ряда конечного ранга.

Проведем обоснование приведенного алгоритма прогноза. Одним из первых вопросов, которые будут для этого рассмотрены, является вопрос о единственности прогнозирования ряда. Справедлива следующая теорема:

Теорема 3.1 Пусть ряд $(f_i)_{i=1}^N$ имеет ранг r , и $(f_i)_{i=1}^{N+1}$ — его продолжение. Пусть также $\tau > r - 1$. Тогда, если система (7) имеет решение, то оно единствено.

Доказательство

Проведем доказательство методом “от противного”. В предположении неединственности решения системы (7) докажем, что ранг исходного ряда не может быть больше $r - 1$.

Рассмотрим систему (7) и соответствующую ей однородную систему:

$$\begin{cases} \sum_{j=1}^r h_j v_1^{(j)} = 0 \\ \dots \\ \sum_{j=1}^r h_j v_{\tau-1}^{(j)} = 0. \end{cases} \quad (8)$$

Пусть система (7) совместна. Эта система имеет единственное решение тогда и только тогда, когда (8) имеет только тривиальные решения. Если решение (7) неединственно, то существуют такие h_1, \dots, h_r , $\sum_{k=1}^r h_k^2 \neq 0$ что: $\sum_{j=1}^r h_j \hat{v}^{(j)} = 0$

(через $\hat{v}^{(j)}$ обозначены векторы, составленные из первых $\tau - 1$ компонент векторов $v^{(j)}$). Поэтому среди $\hat{v}^{(j)}$ имеется не более чем $r - 1$ линейно независимый вектор. Следовательно, каждый столбец многомерной выборки

$$\begin{pmatrix} f_1 & f_2 & \dots & f_{N-\tau+2} \\ f_2 & f_3 & \dots & f_{N-\tau+3} \\ \vdots & \vdots & \ddots & \vdots \\ f_{\tau-1} & f_{\tau+1} & \dots & f_N \end{pmatrix}$$

является линейной комбинацией не более $r - 1$ вектора, и поэтому ранг ряда f_i не может превосходить $r - 1$. Т.о. решение (8) — тривиально и решение (7) единствено.

■

Доказанная теорема дает нам средство исследования свойств продолжимости детерминированных рядов. В частности, очевидным следствием предыдущей теоремы является теорема единственности продолжения ряда.

Теорема 3.2 Пусть ряд f_1, \dots, f_N допускает продолжение f_{N+1} . Тогда это продолжение единственно.

Доказательство

Действительно, пусть это не так, и существует $\hat{f}_{N+1} \neq f_{N+1}$ которое также является продолжением ряда. Рассмотрим систему (7). Она, очевидно, имеет решение (соответствующее f_{N+1}). Однако по Теореме 3.1 это решение должно быть единствено и поэтому $\hat{f}_{N+1} = f_{N+1}$. ■

Свойство единственности продолжения ряда может быть обобщено на случай многошагового продолжения ряда. Рассмотрим соответствующее обобщение Алгоритма 3.1.

Алгоритм 3.2 Пункты 1-4 совпадают с соответствующими пунктами Алгоритма 3.1.

Для определения последовательных значений f_{N+1}, \dots, f_{N+k} имеем последовательность систем

$$\begin{cases} \sum_{j=1}^r h_j v_1^{(j)} &= f_{N-\tau+k+1} \\ \dots \\ \sum_{j=1}^r h_j v_{\tau-1}^{(j)} &= f_{N+k-1}. \end{cases} \quad (9)$$

5. Исследуем систему (9), соответствующую $k = 1$, на совместность.

5.1 Если система (9), соответствующая $k = 1$, несовместна, то исходный ряд не допускает продолжения.

5.2 Если система (9), соответствующая $k = 1$ имеет решение, то она имеет решение при любом k , и продолжение исходного ряда однозначно определяется по формулам:

$$x^{(k)} = \sum_{j=1}^r h_j^*(k) v_n^{(j)},$$

где $h_1^*(k), \dots, h_r^*(k)$ — решение системы (9), соответствующее k -му шагу алгоритма.

Теорема 3.3 Пусть ряд $(f_i)_{i=1}^N$ допускает продолжение f_{N+1} . Тогда он может быть продолжен неограниченно на любое число шагов, причем это продолжение единственно.

Доказательство

Достаточно доказать, что существует и единственno f_{N+2} . Пусть это не так, и продолжения f_{N+2} не существует. Рассмотрим систему (9), соответствующую $k = 2$. Она также не должна иметь решения. Однако система, решаемая на предыдущем шаге (при $k = 1$), лишь правой частью отличается от рассмотренной. Эта система имеет решение, причем по Теореме 3.2 это решение должно быть единственным. Поэтому она разрешима и однозначно разрешима при любой правой части. Т.о., система (9), соответствующая $k = 2$, обязана иметь единственное решение. ■

Таким образом, вышеприведенные теоремы дают исчерпывающую информацию о единственности прогноза дискретного ряда. Заметим однако, что не любой ряд может быть продолжен. В следующей Лемме рассмотрен класс рядов, не допускающих продолжения.

Лемма 3.1 (*о рядах не допускающих продолжения*) *Пусть ряд $(f_i)_{i=1}^N$ имеет ранг r , допускает продолжение f_{N+1} , и пусть значение f_{N+2} не является продолжением этого ряда (в смысле Определения 2.4). Тогда ряд $(f_i)_{i=1}^{N+2}$ не допускает продолжения.*

Доказательство

Сразу отметим, что ряд $(f_i)_{i=1}^{N+2}$ имеет ранг $r + 1$, иначе $(f_i)_{i=1}^{N+2}$ есть продолжение $(f_i)_{i=1}^{N+1}$. Покажем, что не существует f_{N+3} , такого, что ряд $(f_i)_{i=1}^{N+3}$ сохраняет ранг.

Рассмотрим систему собственных векторов $v^{(1)}, \dots, v^{(r)}$, порожденную рядом $(f_i)_{i=1}^N$, и пусть $L^{(r)} = \text{Lin}(v^{(1)}, \dots, v^{(r)})$ — линейная оболочка векторов $v^{(1)}, \dots, v^{(r)}$. Добавленное f_{N+2} порождает точку многомерной выборки $X^{(n+2)} \notin L^{(r)}$. Действительно, по Лемме 3.3 существует \hat{f}_{N+2} — единственное продолжение $(f_i)_{i=1}^{N+1}$, причем $\hat{X}^{(n+2)} \in L^{(r)}$. Однако

$$X^{(n+2)} - \hat{X}^{(n+2)} = (0, \dots, (f_{N+2} - \hat{f}_{N+2}))^T \neq 0,$$

что противоречит единственности решения системы (7). Поэтому гиперплоскость, содержащая $(f_i)_{i=1}^{N+2}$ есть $L^{(r+1)} = \text{Lin}(v^{(1)}, \dots, v^{(r)}, v^{(r+1)})$, где $v^{(r+1)} = (0, \dots, 1)^T$. Существование f_{N+3} эквивалентно совместности системы

$$\begin{cases} \sum_{j=1}^r h_j v_1^{(j)} &= f_{N-\tau+4} \\ &\dots \\ \sum_{j=1}^r h_j v_{\tau-1}^{(j)} &= f_{N+2}. \end{cases} \quad (10)$$

Пусть система (10) совместна. Тогда соответствующая ей однородная систе-

ма имеет только тривиальные решения. Однако это не так, поскольку среди векторов $\hat{v}^{(j)} = (v_1^{(j)}, \dots, v_{\tau-1}^{(j)})^T$ имеется нулевой: $\hat{v}^{(r+1)}$, и т.о. система (10) несовместна.

■

Также представляющим интерес вопросом является выяснение минимального числа точек, по которым возможно восстановить ряд ранга τ . Справедлива Лемма 3.2 которую мы оставляем без доказательства ввиду ее тривиальности.

Лемма 3.2 Для восстановления ряда ранга r всегда достаточно иметь $N = r + \tau$ точек.

Замечание 3.2 Т.к. условия метода предполагают $\tau > r$, то для восстановления ряда ранга r всегда достаточно иметь $N = 2r + 1$ точку.

Попытаемся теперь посмотреть на метод экстраполяции функций дискретного аргумента с несколько иной точки зрения. Доказанные теоремы позволяют утверждать, что для класса функций дискретного аргумента, являющихся решениями линейных конечно-разностных уравнений порядка r с постоянными коэффициентами, рассмотренный метод позволяет однозначно восстанавливать такую функцию по ее первым $2r + 1$ точкам. Верно и обратное: всякая функция, однозначно восстанавливаемая указанным методом по некоторому числу начальных значений, есть решение некоторого линейного конечно-разностного уравнения с постоянными коэффициентами. В этом контексте Теоремы 3.3, 3.2 аналогичны хорошо известным (см. [5]) условиям существования и единственности решений таких уравнений.

4 Примеры прогнозирования детерминированных функций дискретного аргумента.

Пример 4.1 Линейная функция $f_n = n$ является функцией ранга 2. Соответствующее к.р.у. имеет вид $f_{n+2} - 2f_{n+1} + f_n = 0$, с начальными условиями $f_1 = 1, f_2 = 2$. На Рис. 1 показано прогнозирование последовательности $0, 1, 2, 3, 4$ по первым двум главным компонентам на 4 шага вперед.

Пример 4.2 Показательная функция $f_n = 2^n$ также является функцией ранга 2. Она является решением к.р.у. $f_{n+2} - f_{n+1} - 2f_n = 0$, при начальных условиях $f_1 = 1, f_2 = 2$. На Рис. 2 показано прогнозирование последовательности $1, 2, 4, 8, 16$ по первым двум главным компонентам на 1 шаг вперед.

Пример 4.3 Периодическая функция ранга 1, с периодом 2, $f_n = (-1)^n$ прогнозируется по первым 3-м значениям: см. Рис. 3. Соответствующее к.р.у. имеет вид $f_{n+1} + f_n = 0$, с начальным условием $f_1 = -1$.

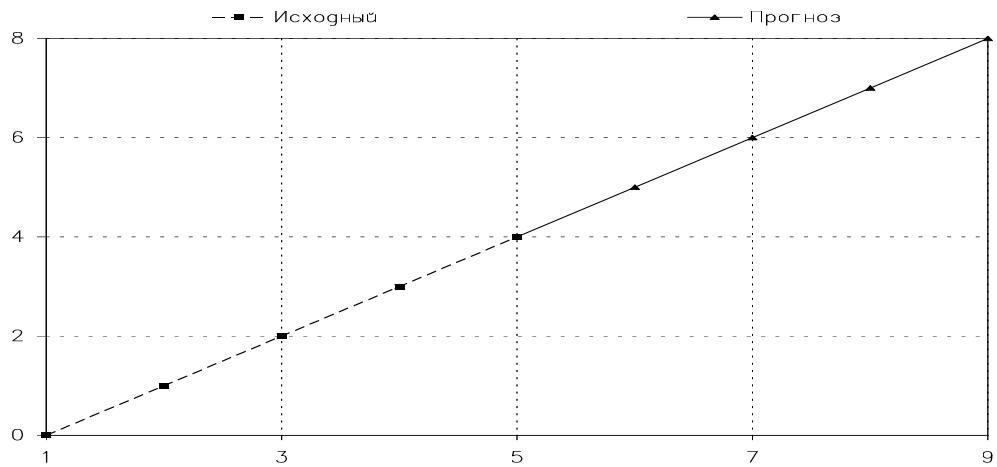


Рис. 1. Прогнозирование линейной функции.

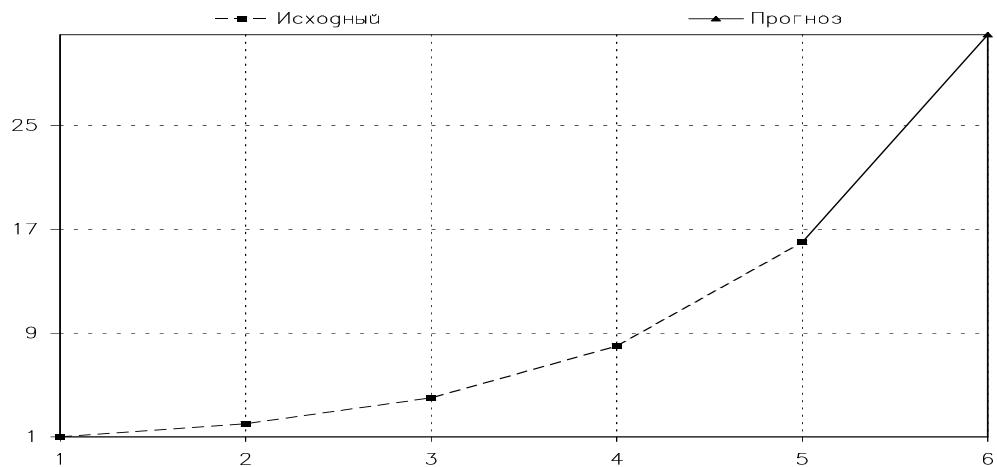


Рис. 2. Прогнозирование показательной функции.

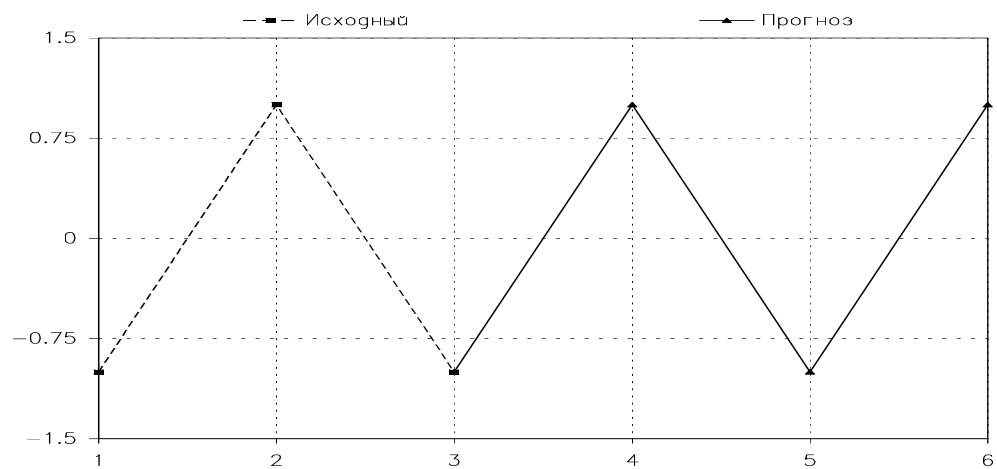


Рис. 3. Прогнозирование простейшей периодической функции.

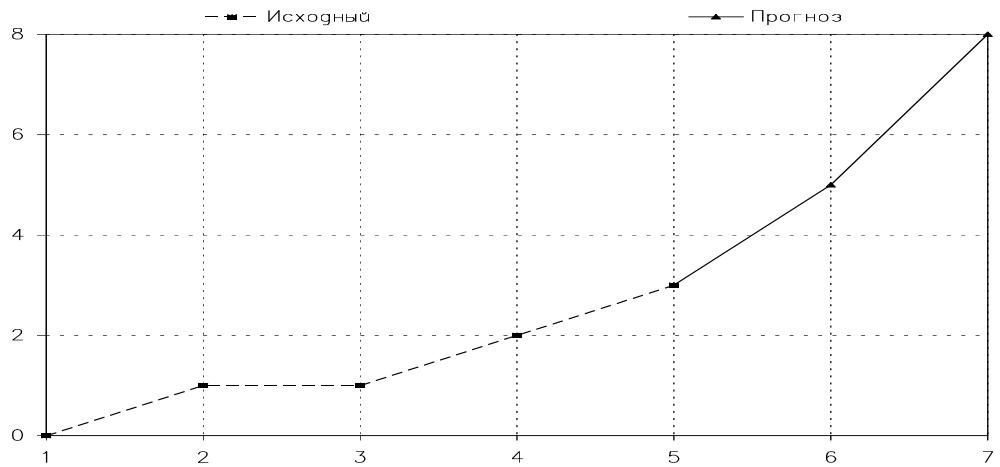


Рис. 4. Прогнозирование последовательности Фибоначчи.

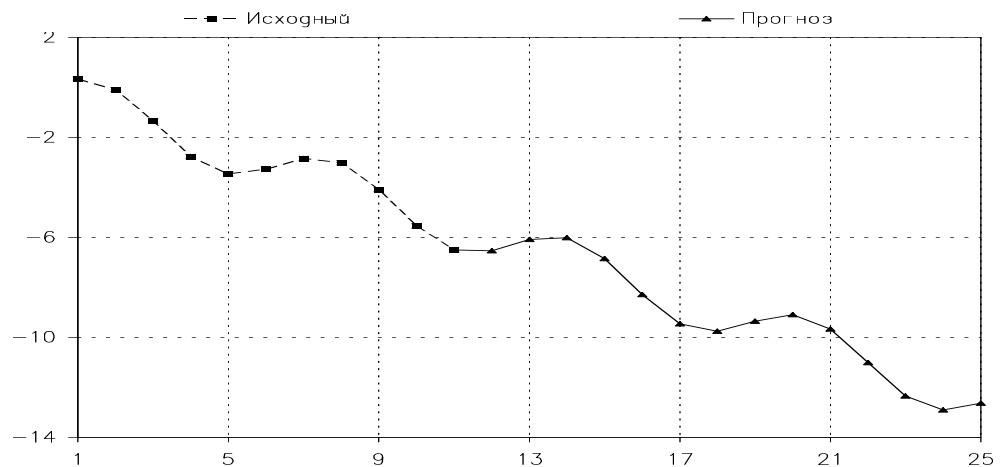


Рис. 5. Прогнозирование суммы линейной и периодической функций.

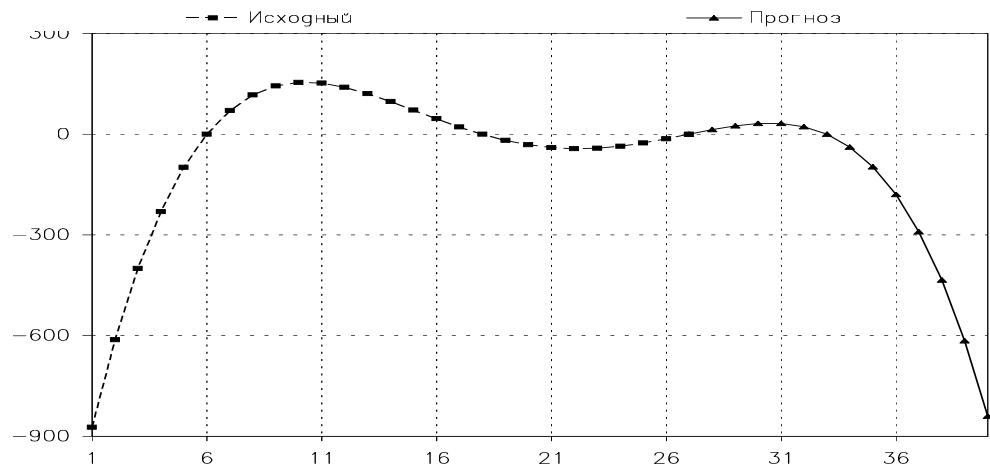


Рис. 6. Прогнозирование полинома 4 порядка.

Пример 4.4 Несколько неожиданным, но эффектным примером прогнозирования функции ранга 2 является прогнозирование последовательности Фибоначчи. Действительно, такая последовательность является решением конечно-разностного уравнения второго порядка

$$f_{n+2} - f_{n+1} - f_n = 0, \quad f_1 = 1, f_2 = 2,$$

и по доказанным нами утверждениям прогнозируется однозначно по 2-м главным компонентам по первым пяти точкам. См. Рис. 4.

Пример 4.5 Примером, имеющим большее отношение к статистическим приложениям, является случай суммы линейной и периодической функций

$$f_n = \sin(n) - \frac{n}{2}, \quad x_n = 1, 2, \dots, 30.$$

Ранг такой функции равен 4, и прогнозирование производится по 4-м главным компонентам, см. Рис. 5.

Пример 4.6 В качестве примера прогнозирования функции ранга 5 взят полином четвертого порядка.

$$f_n = -(x_n - 4)(x_n - 2)(x_n + 1)(x_n + 5), \quad x_n = -7 + n/3.$$

Интересно, что в данном случае метод способен предсказать как “подъем”, так и последующий спад, несмотря на то, что последние наблюдаемые значения относятся к участку функции, где она после локального минимума только начала нарастание. Соответствующий график приведен на Рис. 6.

Отметим, что в последних двух графиках показано большее число точек чем необходимо для прогнозирования (см. Лемма 3.2).

5 Модификация метода: корреляционный прогноз.

Основная идея метода — использование конечномерности выборки, порожденной времененным рядом, продолжает работать и тогда, когда получаемая при гусеничной обработке выборка в дальнейшем центрируется и нормируется. Критерием вырожденности многомерного распределения может быть вырожденность как матрицы вторых моментов, так и корреляционной (ковариационной) матриц. Т.о., если имеется ряд конечного ранга, то у него как ненормированная, так и центрированная нормированная выборки лежат в гиперплоскостях размерности меньшей τ . Опишем подробнее этот вариант прогноза. Введем дополнительно несколько определений:

Определение 5.1 Вектор

$$\bar{x}_i = \frac{1}{n} \sum_{s=1}^n f_{s+i-1}, \quad i = 1, \dots, \tau,$$

будем называть вектором гусеничного среднего.

Определение 5.2 Диагональную матрицу

$$\mathbf{D} = \begin{pmatrix} d_{11} & 0 & \dots & 0 & 0 \\ 0 & d_{22} & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 0 & d_{\tau\tau} \end{pmatrix}, \text{ где } d_{ii} = \sum_{s=1}^n f_{s+i-1}^2 - (\bar{x}_i)^2.$$

будем называть матрицей гусеничного стандарта.

Определение 5.3 Матрицу

$$Z = \mathbf{D}^{-1} Y \quad (11)$$

будем называть центрированной нормированной матрицей наблюдений, соответствующей последовательности f_i .

Уравнение гиперплоскости, содержащей центрированную нормированную выборку, также можно записать в параметрической форме:

$$z(\bar{l}) = \sum_{i=1}^r \bar{l}_i u^{(i)}$$

(под $u^{(i)}$ подразумеваются базисные векторы этой гиперплоскости.) Как и ранее, каждой точке центрированной нормированной многомерной выборки соответствует некоторое значение вектора параметризации \bar{l} , т.е.

$$Z^{(1)} = z(\bar{l}^{(1)}), \dots, Z^{(n)} = z(\bar{l}^{(n)}).$$

Последней, $(n+1)$ -й точке выборки, соответствует значение вектора параметризации $\bar{l}^{(n+1)}$. Восстанавливая значение $\bar{l}^{(n+1)}$ по набору $(f_i)_{i=1}^N$, а затем вычисляя последнюю компоненту точки $Z^{(n+1)}$, мы можем также определить соответствующее значение f_{N+1} . Заметим, однако, что в отличие от алгоритма, описанного в разделе 3, задача восстановления f_{N+1} оказывается нелинейной.

Проведем формальное описание алгоритма:

Алгоритм 5.1 Пусть имеется набор последовательных значений функции дискретного аргумента, образующий ряд $(f_i)_{i=1}^N$ ранга r .

1. В соответствии с (11) сформируем многомерную матрицу наблюдений.
2. Проведем построение корреляционной матрицы.
3. Найдем базис $u^{(1)}, \dots, u^{(r)}$, соответствующий отличным от 0 собственным числам корреляционной матрицы.
4. Запишем систему линейных алгебраических уравнений следующего вида:

$$\begin{cases} \sum_{j=1}^r l_j u_1^{(j)} = z_1^{(n+1)} \\ \dots \\ \sum_{j=1}^r l_j u_{\tau-1}^{(j)} = z_{\tau-1}^{(n+1)} \end{cases}. \quad (12)$$

5. Исследуем систему (7) на совместность.

5.1 Если система (12) несовместна, то исходный ряд не допускает продолжения.

5.2 Если система (12) имеет решение, то исходный ряд имеет продолжение, и это продолжение находится как решение уравнения

$$\frac{z^* - \frac{1}{N-\tau+2} \left(\sum_{s=\tau}^N f_s + z^* \right)}{\sqrt{\frac{1}{N-\tau+2} \left\{ (z^*)^2 + \sum_{s=\tau}^N f_s^2 \right\} - \left\{ \frac{1}{N-\tau+2} \left(z^* + \sum_{s=\tau}^N f_s \right) \right\}^2}} = \sum_{j=1}^r l_j^* u_\tau^{(j)}, \quad (13)$$

где l_1^*, \dots, l_r^* — решение системы (12). Уравнение (13) имеет, вообще говоря, два корня, отличающиеся знаком, который они доставляют левой части уравнения. Поэтому после решения (13) проблем с выбором корня не возникает.

6 Детерминированные ряды, обобщенный прогноз. Уравнение прогноза.

В этом разделе нами будут рассматриваться ряды, не являющиеся рядами конечного ранга. В этом случае исходный ряд может быть полностью восстановлен только по полному набору главных компонент. Однако, возможно восстановление таких рядов “частично”, по интересующему нас поднабору главных компонент. При этом реализуется желание исследователя отбирать информативные главные компоненты. В этой связи мы несколько изменим обозначения, введенные нами раньше.

Вновь рассмотрим систему линейных уравнений (7). Введем следующие обозначения: пусть

$$\mathbf{v} = (v_\tau^{(i_1)}, v_\tau^{(i_2)}, \dots, v_\tau^{(i_r)}), \text{ где } 0 < i_1 < \dots < i_r < \tau, \text{ и}$$

$$\mathbf{V}_* = \begin{pmatrix} v_1^{(i_1)} & v_1^{(i_2)} & \dots & v_1^{(i_r)} \\ v_2^{(i_1)} & v_2^{(i_2)} & \dots & v_2^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{\tau-1}^{(i_1)} & v_{\tau-1}^{(i_2)} & \dots & v_{\tau-1}^{(i_r)} \end{pmatrix}.$$

Заметим, что

$$\mathbf{V}_X = \begin{pmatrix} \mathbf{V}_* \\ \mathbf{v} \end{pmatrix}. \quad (14)$$

Также пусть $Q = (q_i)_{i=1}^{\tau-1} = (x_1^{(n+1)}, \dots, x_{\tau-1}^{(n+1)})^T = (f_{N-\tau+2}, \dots, f_N)^T$ и $\bar{h} = (h_1, \dots, h_r)^T$. В этих обозначениях система (7) запишется как

$$\mathbf{V}_* \bar{h} = Q. \quad (15)$$

В разделе 3 показано, что такая система, вообще говоря, несовместна, хотя и существует широкий класс функций, для которых она разрешима. Расширим понятие решения системы (15).

Определение 6.1 Обобщенным решением системы (15) назовем решение системы

$$\mathbf{V}_*^T \mathbf{V}_* \bar{h} = \mathbf{V}_*^T Q. \quad (16)$$

Определение 6.2 Величину

$$b = \mathbf{v} \bar{h}_*, \quad (17)$$

где \bar{h}_* — решение системы (16), назовем обобщенным продолжением рассматриваемой функции дискретного аргумента. Учитывая (16), можно записать для прогнозируемого выражения x_{N+1} следующую формулу:

$$x_{N+1} = \mathbf{v} (\mathbf{V}_*^T \mathbf{V}_*)^{-1} \mathbf{V}_*^T Q. \quad (18)$$

Из общей теории решения линейных систем (см например [13], [14]) следует, что общее решение (16) может быть записано в виде:

$$\bar{h}_* = C \mathbf{V}_*^T Q + (E - H)w,$$

где $H = C \mathbf{V}_*^T \mathbf{V}_*$, $C = (\mathbf{V}_*^T \mathbf{V}_*)^{-1}$ — обобщенная обратная к $\mathbf{V}_*^T \mathbf{V}_*$, и w — произвольный вектор. Очевидно, обобщенное продолжение будет единствено, если $\mathbf{v} \bar{h}_* = const$. Следующая теорема дает критерий единственности обобщенного продолжения функции дискретного аргумента.

Теорема 6.1 Условие

$$\mathbf{v}(E - H) = 0 \quad (19)$$

является необходимым и достаточным для единственности обобщенного продолжения ряда $(f_i)_{i=1}^N$.

Доказательство опускается ввиду очевидности.

Заметим, что условие (19) может быть записано в следующем виде:

$$\mathbf{v}(E - (\mathbf{V}_*^T \mathbf{V}_*)^{-1} (\mathbf{V}_*^T \mathbf{V}_*)) = 0. \quad (20)$$

Матрица $\mathbf{V}_*^T \mathbf{V}_*$, входящая в равенство (20), есть не что иное как матрица Грамма системы векторов $\hat{v}^{(1)}, \dots, \hat{v}^{(r)}$. (Напомним, что $\hat{v}^{(s)} = Pr_{L_\tau} v^{(s)}$.) Если эта система векторов линейно независима, то $\det(\mathbf{V}_*^T \mathbf{V}_*) \neq 0$, и $\mathbf{V}_*^T \mathbf{V}_*$ обратима в обычном смысле. В этом случае равенство (20) тривиальным образом выполняется. Т.о., справедлива следующая лемма.

Лемма 6.1 Для однозначного продолжения функции f_1, \dots, f_N достаточно, чтобы система векторов $(\hat{v}^{(1)}, \dots, \hat{v}^{(r)})$ была бы линейно независима.

Факты, изложенные в разделе 3 позволяют сделать следующее заключение:

Следствие 6.1 Любой ряд конечного ранга имеет единственное обобщенное продолжение.

Заметим, что в качестве сомножителя Q в правой части системы (16) вместо первых $\tau - 1$ компоненты последнего столбца матрицы данных исходного ряда могут быть взяты соответствующие компоненты матрицы данных восстановленного ряда.

В практических исследованиях нередка ситуация, когда из сложной функции требуется выделить интересующее нас слагаемое. Вопросы разделимости

слагаемых в сумме подробно исследованы в [12]. Рассмотрим вопрос продолжимости слагаемого в составе суммы. Пусть

$$f_i = f_i^{(1)} + f_i^{(2)}, \quad i = 1, \dots, N \quad (21)$$

и $f^{(1)}$ считается информативной частью ряда. Мы собираемся спрогнозировать $f^{(1)}$ в точке $N + 1$, т.е. найти $f_{N+1}^{(1)}$. Далее мы будем отмечать матрицы и векторы, порожденные рядами $f^{(1)}, f^{(2)}$ соответствующим верхним индексом. Пусть набор главных компонент (i_1, i_2, \dots, i_r) задан таким образом, что $f^{(1)}$ полностью восстанавливается по этому набору, иными словами выделяется из суммы, т.е.

$$\begin{aligned} \tilde{X}^{(1)} &= V^{(1)}(V^{(1)})^T(X^{(1)} + X^{(2)}) = V^{(1)}(V^{(1)})^T X^{(1)} + V^{(1)}(V^{(1)})^T X^{(2)} = \\ &= X^{(1)} + V^{(1)}(V^{(1)})^T X^{(2)} = X^{(1)}. \end{aligned}$$

Согласно [12] для этого необходимо и достаточно, чтобы системы собственных векторов и главных компонент, порождаемых матрицами $X^{(1)}$, и $X^{(2)}$, были бы взаимно ортогональны. Запишем формулу (18) для прогнозирования $f_{N+1}^{(1)}$:

$$\begin{aligned} \hat{f}_{N+1}^{(1)} &= \mathbf{v}^{(1)}((\mathbf{V}_*^{(1)})^T \mathbf{V}_*^{(1)})^{-1} (\mathbf{V}_*^{(1)})^T (Q^{(1)} + Q^{(2)}) = \\ &= \mathbf{v}^{(1)}((\mathbf{V}_*^{(1)})^T \mathbf{V}_*^{(1)})^{-1} (\mathbf{V}_*^{(1)})^T Q^{(1)} + \mathbf{v}^{(1)}((\mathbf{V}_*^{(1)})^T \mathbf{V}_*^{(1)})^{-1} (\mathbf{V}_*^{(1)})^T Q^{(2)} = \\ &= f_{N+1}^{(1)} + \mathbf{v}^{(1)}((\mathbf{V}_*^{(1)})^T \mathbf{V}_*^{(1)})^{-1} (\mathbf{V}_*^{(1)})^T Q^{(2)}. \end{aligned}$$

Из последнего выражения следует, что разделимость слагаемого, вообще говоря, не гарантирует его точного прогнозирования. Условия однозначной прогнозируемости слагаемого рассмотрены подробнее в статье [6]. Пока же отметим лишь, что невозможность *точного* прогнозирования тем не менее оставляет возможность приближенного прогнозирования.

7 Аналитическое решение уравнения прогноза.

Рассмотрим выражение (18). Правая часть этого выражения, по существу, представляет собой скалярное произведение некоторого вектора P_v на вектор – отрезок исходного ряда Q , т.е.

$$f_{N+1} = P_v Q,$$

где $P_v = \mathbf{v}(\mathbf{V}_*^T \mathbf{V}_*)^{-1} \mathbf{V}_*^T$. Учитывая очевидное равенство $\mathbf{V}_*^T \mathbf{V}_* + \mathbf{v}^T \mathbf{v} = E$, которое есть просто матричная форма записи условия ортогональности системы собственных векторов, получаем $\mathbf{V}_*^T \mathbf{V}_* = E - \mathbf{v}^T \mathbf{v}$. Если при этом выполняется условие $\|\mathbf{v}^T \mathbf{v}\| < 1$ (в некоторой норме), то обратная к матрице $(\mathbf{V}_*^T \mathbf{V}_*)$ может быть найдена в виде ряда Неймана для оператора $(E - \mathbf{v}^T \mathbf{v})^{-1}$:

$$(\mathbf{V}_*^T \mathbf{V}_*)^{-1} = E + \sum_{k=1}^{\infty} (\mathbf{v}^T \mathbf{v})^k. \quad (22)$$

Обозначая $a = \mathbf{v}\mathbf{v}^T$ и учитывая, что $(\mathbf{v}^T\mathbf{v})^k = a^{k-1}\mathbf{v}^T\mathbf{v}$ получаем для P_v выражение

$$P_v = \frac{1}{1-a} \mathbf{v} \mathbf{V}_*^T. \quad (23)$$

Еще раз отметим, что Q есть просто вектор, составленный из элементов ряда f , начиная с $N-\tau+2$ и заканчивая N . Поэтому формула (23) является в некотором смысле конечно-разностным уравнением, однако, порядок этого уравнения есть $\tau-1$ и поэтому даже для рядов конечного ранга это уравнение не эквивалентно исходному к.р.у. Спектр решений этого уравнения естественно называть *гусеничным спектром* по аналогии с хорошо известным в анализе временных рядов *спектром Прони* (см. [11]).

Нетрудно вычислить норму вектора P_v :

$$\|P_v\|_{III}^2 = P_v P_v^T = \frac{a}{1-a}. \quad (24)$$

Выясним смысл величины a . Легко видеть, что

$$a = \sum_{k=1}^r (v_\tau^{(i_k)})^2. \quad (25)$$

Заметим, что слагаемые в сумме (25) представляют собой квадраты координат вектора $(0, \dots, 0, 1)$ в системе базисных векторов $v^{(i_k)}$. Поэтому $a \leq 1$ для любой системы $v^{(i_k)}$, и $a = 1$ тогда и только тогда, когда вектор $(0, \dots, 0, 1)$ принадлежит линейной оболочке векторов $v^{(i_k)}$. Последнее, впрочем, не выполняется ни для какого ряда конечного ранга (поэтому для них всегда $a < 1$). Другая интерпретация: a — квадрат длины проекции вектора $(0, \dots, 0, 1)$ на линейную оболочку векторов $v^{(i_k)}$ (\sqrt{a} имеет смысл косинуса угла между вектором $(0, \dots, 0, 1)$ и линейной оболочкой векторов $v^{(i_k)}$). Отсюда, в частности следует, что величина a инвариантна относительно поворотов системы векторов $v^{(i_k)}$. В дальнейшем наибольший интерес будут представлять ряды, у которых $a \rightarrow 0$ с ростом τ . Сравнительно легко показать, что к таким рядам принадлежат, например, некоторые периодические ряды. Действительно, рассмотрим функцию дискретного аргумента $f_i = \sin(\omega i + \phi)$. Тогда $f_{i+j} = \sin(\omega i + \phi) \cos(\omega j) + \cos(\omega i + \phi) \sin(\omega j)$, и поэтому $X^{(i)} = \sin(\omega i + \phi) Y_1' + \cos(\omega i + \phi) Y_2'$, где $Y_1' = (1, \cos(\omega), \dots, \cos(\omega(\tau-1)))$, $Y_2' = (0, \sin(\omega), \dots, \sin(\omega(\tau-1)))$. Пусть теперь $\omega = 2\pi k/\tau$ для некоторого $k > 1$. В этом случае Y_1' и Y_2' ортогональны и без ограничения общности могут быть рассмотрены в качестве базисных векторов соответствующей гиперплоскости. Базис главных компонент этой гиперплоскости будет отличаться от базиса Y_1', Y_2' , возможно, лишь поворотом. Непосредственное вычисление a дает нам:

$$a = \frac{\sin^2(\omega(\tau-1))}{\|Y_1'\|^2} + \frac{\cos^2(\omega(\tau-1))}{\|Y_2'\|^2}.$$

Учитывая, что

$$\begin{aligned} \|Y_1'\|^2 &= \frac{1}{2} \left(\tau + \sin^2(\tau\omega) + \frac{\cos(\omega) \cos(\tau\omega) \sin(\tau\omega)}{\sin(\omega)} \right), \\ \|Y_2'\|^2 &= \frac{1}{2} \left(\tau - \sin^2(\tau\omega) - \frac{\cos(\omega) \cos(\tau\omega) \sin(\tau\omega)}{\sin(\omega)} \right), \end{aligned}$$

получаем $a = 1/\tau + o(1/\tau)$ для $\tau \rightarrow \infty$. Другим простым примером такого рода рядов служит константный ряд, либо линейная функция дискретного аргумента. Рассмотрим подробнее последний случай. Пусть τ , для определенности, нечетно. Тогда в качестве базисных векторов могут быть выбраны векторы

$$Y_1' = \frac{1}{\sqrt{2s+1}}(1, \dots, 1),$$

$$Y_2' = \frac{\sqrt{3}}{\sqrt{s(s+1)(2s+1)}}(-s, 1-s, \dots, s-1, s),$$

где $s = (\tau-1)/2$. Вычисляя a , получаем $a = (4s+1)/((s+1)(2s+1))$. Как легко видеть, для всех рассмотренных выше рядов с ростом τ норма прогнозирующего оператора также стремится к нулю.

8 О связи алгоритма прогноза с классическими регрессионными задачами.

В данном разделе мы опишем алгоритм прогнозирования с точки зрения регрессионного анализа. Рассмотрим классическую линейную регрессионную модель:

$$y = \mathbf{V}_* \theta + \varepsilon \quad (26)$$

где $y = (y_i)_{i=1}^{\tau-1}$ — вектор-столбец результатов наблюдений, $\mathbf{V}_* = (v_{ij})_{i=1, \dots, \tau-1}^{j=1, \dots, r}$ — регрессионная матрица, $\theta = (\theta_j)_{j=1, \dots, r}$ — вектор-столбец неизвестных оцениваемых параметров, $\varepsilon = (\varepsilon_i)_{i=1, \dots, \tau-1}$ — вектор-столбец случайных ошибок.

Пусть линейная модель (26) невырождена, т.е. $\det \mathbf{V}_*^T \mathbf{V}_* \neq 0$. Тогда, как известно, МНК оценка $\hat{\theta}$ параметра θ есть решение системы нормальных уравнений:

$$\mathbf{V}_*^T \mathbf{V}_* = \mathbf{V}_*^T y, \quad (27)$$

т.е. $\hat{\theta} = (\mathbf{V}_*^T \mathbf{V}_*)^{-1} \mathbf{V}_*^T y$.

В современных статистических задачах достаточно частой является ситуация, когда в качестве \mathbf{V}_* выбирается матрица с ортогональными либо ортонормальными столбцами. Модели такого типа встречаются, например, в полиномиальной регрессии (приближение ортогональными полиномами). Кроме того, выбор регрессионных матриц такого вида удовлетворяет некоторым свойствам оптимальности (см. например [13], с.207).

Нами рассматривается регрессионная модель вида (26), где вообще говоря $\mathbf{V}_*^T \mathbf{V}_* \neq E$. Предположим, однако, что существует $\mathbf{v} = (v^1, \dots, v^r)$ такие, что матрица

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_* \\ \mathbf{v} \end{pmatrix} \quad (28)$$

уже ортонормирована ($\mathbf{V}^T \mathbf{V} = E$), и таким образом, исходная матрица \mathbf{V}_* близка к ортонормированной. Покажем, что при выполнении некоторых (естественных) условий на \mathbf{v} , матрица плана имеет обратную, которая легко вычисляется аналитически. Более точно, справедлива следующая теорема:

Теорема 8.1 Пусть имеется модель вида (26), причем $\det \mathbf{V}_*^T \mathbf{V}_* \neq 0$ и $\mathbf{V}_*^T \mathbf{V}_* + \mathbf{v}^T \mathbf{v} = E$. Тогда

$$(\mathbf{V}_*^T \mathbf{V}_*)^{-1} = E + \frac{\mathbf{v}^T \mathbf{v}}{1 - a}, \quad (29)$$

где $a = \mathbf{v} \mathbf{v}^T < 1$.

Доказательство

Пусть $a = \mathbf{v} \mathbf{v}^T < 1$. Рассмотрим $(\mathbf{V}_*^T \mathbf{V}_*)^{-1}$. Справедлива следующая цепочка равенств:

$$(\mathbf{V}_*^T \mathbf{V}_*)^{-1} = (E - \mathbf{v}^T \mathbf{v})^{-1} = E + \sum_{k=1}^{\infty} (\mathbf{v}^T \mathbf{v})^k = E + \mathbf{v}^T \mathbf{v} \sum_{k=0}^{\infty} a^k = E + \frac{\mathbf{v}^T \mathbf{v}}{1 - a}, \quad (30)$$

которая по существу представляет собой последовательное вычисление ряда Неймана для $(E - \mathbf{v}^T \mathbf{v})^{-1}$. Покажем, что для каждой модели полного ранга $a < 1$.

Действительно,

$$a = \mathbf{v} \mathbf{v}^T = \sum_{k=1}^r \mathbf{v}_k^2. \quad (31)$$

Однако, слагаемые в сумме (31) представляют собой квадраты координат вектора $(0, \dots, 0, 1)$ в координатной системе, имеющей в качестве базисных векторов столбцы "расширенной" регрессионной матрицы \mathbf{V} . Поэтому $a \leq 1$ для любой регрессионной модели с матрицей $\mathbf{V}_*^T \mathbf{V}_* = E$, и $a = 1$, тогда и только тогда, когда вектор $(0, \dots, 0, 1)$ принадлежит линейной оболочке столбцов \mathbf{V} . Но в этом случае столбцы \mathbf{V}_* линейно зависимы, и $\det \mathbf{V}_*^T \mathbf{V}_* = 0$, что противоречит условию теоремы. ■

Естественно далее перейти к задаче несмещенной оценки линейных функций θ . Пусть вновь имеется модель вида (26), и условия Теоремы 8.1 выполнены. Найдем наилучшую линейную оценку следующей линейной функции $\eta(\theta) = \mathbf{v} \theta$.

Лемма 8.1 МНК оценкой для $\eta(\theta) = \mathbf{v} \theta$ является функция

$$\hat{\eta} = \frac{\mathbf{v} \mathbf{V}_*^T}{1 - a} y. \quad (32)$$

Доказательство

Стандартными выкладками показывается, что $\hat{\eta} = \mathbf{v} (\mathbf{V}_*^T \mathbf{V}_*)^{-1} \mathbf{V}_*^T y$. Учитывая (29) получаем

$$\mathbf{v} (\mathbf{V}_*^T \mathbf{V}_*)^{-1} \mathbf{V}_*^T = \mathbf{v} (E + \frac{\mathbf{v}^T \mathbf{v}}{1 - a}) \mathbf{V}_*^T = (\mathbf{v} + \frac{a}{1 - a} \mathbf{v}) \mathbf{V}_*^T = \frac{\mathbf{v} \mathbf{V}_*^T}{1 - a}. \quad (33)$$

■

Рассмотренная задача, по существу, представляет собой задачу предсказания значения функции регрессии в известной точке.

Отметим, что рассмотренный ранее алгоритм прогнозирования формально может быть представлен в виде (26). Однако в нашем случае матрица плана является, вообще говоря, матрицей случайной и более того, зависимой от ошибки наблюдений. С этой точки зрения алгоритм прогноза представляет собой решение регрессионной задачи вида (26) с переменной матрицей плана. Конечно, следует отметить, что наличие случайной компоненты в собственных векторах делает спорным *непосредственное* применение соответствующих результатов из регрессионного анализа, однако практика показывает, что старшие собственные векторы гусеничной ковариационной матрицы, соответствующие информативным компонентам, часто свободны (или почти свободны) от ошибки. Этот факт позволяет ожидать, что реальные свойства алгоритма прогноза будут близки к свойствам рассмотренных регрессионных задач.

9 Исследование асимптотических свойств уравнения прогноза.

Если ряд конечного ранга не выделяется из суммы (а следовательно, и не прогнозируется точно) при конечном N , он часто может быть выделен и продолжен асимптотически при $N \rightarrow \infty$. Этот случай необходимо рассмотреть подробнее.

Попробуем выяснить смысл вектора параметризации h . Очевидно, что в случае, когда мы рассматриваем обычную гусеничную схему (6), значение h для каждой конкретной точки $X^{(i)}$ есть вектор, составленный из i -х компонент ненормированных векторов главных компонент. Однако в случае прогноза это перестает быть верным, потому что собственные векторы и главные компоненты, вычисленные для ряда $(f_i)_{i=1}^{N+1}$, не совпадают с таковыми для ряда $(f_i)_{i=1}^N$. (Хотя для рядов конечного ранга гиперплоскость, порожденная системой собственных векторов, все же остается неизменной.) В случае прогноза, смысл $h^{(i)}$ тем не менее остается ясным: $h^{(i)}$ — это проекция точки многомерной выборки, (порожденной исходным, либо восстановленным рядом длины $N + 1$) на гиперплоскость, натянутую на систему базисных векторов, выбранных и вычисленных для ряда длины N . Запишем следующее уравнение

$$h = V^T X^{(n+1)} = (V_*^T, \mathbf{v}_*^T) X^{(n+1)}. \quad (34)$$

Это уравнение представляет собой запись проекции $(n + 1)$ -й точки матрицы наблюдений на линейную оболочку системы базисных векторов V . В этом случае $\mathbf{v}_* h$ — “восстановление” последней компоненты вектора $X^{(n+1)}$ по “главной компоненте” h . Качество этого восстановления характеризуется величиной $x_\tau^{(n+1)} - \mathbf{v}_* h$. Заметим, что $V_*^T V_* + \mathbf{v}_*^T \mathbf{v}_* = E$, тогда:

$$(V_*^T V_* + \mathbf{v}_*^T \mathbf{v}_*) h = (V_*^T, \mathbf{v}_*^T) X^{(n+1)}. \quad (35)$$

Рассмотрим уравнение прогноза

$$V_*^T V_* \bar{h} = V_*^T \bar{X}^{(n+1)}, \quad (36)$$

где $\bar{X}^{(n+1)} = P_{L_\tau} X^{(n+1)}$. Найдем разность $\Delta h = \bar{h} - h$:

$$V_*^T V_* \Delta h = \mathbf{v}_*^T (x_\tau^{(n+1)} - \mathbf{v}_* h)$$

и, таким образом, $\Delta h = (V_*^T V_*)^{-1} \mathbf{v}_*^T (x_\tau^{(n+1)} - \mathbf{v}_* h)$.

Пусть $\mathbf{v}_* \rightarrow 0$ покомпонентно. (Это требование неограничительно, поскольку чаще всего асимптотика рассматривается по $N \rightarrow \infty$ и $\tau \rightarrow \infty$, а у всех известных авторам рядов конечного ранга все компоненты стремятся к нулю с ростом τ .) В этом случае $(V_*^T V_*)^{-1} \rightarrow E$, и достаточным условием стремления к нулю погрешности Δh является ограниченность $x_\tau^{(n+1)} - \mathbf{v}_* h$. В свою очередь, малость Δh обуславливает малость разности $\mathbf{v} \bar{h}_* - \mathbf{v} h_*$, поскольку $\mathbf{v} \bar{h}_* - \mathbf{v} h_* = \mathbf{v} \Delta h$.

Таким образом: если при вычислении и выбранном для ряда $(f_i)_{i=1}^N$ наборе собственных векторов результат проектирования и восстановления матрицы X ряда $(f_i)_{i=1}^{N+1}$ хорошо приближает этот ряд в точке $N + 1$, то эта точка хорошо прогнозируется. Более формально этот результат можно сформулировать в виде следующей теоремы:

Теорема 9.1 Пусть $(f_i^{(1)})_{i=1}^{N+1}$ – ряд ранга r , и наблюдается $f_i = f_i^{(1)} + f_i^{(2)}$, $i = 1, \dots, N + 1$ при $N \rightarrow \infty$, $\tau = \tau(N)$. Пусть (i_1, \dots, i_r) – выбираемый набор главных компонент, а $v^{(i_1)}, \dots, v^{(i_r)}$ – соответствующие собственные векторы. Если при этом (в указанных выше обозначениях) выполняется $\mathbf{v}_* \rightarrow 0$, и $x_\tau^{(n+1)} - \mathbf{v}_* h = O(1)$, то $\mathbf{v}_* \bar{h} - \mathbf{v}_* h \rightarrow 0$.

10 Примеры прогнозирования слагаемых в сумме.

Пример 10.1 В качестве примера прогнозирования разделенного слагаемого приведем функцию, являющуюся суммой двух периодических компонент. Пусть $f_i^{(1)} = \sin(\frac{2\pi i}{18})$, а $f_i^{(2)} = \frac{2}{3}\sin(\frac{2\pi i}{6})$ (постоянный множитель $2/3$ выбран из требований наглядности), ряд состоит из 90 элементов. В соответствии с условиями отдельности периодической функции (см. [12]) выбор $\tau = 36$ приводит к разделению периодических слагаемых. На Рис. 7 изображены исходная функция, восстановление по первым двум главным компонентам и прогноз на 18 точек.

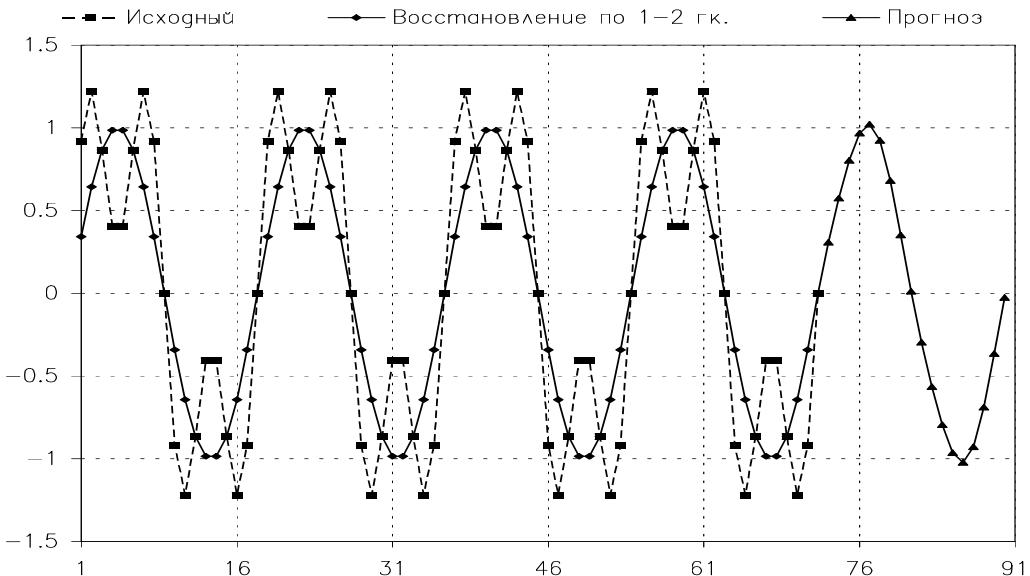


Рис. 7. Прогнозирование слагаемого в сумме.

Пример 10.2 Приведем пример асимптотического выделения и прогнозирования слагаемого. Пусть $f_i^{(1)} = \sin(\frac{2\pi i}{6})$, $f_i^{(2)} = -\frac{i}{3}$, и наблюдается сумма $f_i = f_i^{(1)} + f_i^{(2)}$, $i = 1, \dots, N$. Такие слагаемые не разделяются *точно* при конечных N, τ , однако могут быть выделены приближенно, со все большей точностью при увеличении длины ряда и параметра τ . Ниже приведен результат прогнозирования линейной части такой функции (тренда) при следующих величинах параметров ряда и гусеницы:

1. Рис. 8, $N = 10, \tau = 5$, для восстановления и прогноза использованы 1 и 2 ГК.
2. Рис. 9, $N = 20, \tau = 10$, для восстановления и прогноза использованы 1 и 2 ГК.
3. Рис. 10, $N = 60, \tau = 30$, для восстановления и прогноза использованы 1 и 4 ГК.

В каждом случае для прогноза выбирались главные компоненты, соответствующие линейной части тренда. Для обеспечения наглядности рисунков только последние 10 членов исходного ряда вместе с прогнозом на 5 шагов вперед изображались на рисунке. Хорошо видно “исправление” формы линейного тренда с ростом длины ряда и гусеницы.

11 Прогнозирование рядов со случайной ошибкой.

Схема, изложенная в предыдущем разделе переносится без серьёзных изменений на случай, когда детерминированная функция наблюдается со случайной ошибкой. Пусть

$$f_i = f_i^{(1)} + f_i^{(2)}, \quad i = 1, \dots, N, \quad (37)$$

где $f_i^{(1)}$ — детерминированный ряд конечного ранга (для определенности можно считать его ранг равным r) $f_i^{(2)} = \xi_i$ — случайный ряд (шум), представляющий собой последовательность н.о.р.с.в. с нулевым средним и дисперсией σ^2 . В этом случае естественно рассматривать получаемое продолжение дискретного ряда как оценку значения детерминированной части функции в соответствующей точке. Заметим, что в этом случае для любого N детерминированная часть исследуемой функции однозначно прогнозируется (в обобщенном смысле), и через f_{N+1} будет обозначаться её обобщенное продолжение.

Изучим некоторые свойства прогноза в этом случае. Пусть для прогнозирующего оператора ряда $f_i^{(1)}$ выполнено $a \rightarrow 0$, при $N \rightarrow \infty$. Обозначим этот прогнозирующий оператор как $P_v^{(1)}$. Выясним, какими свойствами обладает прогноз. Рассмотрим действие $P_v^{(1)}$ на сумму $f^{(1)} + f^{(2)}$

$$P_v^{(1)}(Q^{(1)} + Q^{(2)}) = f_{N+1}^{(1)} + P_v^{(1)}Q^{(2)}. \quad (38)$$

Случайная величина $P_v^{(1)}Q^{(2)}$ представляет собой сумму случайных величин с нулевыми средними. Ее второй момент может быть выражен как $E(P_v^{(1)}Q^{(2)})^2 = aE\xi_i^2 = a\sigma^2$, и поэтому $E(P_v^{(1)}Q^{(2)})^2 \rightarrow 0$, при $N \rightarrow \infty$.

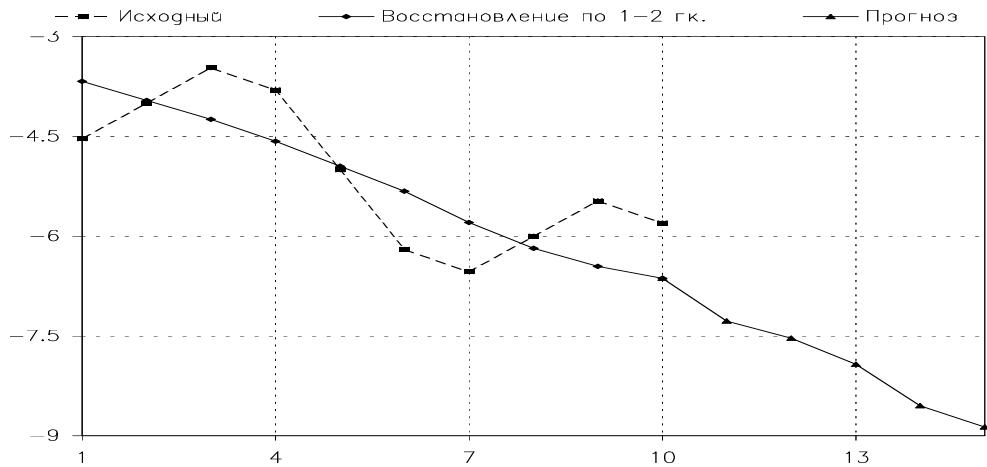


Рис. 8. Асимптотическая сходимость прогноза. Пример 10.2, $N = 10$, $\tau = 5$.

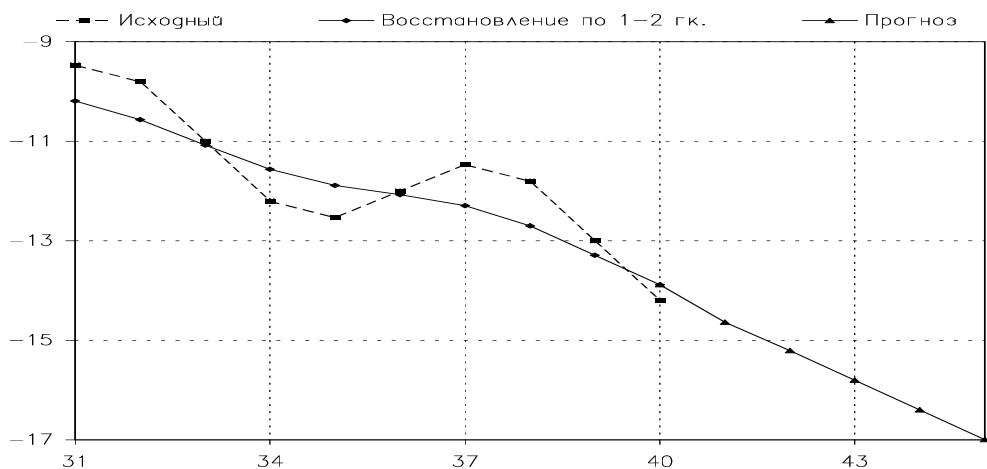


Рис. 9. Асимптотическая сходимость прогноза. Пример 10.2, $N = 20$, $\tau = 10$.

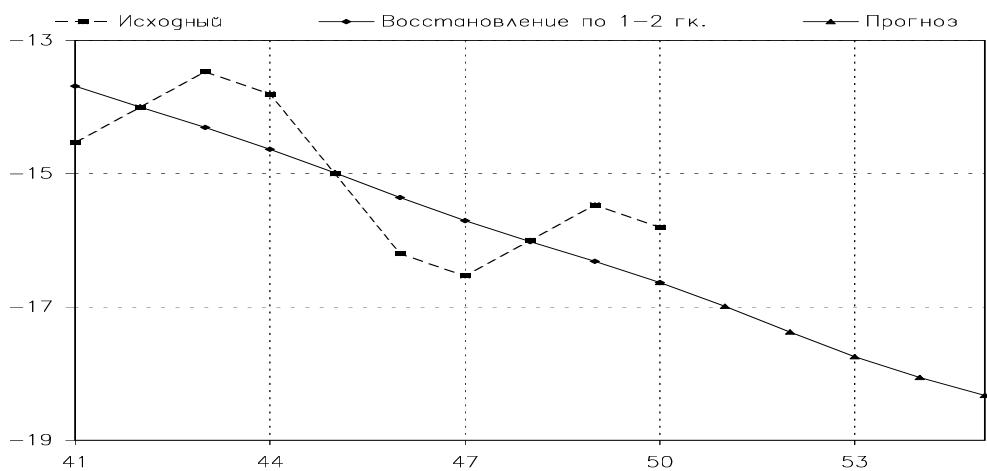


Рис. 10. Асимптотическая сходимость прогноза. Пример 10.2, $N = 60$, $\tau = 30$.

12 Примеры прогнозирования функций со случайной ошибкой.

Пример 12.1 Приведем пример асимптотического прогнозирования функции, наблюдаемой со случайной ошибкой. Пусть $f_i^{(1)} = \sin(\frac{2\pi i}{6}) - \frac{i}{3}$, а $f_i^{(2)}$ — независимые с.в., имеющие стандартное нормальное распределение. Ниже приведен результат прогнозирования детерминированного тренда такого ряда. Параметры ряда и гусеницы были выбраны следующими:

1. $N = 20, \tau = 10$, для Рис. 11, прогноз по 1-3 ГК.
2. $N = 30, \tau = 15$, для Рис. 12, прогноз по 1-3 ГК.
3. $N = 80, \tau = 40$, для Рис. 13, прогноз по 1-4 ГК.

В первых двух случаях для прогноза выбирались 1-3 ГК, поскольку на таких коротких рядах четвертая ГК оказалась неотделимой от шума. Как и в разделе 10, на рисунках приведены только последние 20 членов исходного ряда вместе с прогнозом на 10 шагов вперед.

13 Модификация метода: корреляционный прогноз.

Как уже упоминалось, в анализе главных компонент достаточно распространенным является использование, наряду с ковариационной матрицей, также и корреляционной. В разделе 5 было показано, как записывается алгоритм прогноза рядов конечного ранга, использующий корреляционную матрицу гусеничной многомерной выборки. Не возникает принципиальных трудностей в обобщении этого алгоритма на более широкий класс рядов. Заметим, что алгоритм прогнозирования рядов по центрированной, но не нормированной гусеничной матрице описан также в [17, 16].

Рассмотрим систему

$$U_* \bar{l} = Q, \quad (39)$$

где

$$U_* = \begin{pmatrix} u_1^{(i_1)} & u_1^{(i_2)} & \dots & u_1^{(i_r)} \\ u_2^{(i_1)} & u_2^{(i_2)} & \dots & u_2^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ u_{\tau-1}^{(i_1)} & u_{\tau-1}^{(i_2)} & \dots & u_{\tau-1}^{(i_r)} \end{pmatrix}, \text{ где } 0 < i_1 < \dots < i_r < \tau.$$

Здесь $u_k^{(i_s)}$ есть k -я компонента i_s -го собственного вектора корреляционной матрицы,

$$\bar{l} = (l_1, \dots, l_r)^T,$$

а вектор

$$Q = (z_1^{(n+1)}, \dots, z_{\tau-1}^{(n+1)})^T,$$

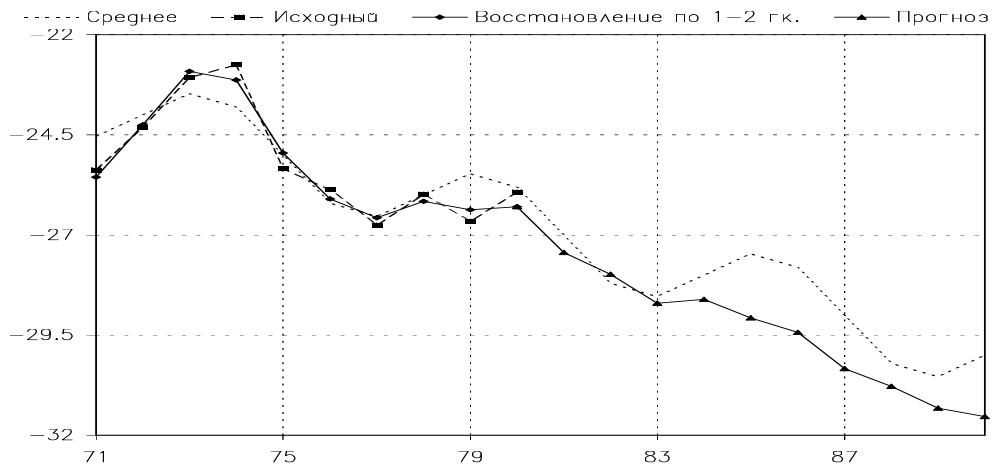


Рис. 11. Асимптотическая сходимость прогноза. Пример 12.1, $N = 20, \tau = 10$.

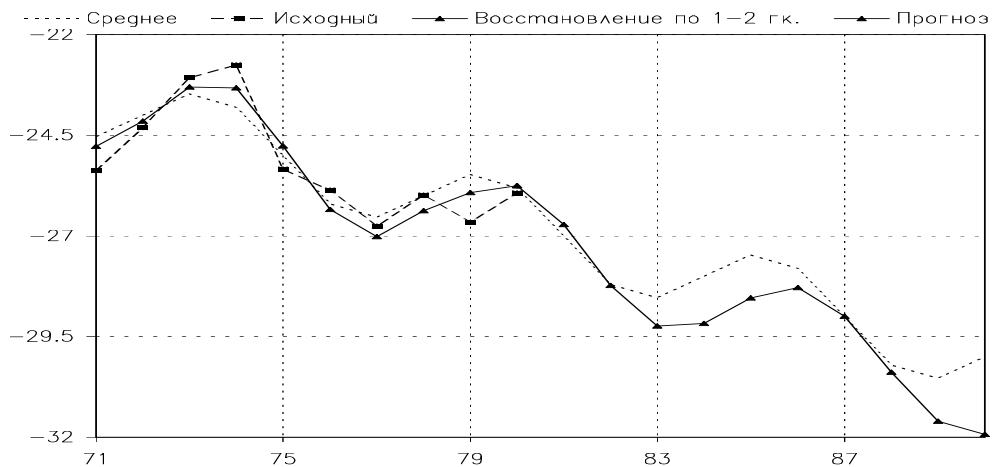


Рис. 12. Асимптотическая сходимость прогноза. Пример 12.1, $N = 30, \tau = 15$.

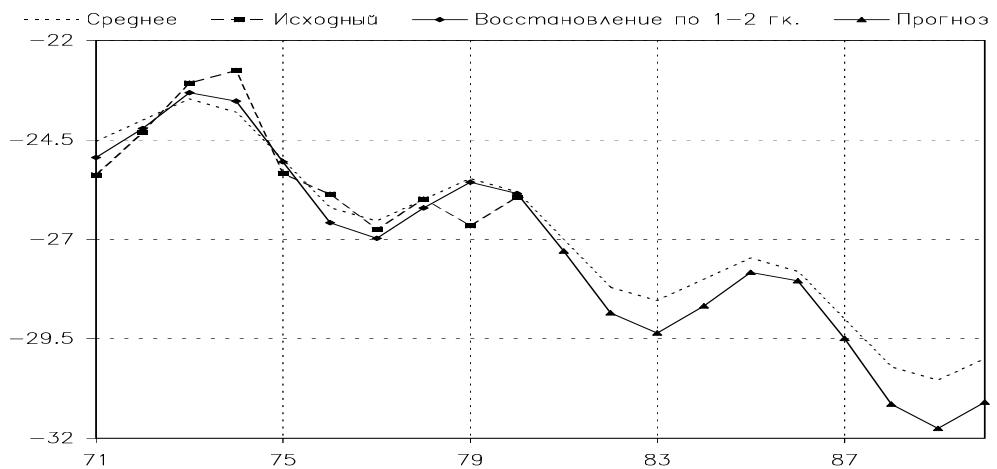


Рис. 13. Асимптотическая сходимость прогноза. Пример 12.1, $N = 80, \tau = 40$.

составлен из первых $(\tau - 1)$ -й компонент $(n + 1)$ -ой точки центрированной нормированной многомерной выборки порожденной рядом $(f_i)_{i=1}^{N+1}$. Пусть также

$$\mathbf{u} = (u_\tau^{(i_1)}, u_\tau^{(i_2)}, \dots, u_\tau^{(i_r)}).$$

Определение 13.1 Обобщенным решением системы (39) назовем решение системы

$$U_*^T U_* \bar{l} = U_*^T Q. \quad (40)$$

Определение 13.2 Решение уравнения

$$\frac{z^* - \frac{1}{N-\tau+2} \left(\sum_{s=\tau}^N f_s + z^* \right)}{\sqrt{\frac{1}{N-\tau+2} \left\{ (z^*)^2 + \sum_{s=\tau}^N f_s^2 \right\} - \left\{ \frac{1}{N-\tau+2} \left(z^* + \sum_{s=\tau}^N f_s \right) \right\}^2}} = \mathbf{u} \bar{l}_*, \quad (41)$$

где \bar{l}_* — решение системы (40), назовем обобщенным корреляционным продолжением рассматриваемой функции дискретного аргумента.

Естественно, как для ковариационного алгоритма прогноза, так и для корреляционного можно рассматривать прогнозирование слагаемых в сумме, изучать асимптотические свойства прогноза и прогнозировать временные ряды со случайной ошибкой. Однако эта тема слишком обширна и не будет здесь рассматриваться.

14 Примеры работы метода на реальных данных.

Приводимые в этом разделе примеры прогнозов реальных временных рядов носят в большой степени иллюстративный характер. Этим обусловлен и выбор конкретных наборов данных, значительная часть которых заимствована из классических работ. Для иллюстрации хорошей работы описываемых алгоритмов на реальных данных применен следующий подход: рассматривается лишь часть имеющегося временного ряда (первые несколько точек) и на основе анализа поведения ряда на этом участке производится его прогнозирование. Полученный прогноз затем сравнивается с реальным поведением ряда (либо с результатом сглаживания этого ряда с помощью гусеницы).

Пример 14.1 В качестве первого примера рассмотрим временной ряд, представляющий собой среднее число яиц на несушку на каждый месяц в США за 1938, 1939 и 1940 гг (см. [7], гл. 45, табл. 45.3.). Исходный временной ряд изображен на Рис. 14. На Рис. 15 можно видеть результаты гусеничного прогноза для этого ряда: по первым 24-м точкам спрогнозированы следующие 12 (длина гусеницы равнялась 12, прогноз произведен по первым 9 ГК.). На том же рисунке построены и соответствующие наблюденные данные. Сравнение наблюденных и предсказанных величин говорит о хорошем качестве прогноза.

Пример 14.2 Вторым примером послужит временной ряд описывающий заболеваемость СПИДом в США с 1984 по 1988 года, 1 точка — 3 мес. Данные приведены в [18], Таблица I. График исходных данных представлен на Рис. 16.

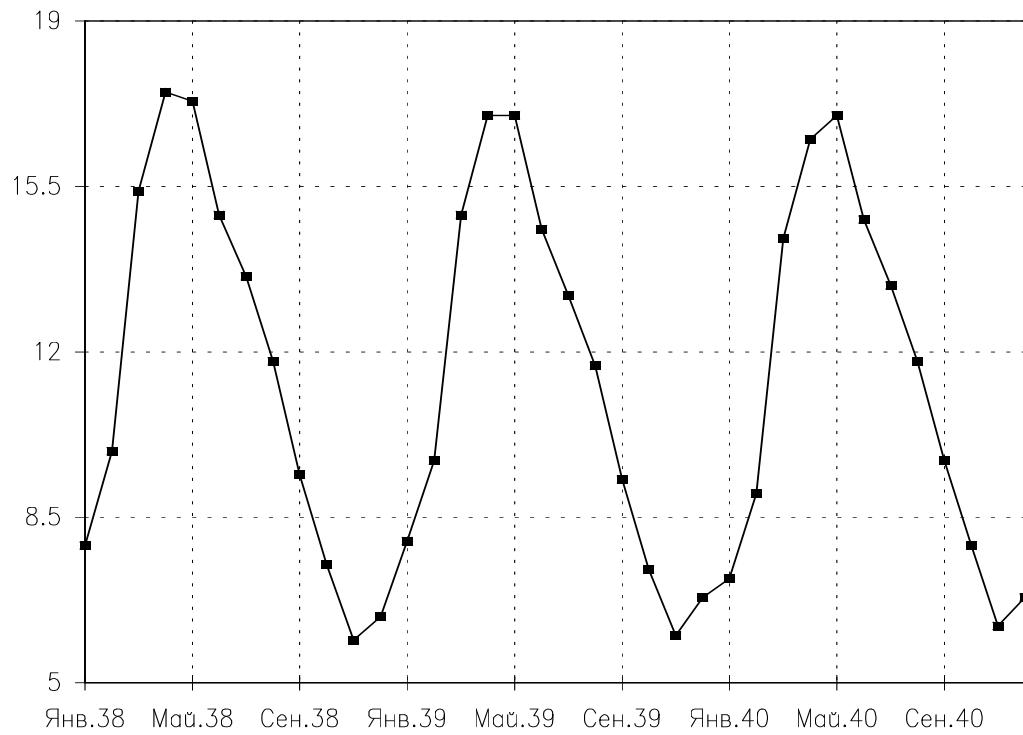


Рис. 14. Среднее число яиц на несушку: исходные данные.

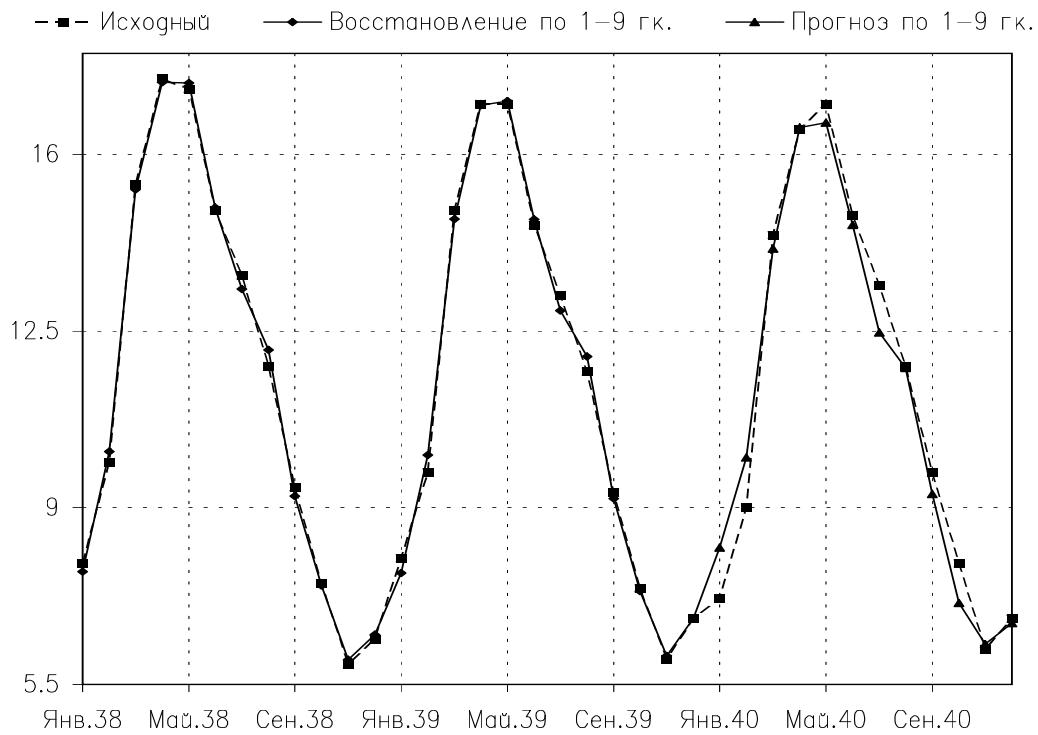


Рис. 15. Среднее число яиц на несушку: гусеничное сглаживание по 24-м точкам и прогноз на 12 точек вперед.

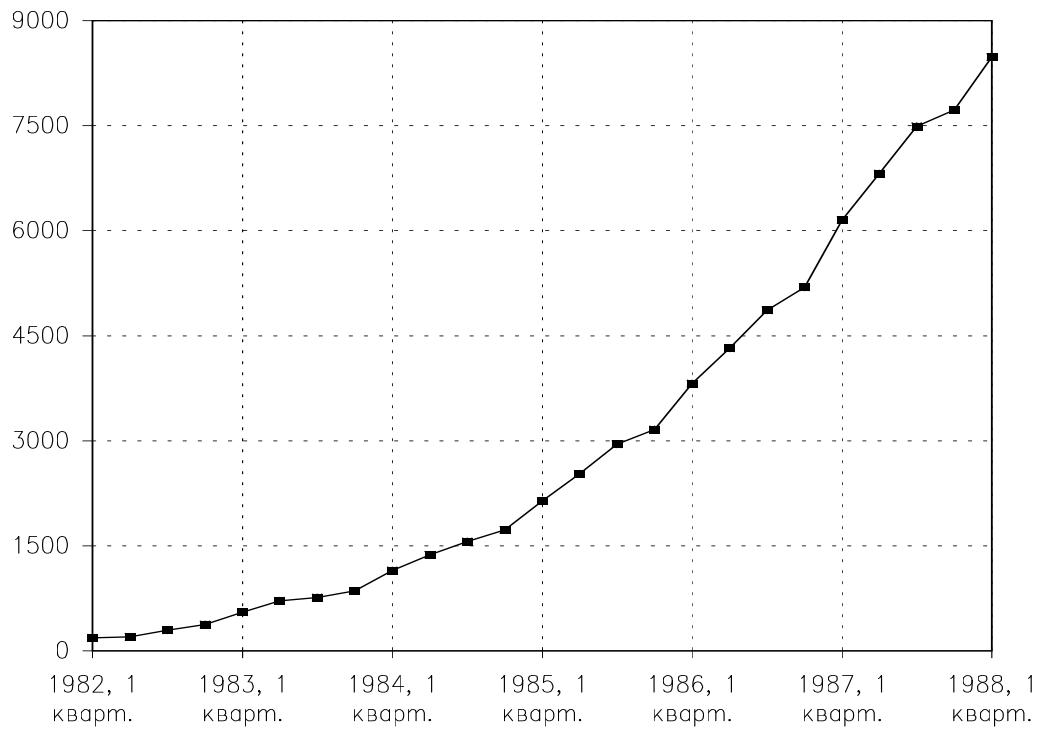


Рис. 16. Заболеваемость СПИДом в США: исходные данные.

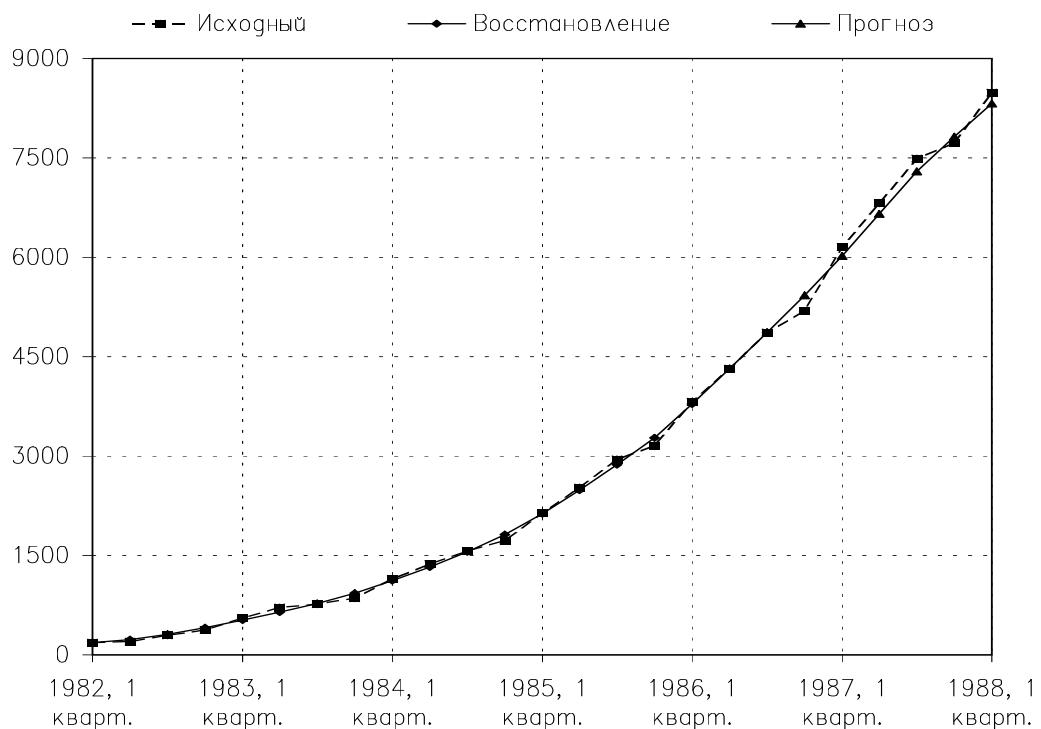


Рис. 17. Заболеваемость СПИДом в США: гусеничное сглаживание по 20-ти точкам и прогноз на 6 точек вперед.

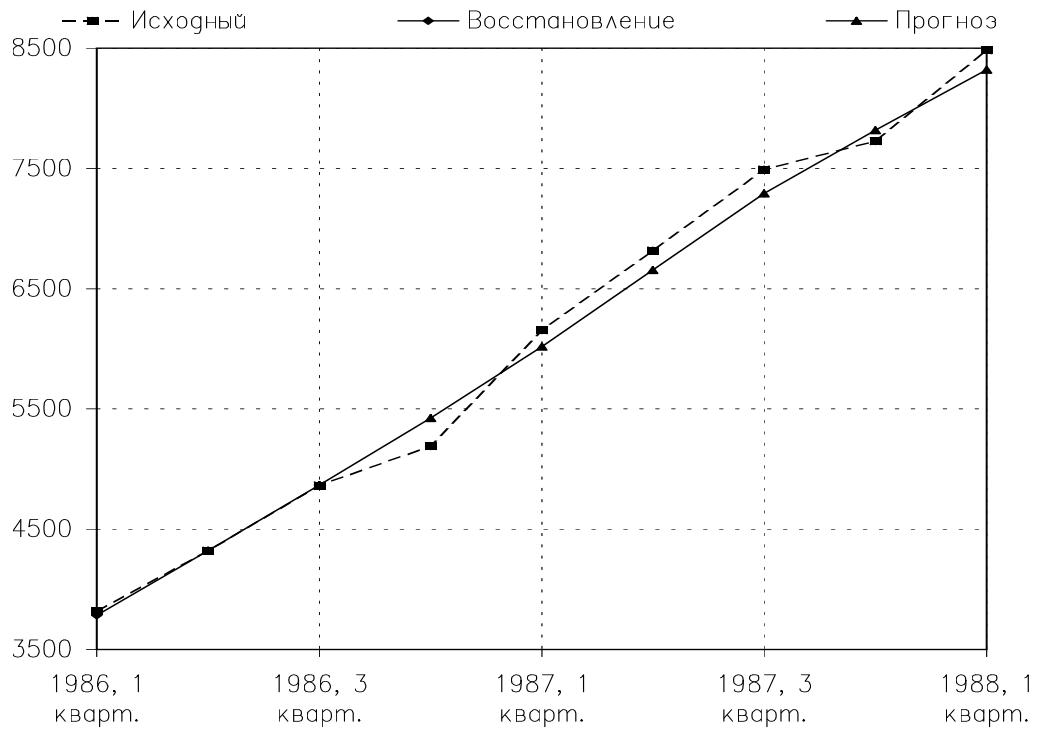


Рис. 18. Заболеваемость СПИДом в США: гусеничное сглаживание по 20-и точкам и прогноз на 6 точек вперед. Изображены последние 8 точек.

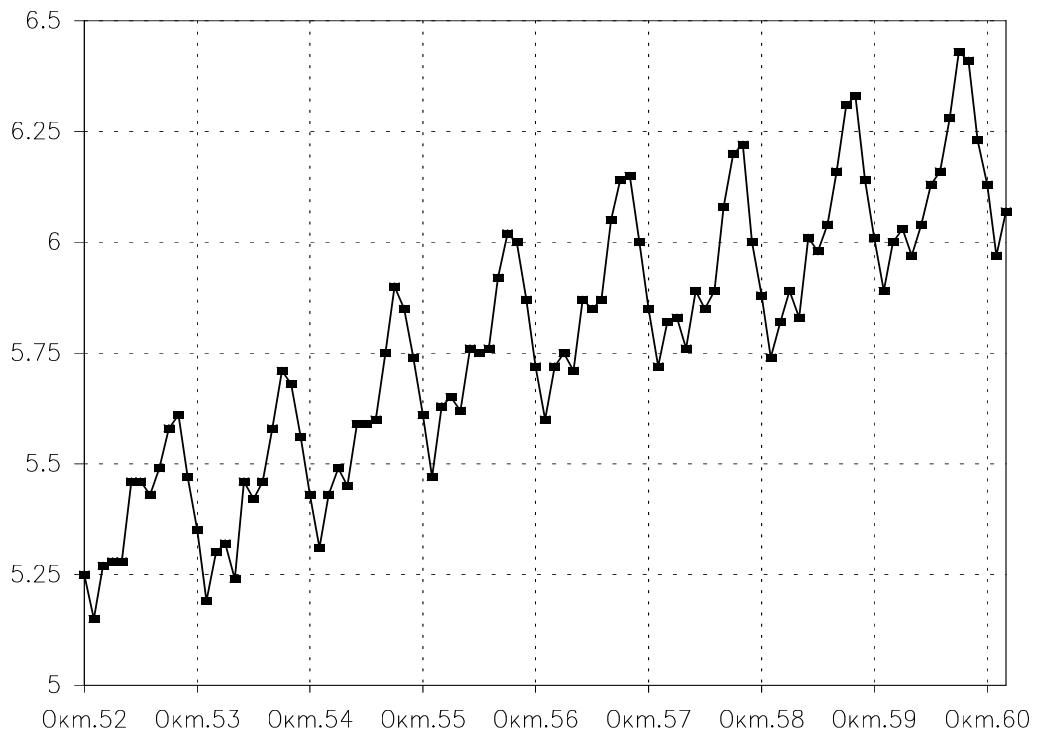


Рис. 19. Объем пассажироперевозок на международных авиалиниях. Исходные данные, ряд прологарифмирован.

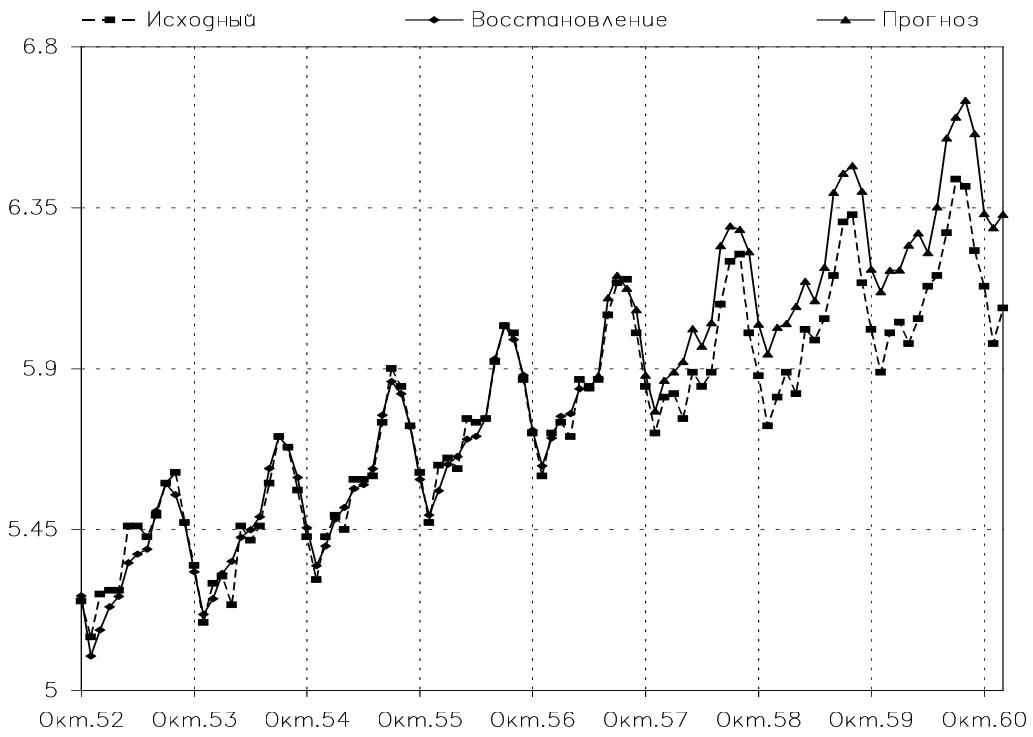


Рис. 20. Объем пассажироперевозок на международных авиалиниях.
Исходный ряд, восстановление и прогноз.

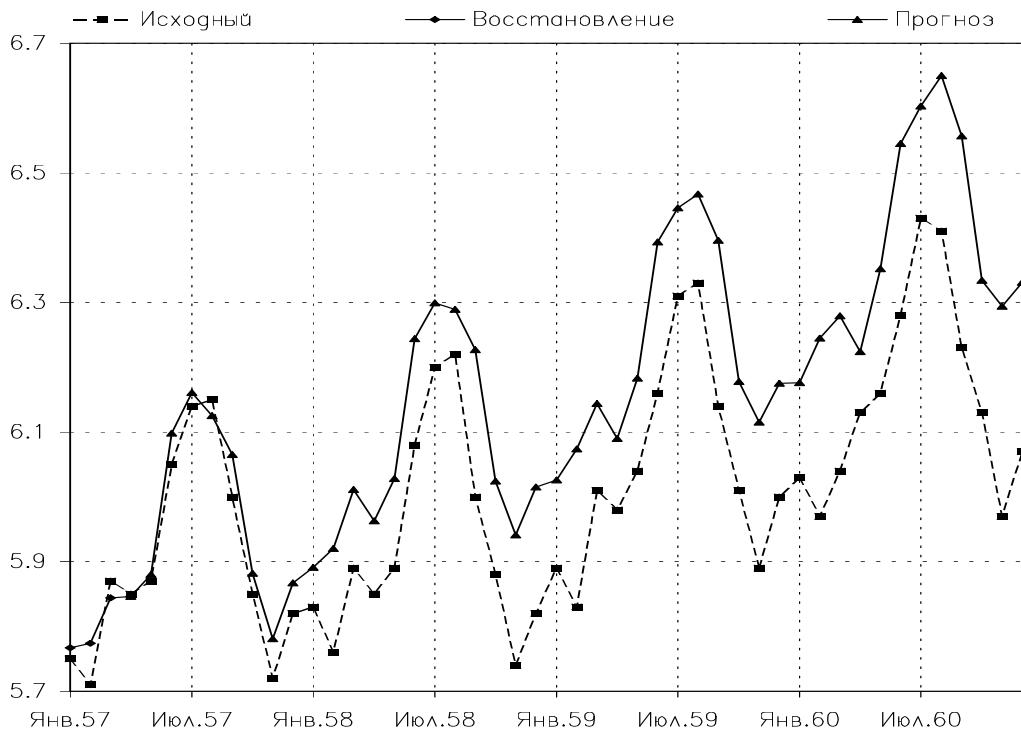


Рис. 21. Объем пассажироперевозок на международных авиалиниях.
Исходный ряд, восстановление и прогноз (масштаб увеличен).

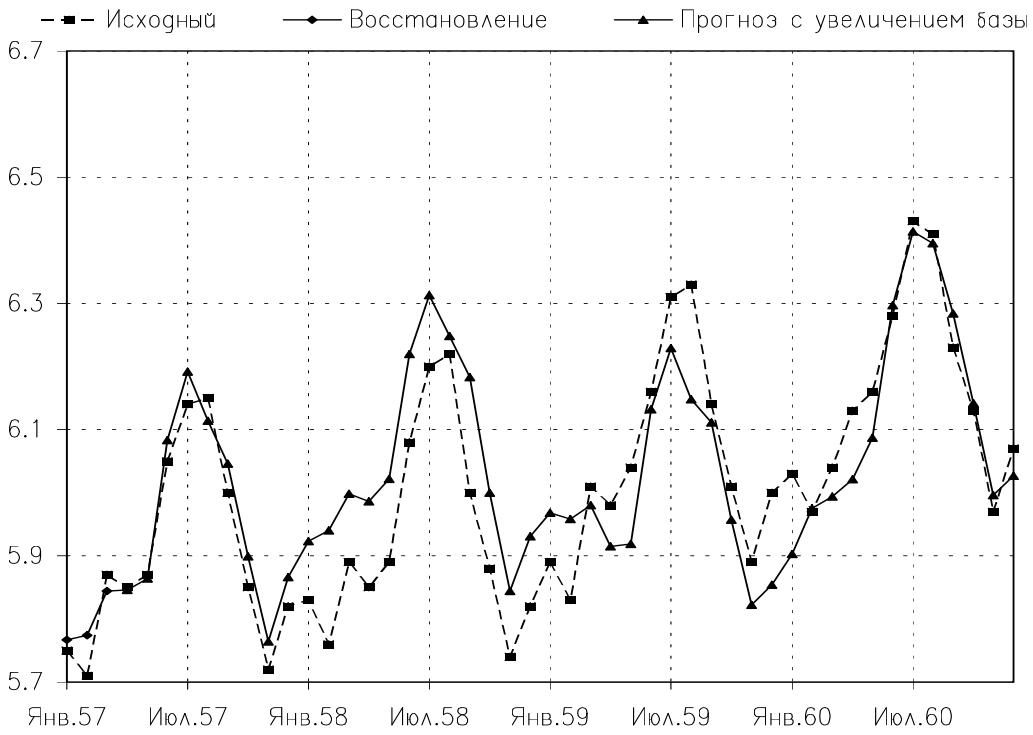


Рис. 22. Объем пассажироперевозок на международных авиалиниях.
Исходный ряд, восстановление и прогноз с увеличением базы.

На Рис. 17 вновь можно видеть результаты гусеничного прогноза для этого ряда: по первым 20-ти точкам спрогнозированы следующие 6 (длина гусеницы 10, прогноз произведен по первым 2-м ГК.). На Рис. 18 изображены последние 8 точек ряда совместно с восстановлением и прогнозом.

Пример 14.3 Следующим пробным рядом, широко использующимся во многих книгах по статистическому анализу временных рядов (см. например [3], гл. 9.) являются данные описывающие количества перевезенных пассажиров на международных авиалиниях с января 1949 по декабрь 1960. Каждая точка – 1 мес., числа – в тыс. человек. Для обработки исходный ряд логарифмировался. Логарифм исходного ряда представлен на Рис. 19.

На Рис. 20, 21 изображен прогноз по участку с октября 1952 по апрель 1957 на 42 точки вперед (по декабрь 1960), на основе первых 8 ГК. Легко заметить, что форма прогнозируемого ряда практически полностью воспроизводит форму исходного, однако, в тренде имеется некоторое расхождение (прогноз приподнят.)

Отметим, что на практике, как правило, прогнозируемое значение корректируется по мере поступления новых данных. Логично предположить, что увеличение интервала, на основе которого производится прогноз, приводит к лучшим результатам (см. например разделы 10, 12). Наблюдается ли подобное явление на реальных данных? Предположим, что мы хотим предсказать объем пассажироперевозок на полгода вперед. Рассмотрим следующую схему прогнозирования. На основе участка с октября 1952 по апрель 1957 проведем прогноз на 6 точек вперед (6 месяцев). Эти 6 точек нанесем на график. Затем увеличим базовый интервал на 1 точку (до мая 1957) и снова произведем прогноз на 6

месяцев вперед. На сей раз только последняя точка попадет на график. Затем еще увеличим базовый интервал и т.д. Результаты применения такого прогноза приведены на Рис. 22. По сравнению с Рис. 21 прогнозируемые значения лежат гораздо плотнее к наблюдаемым. Более того, последние прогнозируемые значения практически точно совпадают с наблюдаемыми, чего не наблюдалось на Рис. 21.

15 Заключение.

В статье авторами сделан основной упор на алгоритмическую сторону проблемы, и изложенные алгоритмы представляют результат развития лишь одного (наиболее продвинутого) из возможных подходов к задаче прогнозирования временных рядов с помощью метода "гусеница". За пределами статьи остались как ряд теоретических вопросов, так и некоторые практические (правильный выбор главных компонент, надежность и устойчивость прогноза и т.д.) Кроме того, некоторая часть результатов (особенно это касается раздела 9) носят предварительный характер и не являются окончательно завершенными.

Литература

- [1] Белонин, М.Д., Голубева, В.А., Скублов, Г.Т., (1982), *Факторный анализ в геологии*. М., Недра.
- [2] Белонин, М.Д., Татаринов, И.В., Калинин, О.М., Шиманский, В.К., Бескровная, О.В., Гранский, В.В., Похитонова, Т.Е., (1971), *Факторный анализ в нефтяной геологии*. М., ВИЭМС.
- [3] Д. Бокс, Д. Дженкинс, (1970), *Анализ временных рядов, прогноз и управление*. М., Наука.
- [4] Ветошева, В.И., Солнцев, В.Н., (1990), Использование многомерных методов анализа для классификации вызванных потенциалов условной негативной волны. *Вестник Ленинградского университета*, сер. 3, вып. 2, N 10, 54-59.
- [5] Гельфонд, А.О., (1967), *Исчисление конечных разностей*. М., Наука.
- [6] Данилов, Д.Л., (1997), Применение метода "Гусеница" для прогнозирования временных рядов. Элементы операторной теории. *Настоящий сборник*.
- [7] Кендалл, М.Дж., Стьюарт, А., (1976), *Многомерный статистический анализ и временные ряды*. М., Наука.
- [8] Кислицин, М.М., (1978), Биотехнические системы в авиационной эргономике. *Вопросы кибернетики*. **51**, М.
- [9] Кислицин, М.М., (1978), Исследование функционального состояния пилотов в процессе летного труда методом многомерной статистики. *Вопросы кибернетики*. **51**, М.
- [10] Кислицин, М.М., (1978), Многомерная статистика временных рядов наблюдений в авиационной эргономике. *Вопросы кибернетики*. **51**, М.
- [11] Марпл-мл., С.Л., (1990), *Цифровой спектральный анализ и его приложения*. М., Мир.

- [12] Некруткин., В.В., (1997), Разложения временных рядов. *Настоящий сборник*.
- [13] Рао, С.Р., (1968), *Линейные статистические методы и их применение*. М., Наука.
- [14] Соболев, С.Л., (1974), *Введение в теорию кубатурных формул*. М., Наука.
- [15] Buchstaber, V.M., (1994), Time Series Analysis and Grassmannians. *Amer. Math. Soc. Transl.* (2), **162**.
- [16] Danilov, D.L., (1996), Principal Components in Time Series Forecast. *1996 Proceedings of the Statistical Computing Section of the American Statistical Association*, 156-160.
- [17] Danilov, D., Solnsev, V.N., Zhigljavsky, A., (1996) Analysis and forecast of time series on the base of the principal components. *Abstracts of Second International Conference on Computing in Economics and Finance* Geneva, Switzerland, 26-28 June 1996.
- [18] Rosenberg, P.S., Gail, M.H., (1991), Backcalculation of flexible linear models of the Human Immunodeficiency Virus infection curve. *Applied Statistics*, 40, 269-282.

III.2 Обработка многомерных временных рядов с помощью метода “Гусеница”

Н.Э. Голяндина Д.Л. Данилов

1 Введение

В статье излагается подход к обработке многомерных временных рядов с помощью алгоритма “Гусеница”. Напомним, что идея метода одномерной гусеницы состоит в переходе от одномерного ряда к многомерной выборке и исследовании свойств этой выборки методами многомерной статистики. Однако эта идея продолжает работать также и в случае, когда исходно рассматривается не один временной ряд, а система. В этом случае в многомерную выборку преобразуется многомерный же временной ряд. Схема метода и основные обозначения приведены в разделе 2. В разделе 3 приводятся некоторые вспомогательные сведения из теории конечно-разностных уравнений, в разделе 4 с помощью достаточно простых рассуждений удается выяснить, какого рода сходства и различия в поведении многомерных временных рядов обнаруживаются этим алгоритмом. Другой вариант многомерной “Гусеницы” предложен в разделе 5. Алгоритм прогнозирования системы временных рядов предложен в разделе 6. Примеры применения алгоритмов приведены в разделе 7.

2 Схема метода “Гусеница” и основные определения.

Для начала опишем общую схему метода. Будем считать, что наблюдается система функций дискретного аргумента $\{(f_i^{(k)})_{i=1}^{N_k}\}$, где $k = 1, \dots, s$. Параметр s , таким образом, имеет смысл количества одномерных числовых последовательностей (либо размерности многомерной числовой последовательности), а N_k — количества элементов в k -ой последовательности. Целью статистической обработки такой системы числовых последовательностей мы ставим обнаружение общих (для всех или для части) закономерностей в их поведении. Статистика может интересовать, присутствует ли в этих рядах общая колебательная компонента, тренд и, если присутствует, то какого характера. После такого предварительного изучения обычно требуется произвести выделение той или иной интересующей нас компоненты, выяснение того, какие из исходных рядов вносят вклад в эту компоненту и др.

Опишем предлагаемый алгоритм.

Будем считать, что $N_k > 2$ для любого k . Выберем n такое, что $0 < n \leq N_k - 1$ для любого k . Это число будем далее называть временем жизни многомерной гусеницы. Для каждого k вычислим $\tau_k = N_k - n + 1$ — длину гусеницы для соответствующего одномерного ряда. Построим последовательность из n векторов в \mathbb{R}^τ , $\tau = \sum_{k=1}^s \tau_k$, следующего вида:

$$\mathcal{X}^{(l)} \in \mathbb{R}^\tau, \quad \mathcal{X}^{(l)} = (X^{(l,1)}, \dots, X^{(l,s)})^T,$$

где $X^{(l,k)} = (x_i^{(l,k)})_{i=1}^{\tau_k}$, и $x_i^{(l,k)} = f_{i+l-1}^{(k)}$. Обозначим

$$\mathbf{X} = (\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(n)}). \quad (1)$$

Будем называть \mathbf{X} нецентрированной матрицей наблюдений, порожденной гусеницей со временем жизни n .

Заметим, что

1. Матрица $\mathbf{X} = (\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(n)})$ может быть записана как

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \vdots \\ \mathbf{X}^{(s)} \end{pmatrix},$$

где $\mathbf{X}^{(k)} = ((X^{(1,k)})^T, \dots, (X^{(n,k)})^T)$.

2. Подматрица $\mathbf{X}^{(k)}$ есть в точности нецентрированная матрица наблюдений, порожденная одномерной гусеницей с длиной τ_k , примененной к k -му ряду. Говоря неформально, подматрицы $\mathbf{X}^{(k)}$ в \mathbf{X} расположены “параллельно”.

Через $\bar{\mathbf{X}}$ условимся обозначать вектор гусеничного среднего

$$\bar{\mathbf{X}} = \begin{pmatrix} \bar{\mathbf{X}}^{(1)} \\ \vdots \\ \bar{\mathbf{X}}^{(s)} \end{pmatrix},$$

где $\bar{\mathbf{X}}^{(k)} = \frac{1}{n} \sum_{i=1}^n X^{(i,k)}$ есть в точности вектор гусеничного среднего для $\mathbf{X}^{(k)}$. Вычитом гусеничного среднего из каждого вектора, образующего нецентрированную матрицу наблюдений, мы получим центрированную матрицу наблюдений, которую будем обозначать через \mathbf{Y} , $\mathbf{Y} = \mathbf{X} - \bar{\mathbf{X}}$.

Проводимый в дальнейшем анализ главных компонент может проводиться как по центрированной, так и по нецентрированной выборкам. Для упрощения выкладок рассмотрим простейший нецентрированный вариант. Применим к \mathbf{X} нецентрированный метод главных компонент. Вычислим нецентральную ковариационную матрицу C , ее систему собственных векторов

$$V = (v^{(1)}, \dots, v^{(\tau)}), \quad (v^{(i)})^T v^{(j)} = \delta_{ij}, \quad C = V \Lambda V^T, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_\tau),$$

(здесь $v^{(i)}$ представляет собой i -ый собственный вектор, а Λ — диагональная матрица собственных чисел) и систему главных компонент:

$$\Phi = V^T \mathbf{X}, \quad \Phi = (\Phi^{(1)}, \dots, \Phi^{(\tau)})^T, \quad \Phi^{(i)} = (\phi_1^{(i)}, \dots, \phi_n^{(i)})^T.$$

После проведения анализа главных компонент обычно предполагается проведение операции восстановления исходной матрицы наблюдений по некоторому поднабору главных компонент, т. е. для $\tilde{V} = (v^{(i_1)}, \dots, v^{(i_r)})$ и $\tilde{\Phi} = \tilde{V}^T \mathbf{X}$ вычисляется матрица $\tilde{\mathbf{X}} = \tilde{V} \tilde{\Phi}$. Указанным способом может быть восстановлена как нецентрированная матрица наблюдений, так и центрированная. (При выделении периодических функций чаще используется первый вариант, в то время как для выделения тренда часто лучше производить анализ по центрированной

выборке.) После получения $\tilde{\mathbf{X}}$ (либо $\tilde{\mathbf{Y}}$) исходные последовательности восстанавливаются усреднением по побочным диагоналям подматриц $\mathbf{X}^{(k)}$, т. е., если $\tau_k \leq n$, то

$$\tilde{f}_m^{(k)} = \begin{cases} \frac{1}{m} \sum_{i=1}^m x_i^{(m-i+1,k)} & 1 \leq m \leq \tau_k, \\ \frac{1}{\tau_k} \sum_{i=1}^{\tau_k} x_i^{(m-i+1,k)} & \tau_k \leq m \leq n, \\ \frac{1}{N_k - m + 1} \sum_{i=1}^{N_k - m + 1} x_{i+m-n}^{(n-i+1,k)} & n \leq m \leq N_k. \end{cases}$$

Как и в одномерной "гусенице", операция получения главных компонент есть изображение исходной выборки в базисе, составленном из выбранных собственных векторов, а восстановление — проектирование исходной выборки на гиперплоскость, порожденную выбранным набором собственных векторов ковариационной матрицы. Однако для многомерной гусеницы выбор одной главной компоненты приводит, вообще говоря, к восстановлению всей системы рядов. Заметим, что в рассматриваемом методе главные компоненты действительно являются общими для всей системы рядов, в то время как собственные вектора состоят из частей длины τ_k , соответствующих отдельным рядам. Это связано с тем, что ковариационная матрица "гусеницы" для системы рядов составлена из блоков, представляющих собой кросс-ковариации многомерных выборок, порожденных каждым из рядов по отдельности.

В теории одномерной гусеницы большое значение играют функции дискретного аргумента, порождающие многомерные выборки, лежащие в гиперплоскостях размерности меньшей, чем τ . Такие функции принято называть функциями конечного ранга (см. раздел III). Переход к многомерному случаю вносит ряд принципиально новых моментов в это явление, что и будет подробно рассмотрено далее. По аналогии со случаем одномерной "гусеницы" введем несколько определений (см. также раздел IV.1).

Определение 2.1 Будем говорить, что система числовых рядов $(f_i^{(k)})_{i=1}^{N_k}$ имеет ранг r , если порождаемая им при гусеничной обработке (с параметром n таким, что $\tau > r$) выборка лежит в гиперплоскости размерности r .

Замечание 2.1 Легко понять, что если нецентрированная матрица наблюдений лежит в гиперплоскости конечной размерности, то и центрированная матрица тоже лежит в некоторой гиперплоскости. Однако размерность последней может быть меньше.

3 Некоторые сведения о конечно-разностных уравнениях с постоянными коэффициентами.

Рассмотрим однородное конечно-разностное уравнение порядка r :

$$f_{t+r} + a_1 f_{t+r-1} + \dots + a_r f_t = 0, \quad t \in \mathbb{Z}. \quad (2)$$

Известно (см. [1]), что вид частных решений такого уравнения определяется видом корней его характеристического полинома:

$$\lambda^r + a_1\lambda^{r-1} + \dots + a_r = 0. \quad (3)$$

Пусть $\lambda_1, \dots, \lambda_m$ — корни этого полинома ($\lambda_i \in \mathbb{C}$), p_1, \dots, p_m — их кратности, и $\sum_{l=1}^m p_l = r$. В этом случае частные решения уравнения (2) легко выписываются:

$$\begin{array}{cccc} \lambda_1^t t^{p_1-1} & \lambda_1^t t^{p_1-2} & \dots & \lambda_1^t \\ \dots & \dots & \dots & \dots \\ \lambda_m^t t^{p_m-1} & \lambda_m^t t^{p_m-2} & \dots & \lambda_m^t \end{array} \quad (4)$$

Общее решение уравнения (2) представляет собой сумму вида

$$\sum_{l=1}^m \sum_{u=0}^{p_l-1} \lambda_l^t t^u A_{l,u}, \quad (5)$$

где, вообще говоря, комплексные коэффициенты $A_{l,u}$ выбираются исходя из соответствия начальным (граничным) условиям. Заметим, что, несмотря на наличие комплексных λ_l , общее решение вещественно, поскольку вещественны коэффициенты (3) и, следовательно, комплексные корни у этого полинома могут встречаться только в сопряженных парах (см. [1]).

Фиксируем некоторое частное решение $f(t) = B\lambda^t t^u$, где $u \leq p_l - 1$, и изучим структуру нецентрированной матрицы наблюдений такой числовой последовательности, порожденную одномерной гусеницей длины τ . Этую матрицу наблюдений обозначим, для определенности, $\mathcal{Z} = (Z^{(1)}, \dots, Z^{(n)})$. Непосредственно проверяется, что

$$f(t+s) = B\lambda^{t+s} (t+s)^u = \sum_{j=0}^u \phi_j(t) \psi_j(s),$$

где $\phi_j(t) = AC_u^j \lambda^t t^j$ и $\psi_j(s) = \lambda^s s^{u-j}$. Рассмотрим набор векторов вида

$$Y_j = \begin{pmatrix} \psi_j(0) \\ \psi_j(1) \\ \dots \\ \psi_j(\tau-1) \end{pmatrix} \quad \text{для } j = 0, \dots, u.$$

Эти векторы являются линейно независимыми. Непосредственно проверяется, что любой вектор рассматриваемой многомерной выборки представляет собой линейную комбинацию Y_j следующего вида:

$$Z^{(q)} = \lambda^q \sum_{j=0}^u BC_u^j q^j Y_j$$

Рассмотрим теперь решение (5). Оно представляет собой сумму функций рассмотренного вида $B\lambda^t t^u$ с конкретными λ и B . Суммируя, получаем

$$X^{(q)} = \sum_{l=1}^m \sum_{u=1}^{p_l-1} \lambda_l^q \sum_{j=0}^u A_{l,u} C_u^j q^j Y_j^{(l)} =$$

$$= \sum_{l=1}^m \lambda_l^q \sum_{u=0}^{p_l-1} \sum_{j=0}^u A_{l,u} C_u^j q^j Y_j^{(l)} = \sum_{l=1}^m \lambda_l^q \sum_{i=0}^{p_l-1} q^i e_{l,i}, \quad (6)$$

где $e_{l,i}$ — линейные комбинации векторов $Y_j^{(l)}$, представляющие собой набор образующих гиперплоскости, содержащей рассматриваемую многомерную выборку. Заметим, что для конкретного решения и фиксированного τ набор этих векторов постоянен и не зависит от длины ряда (конечно, подразумевается, что длина ряда достаточно велика по сравнению с r). Кроме того, необходимо особо отметить, что если в конкретное решение данного конечно-разностного уравнения не войдет частное решение вида $\lambda_l^t t^u$, то и соответствующий вектор $e_{l,i}$ в (6) будет нулевым, а размерность выборки соответственно уменьшится. Однако в дальнейшем мы исключим из рассмотрения такие ситуации, считая, что в этом случае мы уменьшили порядок соответствующего конечно-разностного уравнения, т. е. для данного конкретного ряда чисел рассматриваем порождающее его конечно-разностное уравнение наименьшего порядка.

4 О связи размерности многомерной матрицы данных с корнями характеристических полиномов конечно-разностных уравнений, порождающих числовые ряды.

Вернемся к системе числовых рядов. Пусть имеется система числовых рядов $\{(f_i^{(k)})_{i=1}^{N_k}, k = 1, \dots, s\}$, каждый из которых является рядом конечного ранга т. е.

$$\begin{aligned} f_{t+r_1}^{(1)} + a_1^{(1)} f_{t+r_1-1}^{(1)} + \dots + a_{r_1}^{(1)} f_t^{(1)} &= 0 \\ \dots &\dots \dots \dots \dots \\ f_{t+r_s}^{(s)} + a_1^{(s)} f_{t+r_s-1}^{(s)} + \dots + a_{r_s}^{(s)} f_t^{(s)} &= 0 \end{aligned} \quad (7)$$

Исследуем вопрос о структуре многомерной матрицы данных \mathbf{X} . Будем обозначать через $\lambda_l^{(k)}$ l -е характеристическое число k -го ряда, m_k — количество различных характеристических чисел k -го уравнения и $p_l^{(k)}$ — кратность $\lambda_l^{(k)}$. Для каждой подматрицы $\mathbf{X}^{(k)}$ проведем разложение вида (6). Получим следующее представление $X^{(q,k)}$:

$$X^{(q,k)} = \sum_{l=0}^{m_k} (\lambda_l^{(k)})^q \sum_{i=0}^{p_l^{(k)}-1} q^i e_{l,i}^{(k)}. \quad (8)$$

Обозначая через $E_{l,i}^{(k)}$ вектор размерности τ , у которого компоненты, соответствующие подматрице $\mathbf{X}^{(k)}$, совпадают с компонентами $e_{l,i}^{(k)}$, а остальные равны 0, получим для величины

$$\mathcal{X}^{(q)} = \begin{pmatrix} X^{(q,1)} \\ \vdots \\ X^{(q,s)} \end{pmatrix}$$

следующее выражение

$$\mathcal{X}^{(q)} = \sum_{k=1}^s \sum_{l=0}^{m_k} (\lambda_l^{(k)})^q \sum_{i=0}^{p_l^{(k)}-1} q^i E_{l,i}^{(k)}, \quad (9)$$

где суммирование ведется по всем линейно независимым решениям. Представление (9) фактически является параметрической формой записи последовательности точек гиперплоскости, содержащей данные.

Сформулируем несколько непосредственных выводов.

1. Многомерная матрица \mathbf{X} также, как и все ее подматрицы, лежит в гиперплоскости, размерности меньшей τ .
2. Размерность гиперплоскости и ее расположение определяются видом конечно-разностных уравнений и не зависят от количества элементов в рядах.
3. $\dim(\mathbf{X}) < \tau \leq r = \sum_{k=1}^s r_k$.

Исследуем вопрос об обнаружении общих составляющих системы числовых рядов. Пусть в каждом из s рядов присутствует ровно r_k линейно независимых решений и среди характеристических чисел разных рядов нет совпадающих. Тогда размерность гиперплоскости, содержащей \mathbf{X} , равна в точности r . (Это непосредственно следует из разложения (9)). Пусть теперь имеется \tilde{k} рядов, имеющих слагаемое с совпадающим характеристическим числом (для определенности равным $\tilde{\lambda}$). Тогда слагаемые в (9), соответствующие сомножителям $\tilde{\lambda}^t t^l$, можно сгруппировать. При этом вместо \tilde{k} векторов $E_{l,i}^{(k)}$, соответствующих различным l , можно записать один, являющийся линейной комбинацией этих $E_{l,i}^{(k)}$. В этом случае многомерная матрица наблюдений будет лежать в гиперплоскости размерности меньшей, чем τ и равной $\tau - \tilde{k} + 1$. Это означает, что в разложении по системе собственных векторов нецентральной ковариационной матрицы такой системы рядов меньшее число собственных чисел будет отлично от нуля. Поскольку собственные векторы нецентральной ковариационной матрицы также образуют базис гиперплоскости, то они будут представлять собой линейные комбинации $E_{l,i}^{(k)}$. Наличие же решений с совпадающими характеристическими числами λ повлечет существование главных компонент, “общих” для некоторой части рядов.

5 О возможной модификации метода многомерной “Гусеницы”

Полученная в предыдущем разделе интерпретация метода позволяет предложить некоторую модификацию этого метода. Этот новый метод обработки систем временных рядов использует тот же принцип, что и ранее описанный, однако в вычислительном плане имеет меньше недостатков. Действительно, рассмотрим ситуацию, когда среди рядов $f^{(k)}$ имеется ряд, чья длина много меньше, чем у остальных. Тогда для обеспечения условия $n = N_k - \tau_k + 1$ величина n должна быть относительно малой, а τ_k (у всех рядов, кроме маленького) — достаточно большой. Это влечет образование (и в дальнейшем обработку) ковариационных матриц большой размерности, что является серьезным препятствием в реализации метода, особенно в случае многих рядов.

Как видно из структуры собственных векторов и главных компонент метода, препятствие проистекает из желания выделять именно общие главные компоненты (условие $n = N_k - \tau_k + 1$). Опишем метод, позволяющий выделять общие собственные вектора системы рядов.

Вновь рассмотрим систему функций дискретного аргумента: $\{(f_i^{(k)})_{i=1}^{N_k}\}$ при $k = 1, \dots, s$. Зададим длину многомерной гусеницы τ , тогда $n_k = N_k - \tau + 1$ — время жизни гусеницы для соответствующего одномерного ряда. Положим $\mathbf{n} = \sum_{k=1}^s \mathbf{n}_k$ и построим соответствующую исходной системе рядов многомерную матрицу наблюдений \mathbf{X} размерности $\tau \times \mathbf{n}$, такую что

$$\mathbf{X} = (\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(s)}),$$

где $\mathbf{X}^{(k)}$ — $\tau \times n_k$ матрица наблюдений, порожденная k -ым одномерным рядом при длине гусеницы $\tau_k = \tau$. В отличие от ранее описанного метода, подматрицы наблюдений отдельных рядов расположены в \mathbf{X} последовательно, одна за другой.

Формализм, примененный в разделе 4 к предыдущему методу, может быть применен и здесь. Структура матрицы \mathbf{X} , однако, такова, что, вместо представления точки выборки линейной комбинацией базисных вектор-столбцов (см. (8)), разумнее произвести аналогичное разложение по базисным вектор-строкам. При этом общий вывод сохраняется: присутствие в составе рядов слагаемых одинакового вида влечет сокращение размерности гиперплоскости, содержащей данные. Интересным отличием метода является то, что условие согласования для \mathbf{X} гораздо слабее, чем в предыдущем случае (фактически требуется лишь выполнение $\tau \leq N_k$), что позволяет при обработке обходиться матрицами меньшей размерности. Это, конечно, не гарантирует, что при конкретной обработке не возникнет необходимости сделать τ достаточно большим.

6 Об алгоритмах прогнозирования многомерных рядов.

Вновь вернемся к рассмотрению наблюдаемой системы функций дискретного аргумента $\{(f_i^{(k)})_{i=1}^{N_k}\}$, где $k = 1, \dots, s$. Будем считать, что к рассмотренной системе рядов применена многомерная гусеница, описанная в разделе 2, с временем жизни n , в связи с чем для каждого одномерного ряда определена длина гусеницы $\tau_k = N_k - n + 1$. По аналогии со случаем одномерной "гусеницы" введем следующие определения:

Определение 6.1 Будем говорить, что система числовых рядов $(f_i^{(k)})_{i=1}^{N_k+1}$ является продолжением системы $(f_i^{(k)})_{i=1}^{N_k}$, если выборка, порожденная ей при гусеничной обработке с теми же τ_1, \dots, τ_k , лежит в той же гиперплоскости, что и у исходной системы рядов.

Опишем идею предлагаемого алгоритма экстраполяции системы функций дискретного аргумента. Пусть имеется система числовых рядов $\{(f_i^{(k)})_{i=1}^{N_k}\}$ $k = 1, \dots, s$. Рассмотрим базис гиперплоскости $v^{(1)}, \dots, v^{(r)}$, содержащей многомерную выборку, порожденную этим рядом. Не ограничивая общности, можно считать, что в качестве этого базиса выбрана система собственных векторов,

соответствующих отличным от нуля собственным числам матрицы вторых моментов многомерной выборки. Рассмотрим уравнение этой гиперплоскости, заданное в параметрической форме:

$$x(\bar{t}) = \sum_{i=1}^r t_i v^{(i)}, \quad (10)$$

при этом $\bar{t} = (t_1, \dots, t_r)$ — значение параметризации, соответствующее точке $x(\bar{t})$. Каждой точке многомерной выборки соответствует некоторое значение вектора параметризации \bar{t} , т. е.

$$X^{(1)} = x(\bar{t}^{(1)}), \dots, X^{(n)} = x(\bar{t}^{(n)}).$$

Пусть система числовых рядов $(f_i^{(k)})_{i=1}^{N_k+1}$, $k = 1, \dots, s$, есть продолжение системы $(f_i^{(k)})_{i=1}^{N_k}$. Фиксируем базис гиперплоскости $v^{(1)}, \dots, v^{(r)}$, вычисленный по системе $(f_i^{(k)})_{i=1}^{N_k}$. Рассмотрим выборку, порожденную системой $(f_i^{(k)})_{i=1}^{N_k+1}$, и разложим ее по базису $v^{(1)}, \dots, v^{(r)}$. В этом случае последней $(n+1)$ -й точке выборки также соответствует некоторое значение вектора параметризации $\bar{t}^{(n+1)}$. Если окажется возможным восстановить значение $\bar{t}^{(n+1)}$ только по набору $f_1^{(k)}, \dots, f_{N_k}^{(k)}$, $k = 1, \dots, s$, то все компоненты точки $X^{(n+1)}$ легко могут быть вычислены по формуле:

$$X^{(n+1)} = \sum_{i=1}^r \bar{t}_i^{(n+1)} v^{(i)},$$

откуда нетрудно перейти к исходному ряду. Указанная идея лежит в основе следующего алгоритма экстраполяции функции дискретного аргумента.

Алгоритм 6.1 Пусть имеется набор последовательных значений s функций дискретного аргумента, образующих систему ранга r , и количество элементов этих функций N_k , $k = 1, \dots, s$, достаточно велико.

1. В соответствии с (1) сформируем многомерную матрицу наблюдений.
2. Проведем построение матрицы вторых моментов .
3. Найдем базис $v^{(1)}, \dots, v^{(r)}$, соответствующий отличным от нуля собственным числам матрицы вторых моментов .
4. Запишем систему линейных алгебраических уравнений следующего вида:

$$\left\{ \begin{array}{ll} \sum_{j=1}^r h_j v_1^{(j)} & = f_{N_1-\tau_1+2}^{(1)} \\ & \dots \\ \sum_{j=1}^r h_j v_{\tau_1-1}^{(j)} & = f_{N_1}^{(1)} \\ \sum_{j=1}^r h_j v_{\tau_1+1}^{(j)} & = f_{N_2-\tau_2+2}^{(2)} \\ & \dots \\ \sum_{j=1}^r h_j v_{\tau_1+\tau_2-1}^{(j)} & = f_{N_2}^{(2)} \\ & \dots \\ \sum_{j=1}^r h_j v_{\tau_1+\tau_2+\dots+\tau_{s-1}+1}^{(j)} & = f_{N_s-\tau_s+2}^{(s)} \\ & \dots \\ \sum_{j=1}^r h_j v_{\tau_s}^{(j)} & = f_{N_s}^{(s)}. \end{array} \right. \quad (11)$$

5. Исследуем систему (11) на совместность.

5.1 Если система (11) несовместна, то исходный ряд не допускает продолжения.

5.2 Если система (11) имеет решение, то исходный ряд имеет продолжение, которое записывается как:

$$\begin{aligned} f_{N_1+1}^{*(1)} &= \sum_{j=1}^r h_j^* v_{\tau_1}^{(j)} \\ &\dots \\ f_{N_s+1}^{*(s)} &= \sum_{j=1}^r h_j^* v_{\tau}^{(j)}, \end{aligned}$$

где h_1^*, \dots, h_r^* — решение системы (11).

Естественно, возможно исследовать свойства этого алгоритма, выяснить условия существования продолжения и пр., однако в данной статье мы сосредотачиваемся исключительно на алгоритмической части и не будем развивать полную теорию.

Алгоритм, определяемый формулами (11) и (12), способен прогнозировать (т. е. строить продолжение) только системы конечного ранга. В случае бесконечного ранга система (11), вообще говоря, несовместна. Однако в практических ситуациях преимущественно встречаются именно такие системы рядов, например, ряды со случайной ошибкой.

Рассмотрим следующее обобщение алгоритма 6.1. Расширим понятие решения системы (11). Введем следующие обозначения. Рассмотрим набор индексов $0 < i_1 < \dots < i_r < \tau$. Определим

$$\mathbf{v} = \begin{pmatrix} v_{\tau_1}^{(i_1)} & v_{\tau_1}^{(i_2)} & \dots & v_{\tau_1}^{(i_r)} \\ v_{\tau_2}^{(i_1)} & v_{\tau_2}^{(i_2)} & \dots & v_{\tau_2}^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{\tau_s}^{(i_1)} & v_{\tau_s}^{(i_2)} & \dots & v_{\tau_s}^{(i_r)} \end{pmatrix}$$

и

$$\mathbf{V}_* = \begin{pmatrix} v_1^{(i_1)} & v_1^{(i_2)} & \dots & v_1^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{\tau_1-1}^{(i_1)} & v_{\tau_1-1}^{(i_2)} & \dots & v_{\tau_1-1}^{(i_r)} \\ v_{\tau_1+1}^{(i_1)} & v_{\tau_1+1}^{(i_2)} & \dots & v_{\tau_1+1}^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{\tau_1+\tau_2-1}^{(i_1)} & v_{\tau_1+\tau_2-1}^{(i_2)} & \dots & v_{\tau_1+\tau_2-1}^{(i_r)} \\ \vdots & \vdots & \vdots & \vdots \\ v_{\tau-1}^{(i_1)} & v_{\tau-1}^{(i_2)} & \dots & v_{\tau-1}^{(i_r)} \end{pmatrix}.$$

Также положим

$$Q = (q_i)_{i=1}^{\tau-s} = \left(f_{N_1-\tau_1+2}^{(1)}, \dots, f_{N_1}^{(1)}, f_{N_2-\tau_2+2}^{(2)}, \dots, f_{N_2}^{(2)}, \dots, f_{N_s-\tau_s+2}^{(s)}, \dots, f_{N_s}^{(s)} \right)^T$$

и

$$\bar{h} = (h_1, \dots, h_r)^T.$$

В этих обозначениях система (11) запишется как

$$\mathbf{V}_* \bar{h} = Q. \quad (12)$$

Мы уже упомянули, что такая система, вообще говоря, несовместна, хотя и существует широкий класс систем рядов, для которых она разрешима. Расширим понятие решения системы (12).

Определение 6.2 Обобщенным решением системы (12) назовем решение системы

$$\mathbf{V}_*^T \mathbf{V}_* \bar{h} = \mathbf{V}_*^T Q. \quad (13)$$

Определение 6.3 Величину

$$b = \mathbf{v} \bar{h}_*, \quad (14)$$

где \bar{h}_* — решение системы (13), назовем обобщенным продолжением рассматриваемой функции дискретного аргумента. Учитывая (13), можно записать для прогнозируемых значений системы в точках $N_k + 1$ следующую формулу:

$$\begin{pmatrix} f_{N_1+1}^* \\ \vdots \\ f_{N_s+1}^* \end{pmatrix} = \mathbf{v} (\mathbf{V}_*^T \mathbf{V}_*)^{-1} Q. \quad (15)$$

Из общей теории решения линейных систем (см., например, [2, 3]) следует, что общее решение (13) может быть записано в виде:

$$\bar{h}_* = C \mathbf{V}_*^T Q + (E - H) w,$$

где $H = C \mathbf{V}_*^T \mathbf{V}_*$, $C = (\mathbf{V}_*^T \mathbf{V}_*)^{-1}$ — обобщенная обратная к $\mathbf{V}_*^T \mathbf{V}_*$, а w — произвольный вектор. Очевидно, продолжение (15) единственno, если \mathbf{V}_*^T — неособенная матрица.

Аналогичным образом можно строить алгоритмы для прогноза системы временных рядов по центрированной, либо центрированной и нормированной матрице наблюдений. Однако, этот вопрос слишком обширен и не будет здесь излагаться.

7 Иллюстративные примеры

В этом параграфе мы представим два примера применения метода "Гусеница" к многомерным временным рядам. Основная цель, которую мы будем ставить при применении многомерной гусеницы, — это сравнительное описание рядов, в основном с точки зрения наличия или отсутствия в них той или иной периодики и сравнения фаз найденных периодик.

Примеры применения одномерной "Гусеницы" подробно описаны в разделе 1.2. Хотя часть графических представлений, описанных там, может быть применена и для многомерного случая, многомерные ряды порождают свою специфику. Например, изображение собственных векторов будет разрывным (каждая часть собственного вектора длины M относится кциальному ряду). То же самое относится и к изображению средних, дисперсий, корреляций. Общее поведение временных рядов описывают главные компоненты, которые после восстановления, использующего матрицу собственных векторов, дают различный результат восстановления для разных рядов.

Другой специфической особенностью многомерной гусеницы является то, что ряды могут быть не только в разных масштабах, но и в разных единицах измерения. Если применяется вариант гусеницы с центрированием и нормированием, то это важно только для одновременного изображения рядов и их компонент. Если же используется вариант без центрирования и нормирования, то важно сделать ряды соизмеримыми, например, стандартизовать их (вычтя среднее по ряду и разделив на стандартное отклонение).

В данном параграфе будем использовать только вариант метода с центрированием и нормированием, также будет использоваться и предварительная стандартизация ряда.

7.1 Безработица в США

Рассмотрим данные US_UNEMPL, содержащие четыре временных ряда длиной 408, которые представляют собой месячные данные по безработице в США с 1948 по 1981 год. Эти временные ряды приведены в книге D.F.Andrews and H.M.Herzberg, Data: A Collection of Problems from Many Fields for the Student and Research Worker, 1985, Springer-Verlag. Первый ряд — безработица среди мужчин в возрасте до 20 лет, третий ряд — безработица среди мужчин в возрасте после 20 лет, второй и четвертый ряды — аналогичные характеристики для женщин.

Для того, чтобы было удобно изображать ряды на одном графике, мы стандартизуем их. Чтобы графики не накладывались друг на друга, будем пользоваться стандартным приёмом, состоящим в добавлении констант к значениям рядов. Поэтому все графики нужно рассматривать не с точки зрения абсолютных значений, а с точки зрения сравнения поведения функций. На всех рисунках, где изображены составляющие четырех рядов, первый ряд изображен снизу, второй — выше и т.д. Анализируемые в дальнейшем ряды после преобразований изображены на рис. 1.

На рис. 2 представлена корреляционная матрица матрицы наблюдений, построенной по ряду с параметром длина гусеницы, равным 24. Так как мы рассматриваем 4 ряда, то корреляционная матрица имеет размер $96 = 24 \times 4$ на

96. Матрица разделена на 16 квадратов, диагональные соответствуют корреляционным соотношениям внутри ряда, а внедиагональными — между рядами. Можно заметить четкую годичную (12-точечную) периодичность внутри первых двух рядов, более слабое присутствие этой периодичности в третьем ряде и почти отсутствие таковой в четвертом. Так же смещение диагональных полос в квадратах, соответствующих корреляционным соотношениям первого и второго с третьим рядом может свидетельствовать о смещении фазы годовой периодики в третьем ряде.

На рис. 3 изображены первые десять главных компонент рядов. Отнесем первые две к тренду, ГК 3 и 8 представляют некоторую регулярность, но также могут быть отнесены к тренду, ГК 4–5 и ГК 6–7 соответствуют годовой и полугодовой гармоническим составляющим (периодикам), ГК 9–10 — четырехмесячной периодике.

Рис. 4, на котором изображены собственные вектора, соответствующие главным компонентам, отнесенными к тренду, показывает разницу между рядами с точки зрения тренда. На рис. 5 изображены собственные вектора, соответствующие главным компонентам, отнесенными к годичной периодике. Можно увидеть, что в четвертом ряде годичная гармоническая составляющая не присутствует. Так же можно заметить, что у первого и второго рядов фазы периодики близки и отличаются от фазы третьего ряда.

Рис. 6–12 представляют собой различные результаты восстановления рядов по выбранным главным компонентам. На рис. 6 и 7 изображены два варианта, трендов на фоне исходных рядов, менее и более подробный. Восстановление по составляющей тренда, имеющей некоторую регулярность, показано на рис. 10. Видно, что она практически отсутствует во втором ряде.

Рис. 8 и 9 показывают гармонические составляющие рядов с годовым и полугодовым периодом. Как мы и предполагали по виду собственных векторов, наблюдается отличие фазы годовой периодики третьего ряда от фазы первого и второго, а также отсутствие периодик в четвертом ряде. Один из моментов, на который хотелось бы обратить внимание при интерпретации гармонических составляющих в многомерном ряде, это то, что наблюдается эффект "отражения" гармонической составляющей одного ряда в шумовой составляющей другого. Поэтому вопрос отличия слабой гармонической составляющей от результата такого "отражения" должен ставиться отдельно. На рис. 11 приведена годовая периодическая составляющая рядов, построенная с учетом годовой, полугодовой и четырехмесячных гармоник. Рис. 12 показывает довольно хорошее описание рядов первыми десятью главными компонентами.

7.2 Боевые потери войск союзников в войне во Вьетнаме

Перейдем теперь к одновременному анализу двух временных рядов, содержащих ежемесячные данные по боевым потерям во Вьетнаме, 1966–1971 года, 72 временные точки, с американской (первый ряд) и с южно-вьетнамской (второй ряд) стороны. По-отдельности эти ряды анализируются в разделе I.2. Временные ряды описаны в M.F.Janowitz, B.Schweizer Ordinal and Percentile Clustering, Mathematical Social Sciences 18 (1989) 135-186, Table 10 p.177 (со ссылкой на J.A.Hartigan, Clustering Algorithms , Wiley, N-Y, 1975, p.138).

Так как ряды соизмеримы, не будем проводить никаких преобразований. На рис. 13 изображены оба исходных ряда.

Рассмотрим два способа выбора длины гусеницы M .

- $M = 48$. На рис. 15 представлена корреляционная матрица матрицы данных, соответствующей ряду. Отметим, что в корреляционной матрице присутствуют отрицательные корреляции. Граница между отрицательными и положительными значениями выделена черной линией.

Метод главных компонент, примененный к этой матрице, дает следующий результат. Первую и четвертую главные компоненты можно отнести к тренду, ГК 2–3 дают годовую периодику.

На рис. 14 изображен тренд на фоне исходного ряда. Рис. 16 и 17 показывают результаты восстановления годовой периодики ряда, а также восстановление с учетом тренда и годовой составляющей. Видно, что результаты в целом подобны тем, которые были получены с помощью применения одномерного метода "Гусеница" по-отдельности к каждому ряду в разделе I.2.

- $M = 4$. Покажем теперь, какой эффект может быть получен с помощью маленькой длины гусеницы с точки зрения сравнения рядов. На рис. 18 представлен результат восстановления ряда по среднему и первой главной компоненте. Видно, что то, что получается в результате, не является трендом. Будем интерпретировать первую главную компоненту, как компоненту, описывающую общее в поведении двух рядов. Рис. 19 показывает, что вторая главная компонента содержит в себе различие между рядами. В целом первые две главные компоненты дают вклад 71.5% и достаточно хорошо описывают поведение рядов. Таким образом, маленькая длина гусеницы позволила выразить совместное поведение двух рядов через общее между ними и их различие.

Литература

- [1] Гельфанд, А.О. Исчисление конечных разностей. Москва, “Наука”, 1967.
- [2] Rao, C.R., Линейные статистические методы и их применения. М., Наука, 1968.
- [3] Соболев, С.Л., Введение в теорию кубатурных формул. М., Наука, 1974.

III.3 Метод главных компонент для обработки многомерных полей.

Д.Л. Данилов

1 Введение

Целью данного раздела является описание и исследование некоторых свойств метода обработки многомерных случайных полей, использующего технику главных компонент. Алгоритмически метод является естественным развитием и обобщением методов типа "Гусеница", используемых для одномерных и многомерных временных рядов. Как известно, основой метода "Гусеница" является преобразование исходного набора данных (будь то одномерный или многомерный ряд) в многомерную выборку, с теоретической точки зрения представляющей собой двумерное поле. Эта выборка в дальнейшем раскладывается по системе собственных векторов её ковариационной (корреляционной) матрицы. Соответствующий алгоритм часто носит название сингулярного разложения матрицы. Идею преобразования исходного набора данных в многомерную выборку удается эффективно реализовать в задаче обработки многомерных массивов данных. Далее мы будем подразумевать, что располагаем вещественным n -мерным набором данных, т.е. таблицей чисел $\Xi = \{\xi_{i_1, \dots, i_n}, i_k = 1, \dots, N_k; k = 1, \dots, n\}$. Наиболее непосредственной аналогией является дискретизация некоторой вещественной функции n -аргументов, возможно, наблюданной со случайной ошибкой, заданной в области прямоугольной формы на равномерной по каждому аргументу сетке. Мы будем называть данные такого вида n -мерным случайным полем. Изложение будет проведено сначала в частном случае $n = 2$, а затем обобщено на случай произвольного n .

2 Изложение основной идеи. "Гусеница" для двумерных полей.

В этом разделе приводится изложение основных идей предлагаемого метода на простейшей схеме: схеме обработки двумерных полей. Пусть имеется вещественное двумерное случайное поле $\Xi = \{\xi_{i_1, i_2}, i_1 = 1, \dots, N_1; i_2 = 1, \dots, N_2\}$. Рассмотрим следующую (неформальную) процедуру преобразования этого набора данных в матрицу наблюдений. Фиксируем числа $\tau_k, 1 \leq \tau_k \leq N_k$. (В дальнейшем в основном будет рассматриваться случай $1 < \tau_k < N_k$.) Рассмотрим набор индексов $(i_1, i_2): i_1 = 1, \dots, \tau_1; i_2 = 1, \dots, \tau_2$, этот набор далее будем называть "шаблоном". Шаблону соответствует участок поля $\xi_{i_1, i_2}, i_1 = 1, \dots, \tau_1; i_2 = 1, \dots, \tau_2$. Перенумеруем элементы этого участка построчно, получая вектор $X^{(1)} = (X_1^{(1)}, \dots, X_{\tau_1 \tau_2}^{(1)})^T = (\xi_{1,1}, \xi_{1,2}, \dots, \xi_{1,\tau_2}, \xi_{2,1}, \dots, \xi_{2,\tau_2}, \dots, \xi_{\tau_1,\tau_2})^T$. После этого переместим шаблон на единицу вправо, рассмотрим участок исходного поля $\xi_{i_1, i_2}, i_1 = 1, \dots, \tau_1; i_2 = 2, \dots, \tau_2 + 1$, элементы которого составят второй вектор P -мерной матрицы наблюдений ($P = \tau_1 \tau_2$). Этую процедуру продолжим до тех пор, пока правый край шаблона не дойдет до правого края поля.

После этого процедура повторяется со сдвигом на единицу вниз и т.д. Результатом такой операции является матрица наблюдений, состоящая из $Q = (N_1 - \tau_1 + 1)(N_2 - \tau_2 + 1)$ столбцов.

$$\mathbf{X} = (\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(Q)}). \quad (1)$$

Ниже изображены 4 этапа генерации матрицы наблюдений для двумерного поля $N_1 = 3, N_2 = 4$ с шаблоном $\tau_1 = 2, \tau_2 = 3$.

$$\begin{array}{c} \left(\begin{array}{cccc} \lceil \xi_{1,1} & \xi_{1,2} & \xi_{1,3} \rceil & \xi_{1,4} \\ \lfloor \xi_{2,1} & \xi_{2,2} & \xi_{2,3} \rfloor & \xi_{2,4} \\ \xi_{3,1} & \xi_{3,2} & \xi_{3,3} & \xi_{3,4} \end{array} \right) \rightarrow \left(\begin{array}{cccc} \xi_{1,1} & \lceil \xi_{1,2} & \xi_{1,3} & \xi_{1,4} \rceil \\ \xi_{2,1} & \lfloor \xi_{2,2} & \xi_{2,3} & \xi_{2,4} \rfloor \\ \xi_{3,1} & \xi_{3,2} & \xi_{3,3} & \xi_{3,4} \end{array} \right) \\ \swarrow \\ \left(\begin{array}{cccc} \xi_{1,1} & \xi_{1,2} & \xi_{1,3} & \xi_{1,4} \\ \lceil \xi_{2,1} & \xi_{2,2} & \xi_{2,3} \rceil & \xi_{2,4} \\ \lfloor \xi_{3,1} & \xi_{3,2} & \xi_{3,3} \rfloor & \xi_{3,4} \end{array} \right) \rightarrow \left(\begin{array}{cccc} \xi_{1,1} & \xi_{1,2} & \xi_{1,3} & \xi_{1,4} \\ \xi_{2,1} & \lceil \xi_{2,2} & \xi_{2,3} & \xi_{2,4} \rceil \\ \lfloor \xi_{3,1} & \xi_{3,2} & \xi_{3,3} & \xi_{3,4} \rfloor \end{array} \right). \end{array}$$

В этом случае матрица наблюдений имеет вид

$$\mathbf{X} = \begin{pmatrix} \xi_{1,1} & \xi_{1,2} & \xi_{2,1} & \xi_{2,2} \\ \xi_{1,2} & \xi_{1,3} & \xi_{2,2} & \xi_{2,3} \\ \xi_{1,3} & \xi_{1,4} & \xi_{2,3} & \xi_{2,4} \\ \xi_{2,1} & \xi_{2,2} & \xi_{3,1} & \xi_{3,2} \\ \xi_{2,2} & \xi_{2,3} & \xi_{3,2} & \xi_{3,3} \\ \xi_{2,3} & \xi_{2,4} & \xi_{3,3} & \xi_{3,4} \end{pmatrix}.$$

Полученная таким образом матрица наблюдений обрабатывается посредством обычного нецентрального метода главных компонент, а именно,

1. Вычисляется нецентрированная ковариационная матрица $C = \frac{1}{Q}\mathbf{X}\mathbf{X}^T$.
2. Производится разложение C по собственным векторам:

$$V = (v^{(1)}, \dots, v^{(P)}), \quad (v^{(i)})^T v^{(j)} = \delta_{ij}, \quad C = V\Lambda V^T, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_P).$$

Здесь $v^{(i)}$ - представляет собой i -ый собственный вектор, а Λ - диагональная матрица собственных чисел.

3. Вычисляется система главных компонент

$$\Phi = V^T \mathbf{X}, \quad \Phi = (\Phi^{(1)}, \dots, \Phi^{(P)})^T, \quad \Phi^{(s)} = (\phi_1^{(s)}, \dots, \phi_Q^{(s)}).$$

Одной из существенных особенностей методов типа "гусеница" является возможность интерактивного визуального анализа собственных векторов и главных компонент рассматриваемой матрицы данных. Однако здесь мы имеем первую принципиально новую черту: визуальный анализ собственных векторов самих по себе (как функций одного аргумента) оказывается малоинформативен. Для обеспечения возможности визуального анализа необходим переход от собственного вектора $v^{(i)}$, как такового, к его двумерному "изображению". Эта процедура заключается в построении двумерного поля $\tilde{\phi}_{s_1, s_2}^{(i)}$, такого, что $v_m^{(i)} = \phi_{s_1, s_2}^{(i)}$ для $m = s_2 + \tau_2(s_1 - 1)$. Полученное "собственное поле" $\tilde{\phi}_{s_1, s_2}^{(i)}$ по

размерам соответствует исходному "шаблону". Аналогичная процедура применима к главной компоненте, при этом размеры поля главной компоненты дополнительны размерам шаблона.

После анализа собственных векторов и главных компонент и отбора информативных наступает очередь восстановления исходной матрицы. Это производится обычным для техники главных компонент образом. Если мы обозначим через \mathfrak{J} набор индексов, соответствующих интересующим нас главными компонентам, то процедура восстановления записывается так:

$$\hat{\mathbf{X}} = \sum_{t \in \mathfrak{J}} v^{(t)} \Psi^{(t)}. \quad (2)$$

Однако $\hat{\mathbf{X}}$ — это восстановленная матрица наблюдений, а мы хотим восстановить исходное поле $\hat{\Xi}$. Процедура восстановления исходного поля имеет следующий вид

$$\hat{\xi}_{i_1, i_2} = \frac{1}{|\mathfrak{J}(i_1, i_2)|} \sum_{(j_1, j_2) \in \mathfrak{J}(i_1, i_2)} \hat{x}_{j_1, j_2}. \quad (3)$$

Набор индексов $\mathfrak{J}(i_1, i_2)$ перечисляет все (j_1, j_2) такие, что $x_{j_1, j_2} = \xi_{i_1, i_2}$ в исходной матрице данных \mathbf{X} .

3 Примеры обработки двумерных полей.

Свойства предлагаемого алгоритма логично рассматривать сначала на очень простых наборах данных. В качестве первого примера рассмотрим следующую функцию двух дискретных аргументов: $\Xi(i_1, i_2) = \sin(\pi i_1/9)\sin(\pi i_2/9) + \frac{4}{9}\sin(\pi i_1/3)\sin(\pi i_2/3)$, где $i_1 = 1, \dots, 36, i_2 = 1, \dots, 36$. На Рис. 1 изображен вид этой функции. Для последующей обработки параметры гусеницы были выбраны равными $\tau_1 = \tau_2 = 9$. Представление о поведении собственных чисел этой матрицы может быть получено из Рис. 2. Как мы видим, всего 8 собственных чисел этой матрицы отличны от нуля и они сгруппированы двумя "ступенями", по предположительно соответствующими колебательным компонентам разной частоты. Это говорит о том, что в случае "Гусеницы" для полей, так же как и для обычной "Гусеницы" имеются наборы данных, порождающие матрицы наблюдений, лежащие в подпространствах малой размерности. На Рис. 3 мы видим старшие 4 главные компоненты, соответствующие первому (низкочастотному) слагаемому функции Ξ , на Рис. 4 построены главные компоненты с 5 по 8, соответствующие второму слагаемому. Рис. 5 и 6 показывают восстановления по указанным наборам ГК, полностью выделяющих первое и второе слагаемые исходной функции.

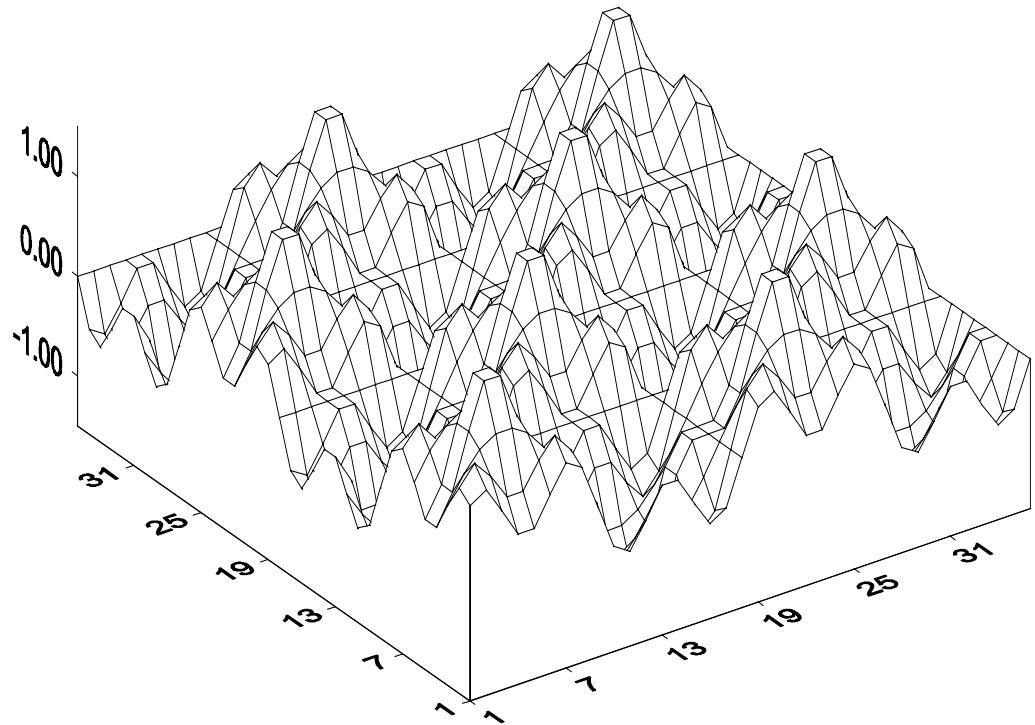


Рис. 1. Функция $\Xi(i_1, i_2) = \sin(\pi i_1/9)\sin(\pi i_2/9) + \frac{4}{9}\sin(\pi i_1/3)\sin(\pi i_2/3)$.

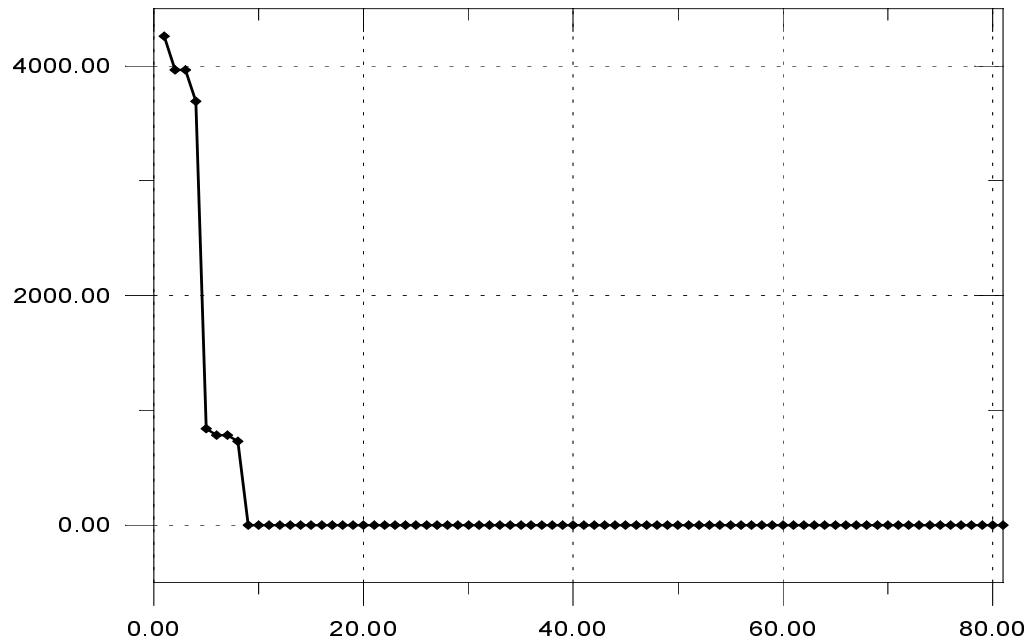


Рис. 2. Собственные числа нецентральной ковариационной матрицы порожденной функцией $\Xi(i_1, i_2) = \sin(\pi i_1/9)\sin(\pi i_2/9) + \frac{4}{9}\sin(\pi i_1/3)\sin(\pi i_2/3)$.

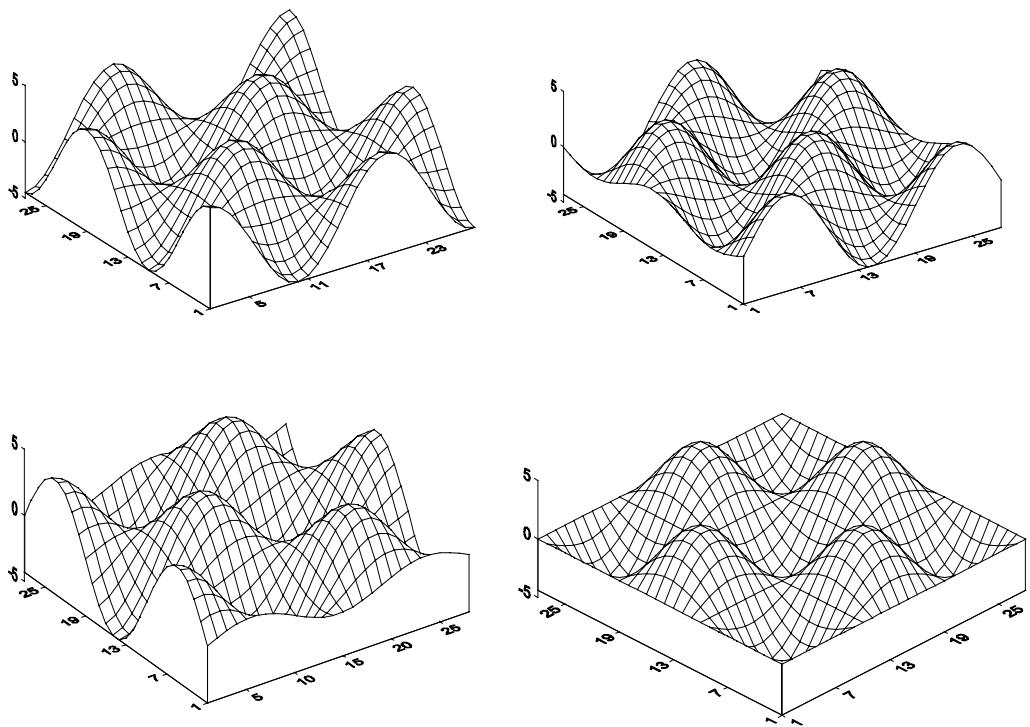


Рис. 3. Главные компоненты, соответствующие первым 4-м собственным числам.

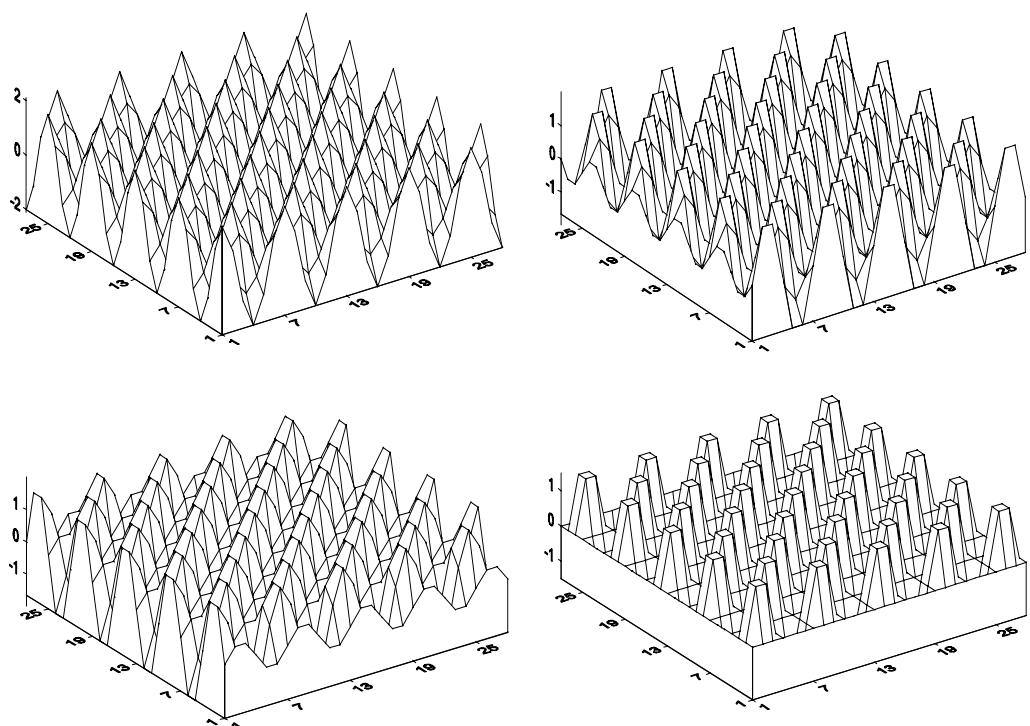


Рис. 4. Главные компоненты, соответствующие собственным числам с 5 по 8.

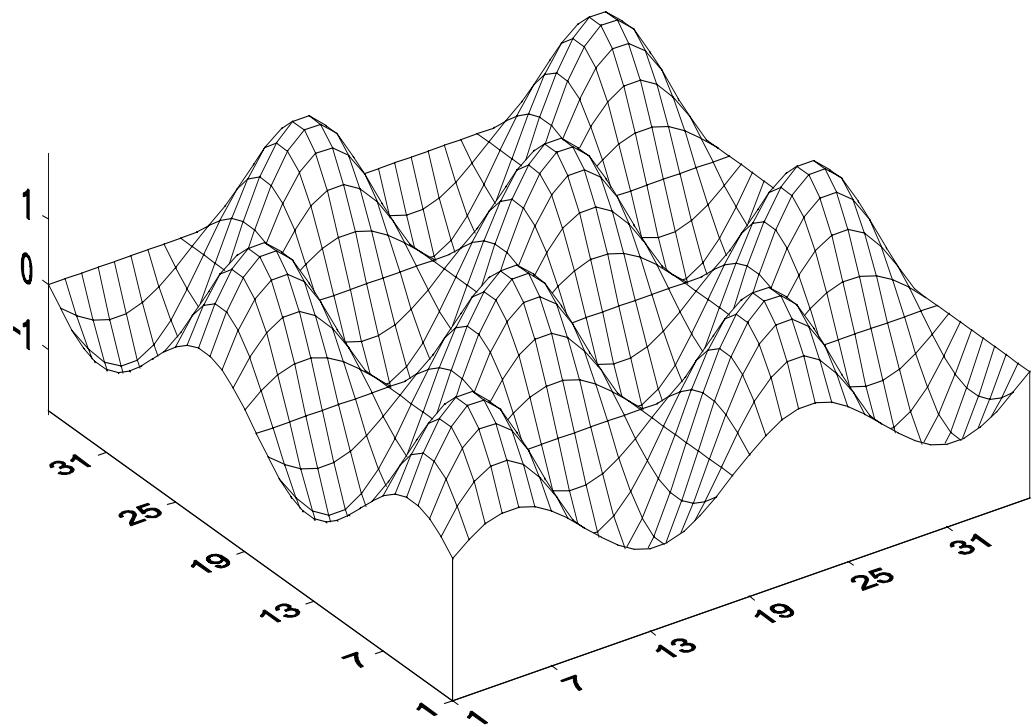


Рис. 5. Восстановление исходного поля по главным компонентам с 1 по 4.

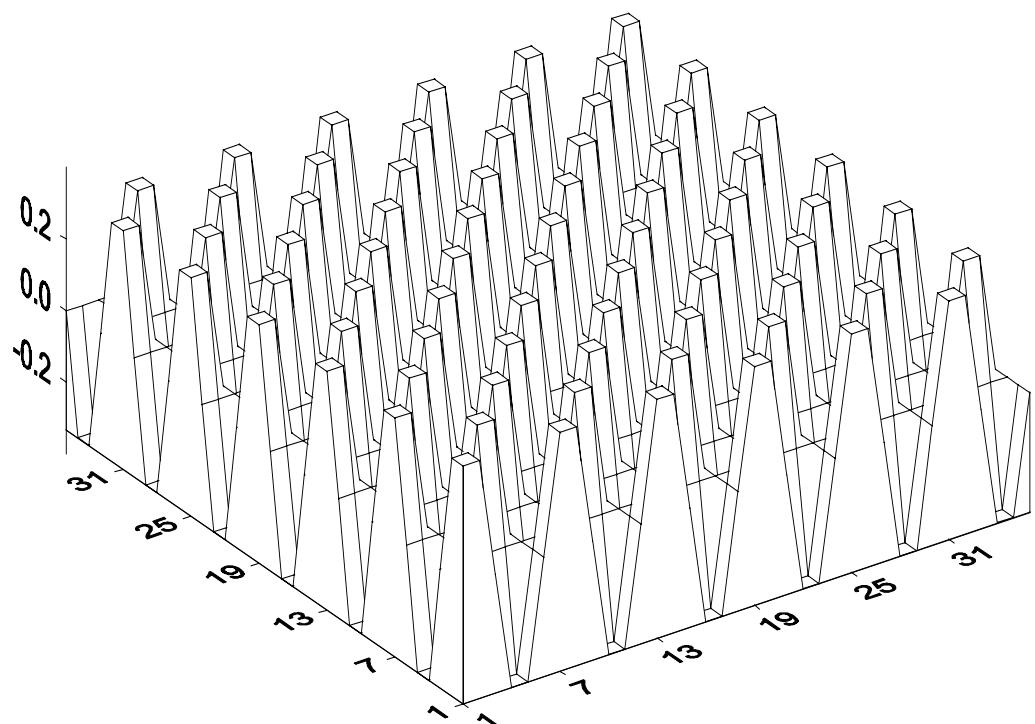


Рис. 6. Восстановление исходного поля по главным компонентам с 5 по 8.

Алгоритм может использоваться для отделения информативной части поля от шума. В качестве примера рассмотрим следующую функцию $\Xi(i_1, i_2) = \sin(\pi i_1/9)\sin(\pi i_2/9) + \frac{4}{9}\sin(\pi i_1/3)\sin(\pi i_2/3) + (i_1 + 2i_2)/54 + \eta_{i_1, i_2}$, где $1 \leq i_1, i_2 \leq 36$, а η_{i_1, i_2} - независимые стандартные нормальные случайные величины. Эта функция изображена на Рис. 7. Для последующей обработки параметры гусеницы были выбраны равными $\tau_1 = \tau_2 = 18$. Для этого примера мы приводим собственные вектора (а не главные компоненты), поскольку их размерность невелика и они визуально воспринимаются проще. На Рис. 8 изображены 1 и 10 собственные вектора, предположительно соответствующие медленно меняющейся части функции. На Рис. 10 мы видим собственные вектора с номерами 2-5, соответствующие первому (низкочастотному) слагаемому функции Ξ , на Рис. 12 построены собственные вектора с 6 по 9, соответствующие второму слагаемому. Рис. 9, 11 и 13 показывают восстановления по указанным наборам ГК, полностью выделяющих линейную часть, первое и второе слагаемые исходной функции.

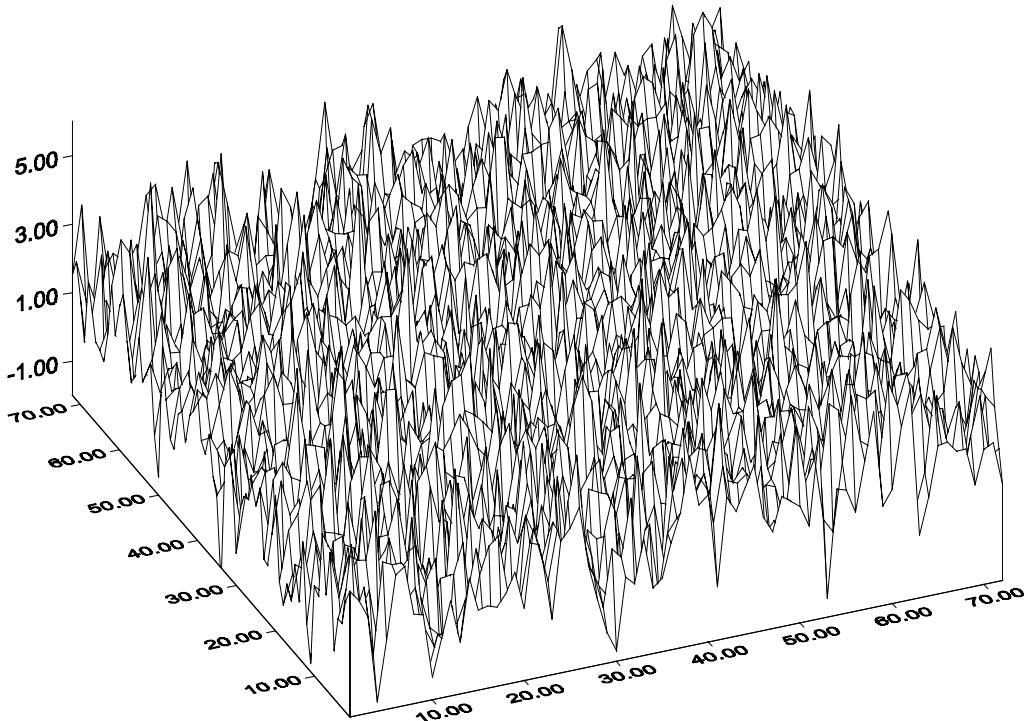


Рис. 7. Функция
 $\Xi(i_1, i_2) = \sin(\pi i_1/9)\sin(\pi i_2/9) + \frac{4}{9}\sin(\pi i_1/3)\sin(\pi i_2/3) + (i_1 + 2i_2)/54 + \eta_{i_1, i_2}$.

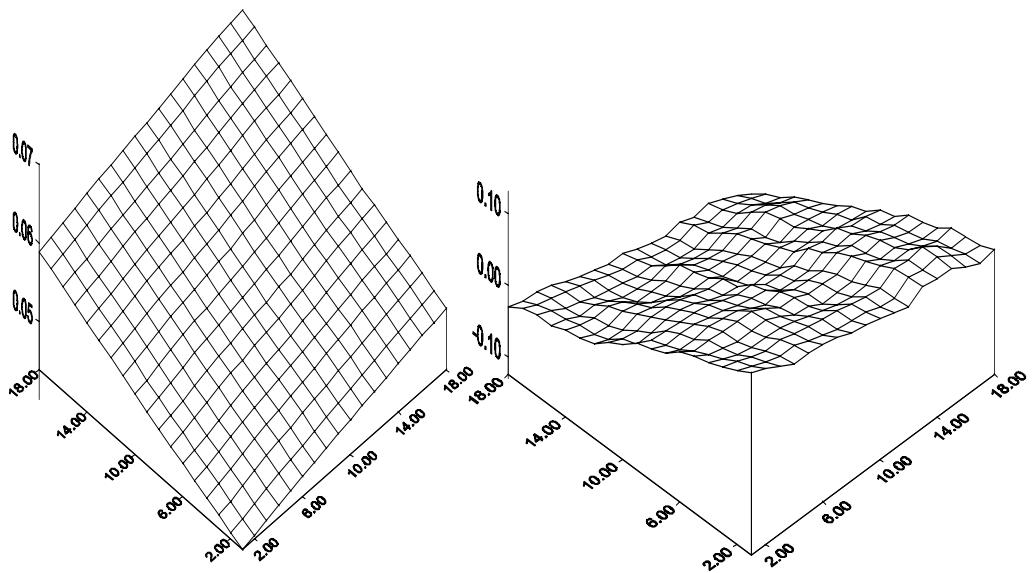


Рис. 8. Собственные вектора 1 и 10.

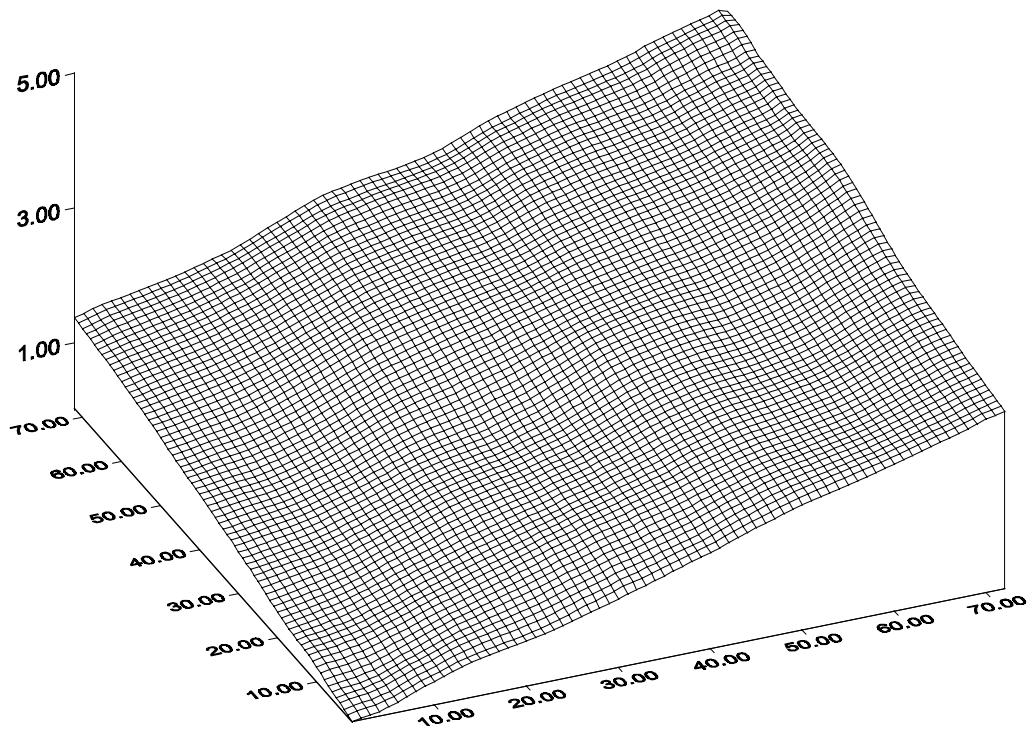


Рис. 9. Восстановление по главным компонентам 1 и 10.

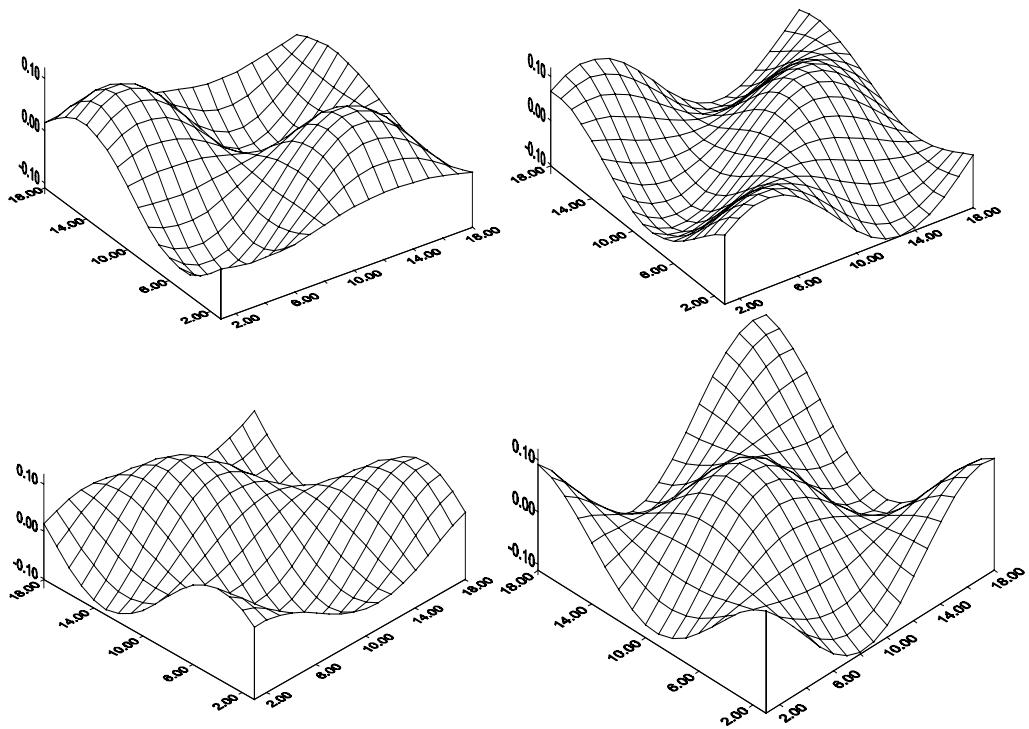


Рис. 10. Собственные вектора со 2 по 5.

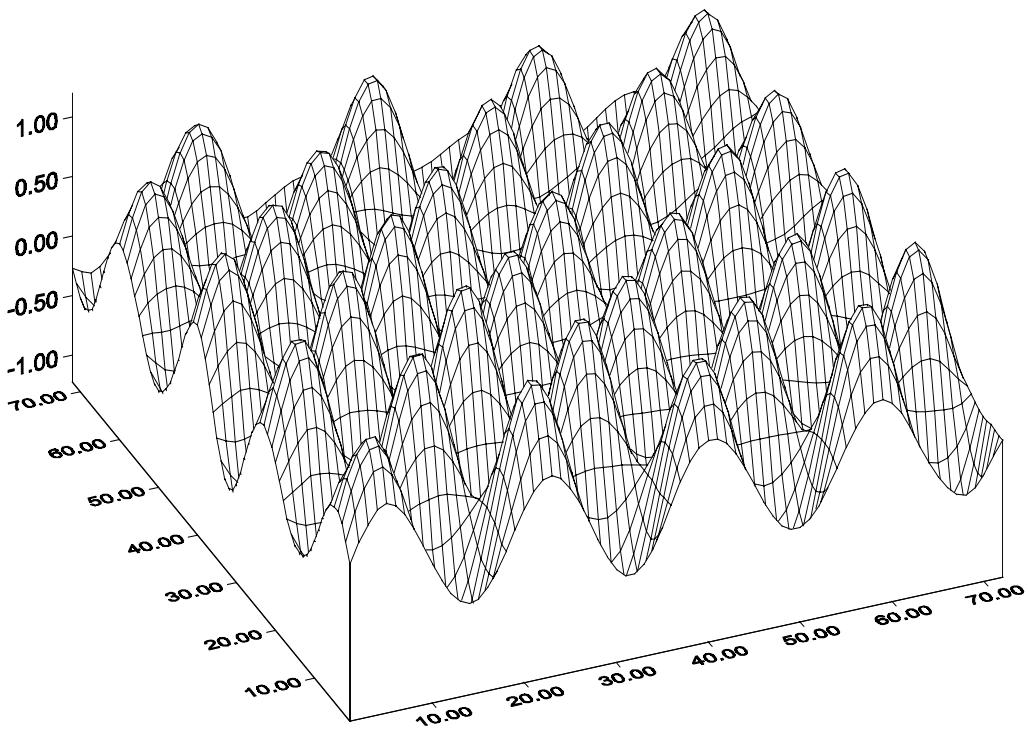


Рис. 11. Восстановление по главным компонентам со 2 по 5.

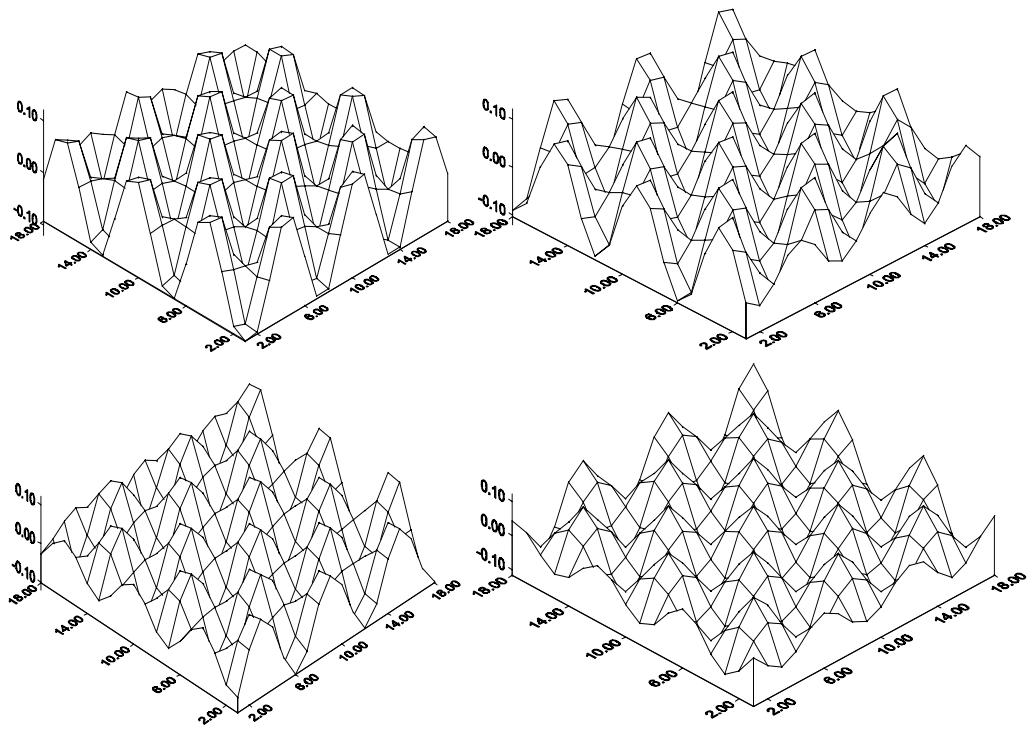


Рис. 12. Собственные вектора с 6 по 9.

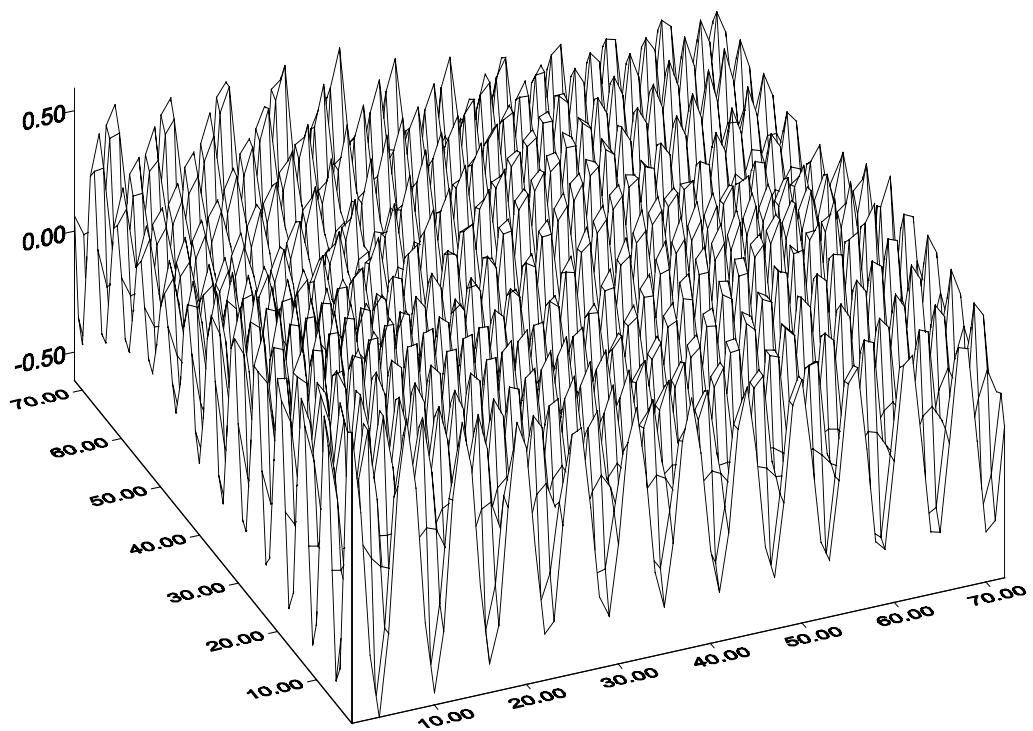


Рис. 13. Восстановление по главным компонентам с 6 по 9.

Еще один пример показывает как работает алгоритм в типичных задачах теории обработки изображений. Так, на Рис. 14 изображен тестовый пример "белый прямоугольник". Размерность исходного поля 120×90 точек, прямоугольник имеет размеры 20×30 точек. Изображение двуцветно. На Рис. 15 приведено изображение восстановления этого поля при гусеничной обработке с длинами 30×30 точек. На Рис. 16 приведено, фактически, то же изображение, однако приведенное к исходной двуцветной палитре: весь интервал значений разделен пополам, значения, меньшие половины, кодированы черным, остальные – белым. Следующий рисунок показывает то же исходное поле с наложенным мультипликативным бинарным шумом. Более точно: для каждого пикселя производилось бернульевское испытание с вероятностью успеха p , и в случае успеха цвет пикселя инвертировался. Для данного изображения $p = 1/3$. На Рис. 18 можно видеть результат восстановления этого поля по 6 первым главным компонентам, которые были выбраны по результатам визуального анализа как информативные. Следующий рисунок показывает результат приведения этого изображения к двуцветной палитре.

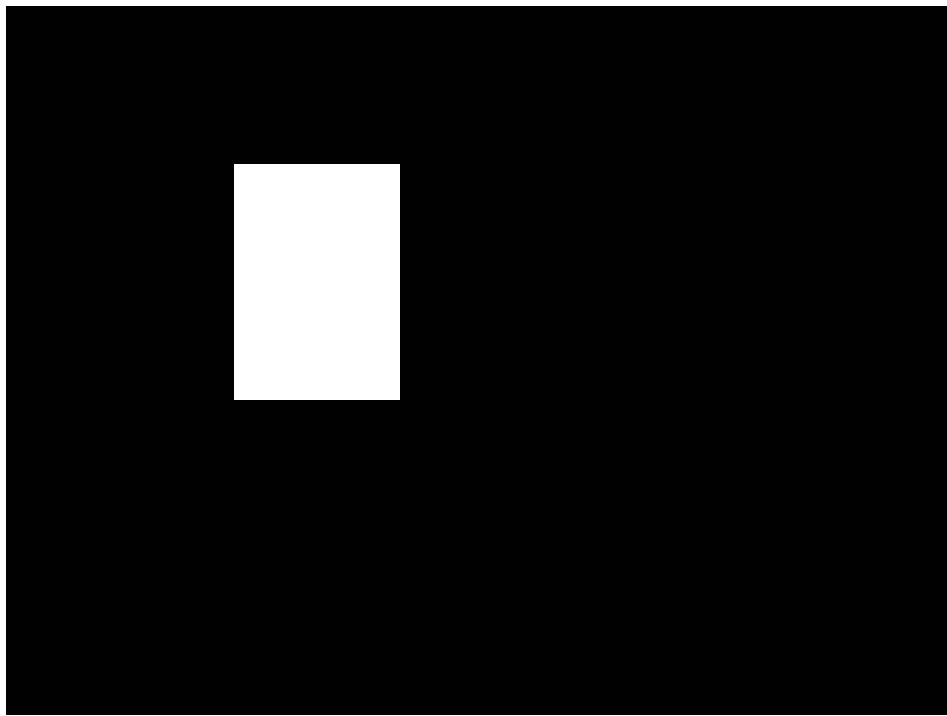


Рис. 14. Тестовый пример "белый прямоугольник на черном фоне".

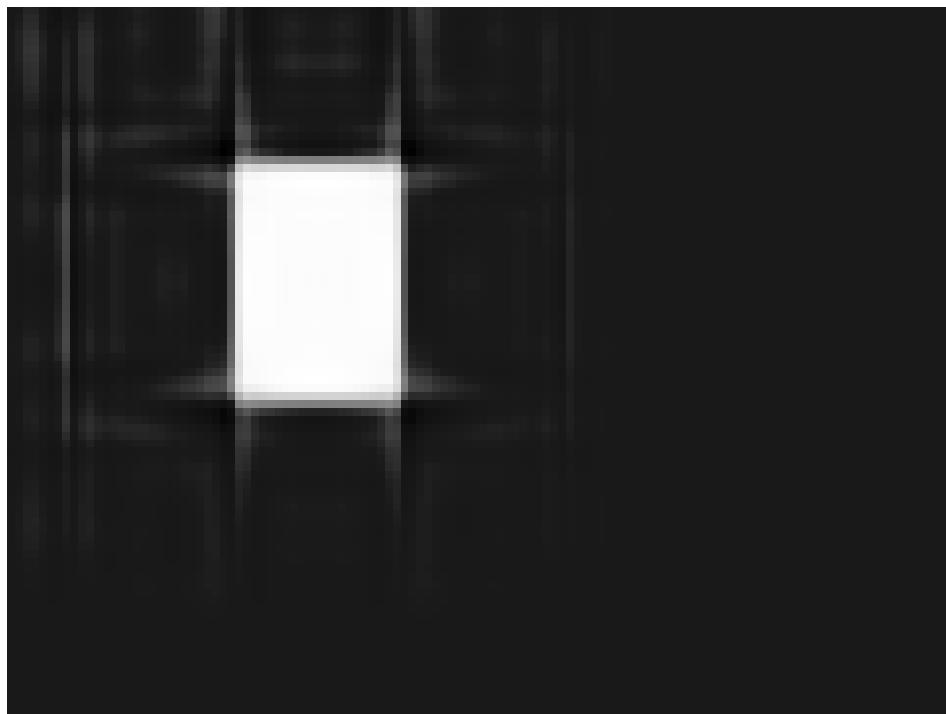


Рис. 15. Тестовый пример ”белый прямоугольник на черном фоне”.
Восстановление по первым 50 ГК.



Рис. 16. Тестовый пример ”белый прямоугольник на черном фоне”.
Восстановление по первым 50 ГК и возврат к исходным цветам.

Глава IV Сравнительный анализ метода “Гусеница” с классическими методами

Д.Л. Данилов

1 Объем пассажироперевозок на международных авиалиниях

Рассмотренный набор данных представляет собой количества перевезенных пассажиров на международных авиалиниях с января 1949 по декабрь 1960. Каждая точка соответствует объему пассажироперевозки за один месяц, числа даны в тысячах человек. Количество отсчетов равно 144. Данные впервые опубликованы в статье [4], таблица С 10. Этот временной ряд характерен для экономических данных. Даже простой визуальный анализ, а тем более обработки различными статистическими методами, показывают наличие тренда и интерпретируемых периодических составляющих, таких как сезонная компонента. Ряд очень хорошо известен и неоднократно рассматривался в классической литературе по обработке временных рядов, например в [1], Гл. 9. Исходный ряд представлен на Рис. 1.1. Непосредственное изучение формы ряда показывает увеличение амплитуды колебаний с течением времени. Это делает затруднительным непосредственное применение, например, анализа Фурье или подгонки авторегрессионных моделей. Заметим, что метод “Гусеница” с успехом работает по исходному ряду. Однако целью нашей работы является сравнительный анализ различных методов, поэтому для обработки был рассмотрен логарифмированный ряд (см. Рис. 1.1), в котором амплитуда колебаний стабилизована. Именно этот ряд и является стандартным.

1.1 Анализ методом “Гусеница”

Рассмотрим результаты применения “некентральной гусеницы” с длиной $\tau = 72$. В данном случае выбрана большая длина гусеницы, поскольку сам ряд также достаточно длинный. Главные компоненты приведены на Рис. 2.1-7.1. Результаты анализа главных компонент кратко могут быть представлены следующей таблицей.

ГК, №	Интерпретация	Рис. №
1	линейный тренд	2.1
2-3	годовая периодика (сезонная компонента)	2.1
4-5	полугодовая периодика (гармоника сезонной комп.)	2.1-3.1
6	нелинейный тренд	3.1
7-8	колебание с периодом прибл. 50 мес.	3.1
9-10	4 мес. колебание (гармоника сезонной компоненты)	4.1
11-12	3 мес. периодика (гармоника сезонной компоненты)	4.1
13-14	2.4 мес. периодика (гармоника сезонной компоненты)	5.1
15-16	27 мес. колебание.	5.1
17-18	18 мес. колебание.	6.1

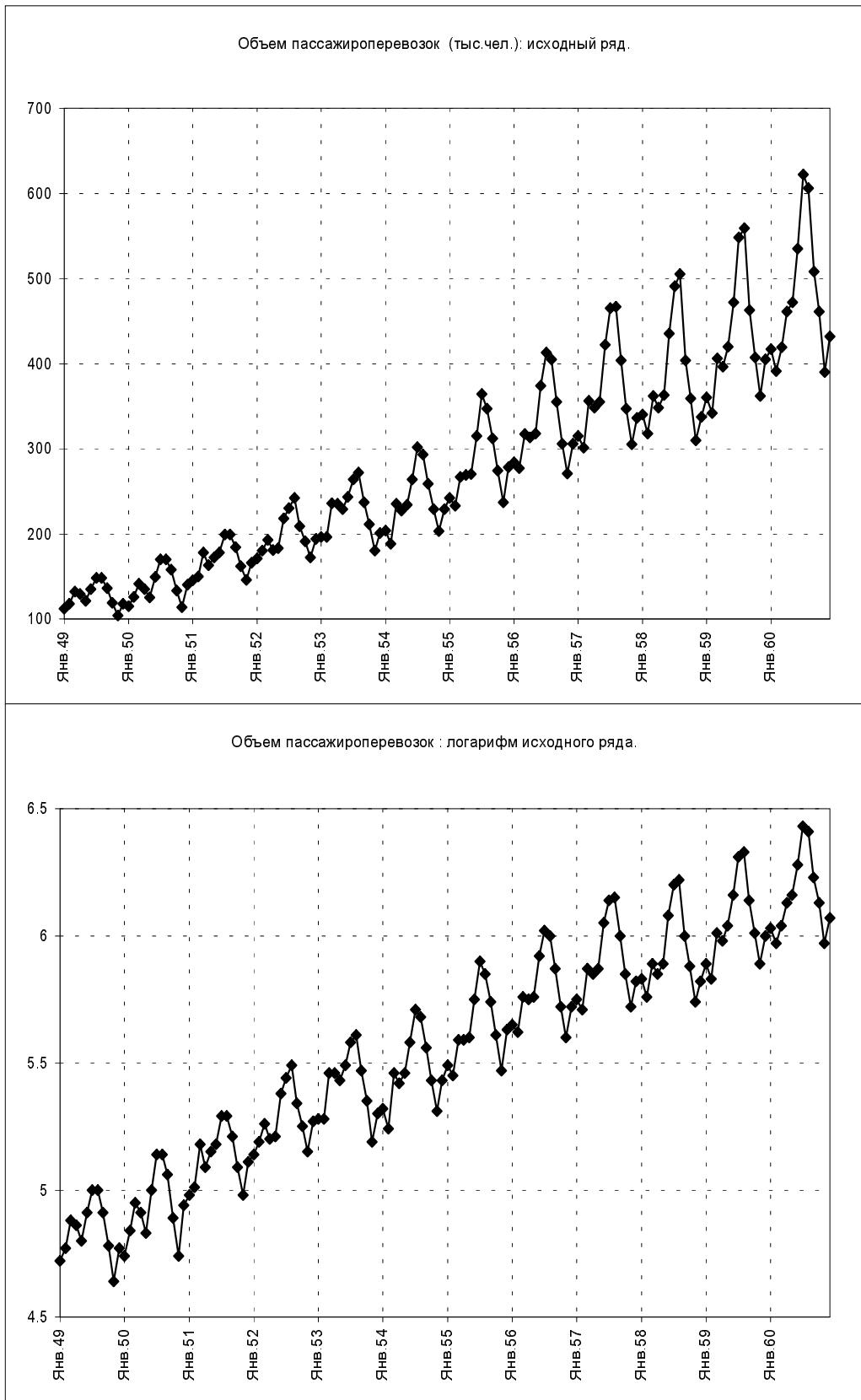


Рис. 1.1. Исходный и логарифмированный ряды.

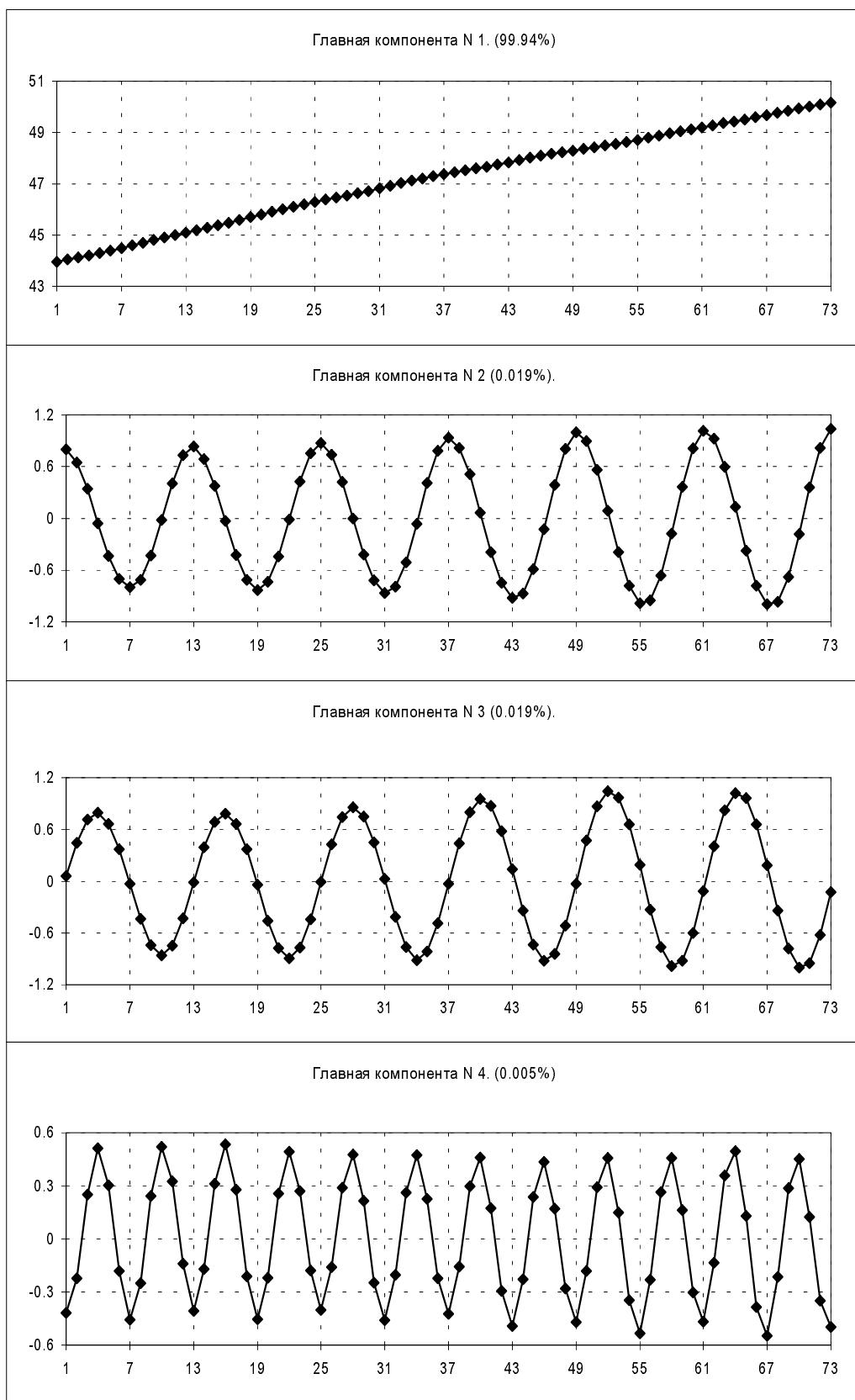


Рис. 2.1. Главные компоненты № 1-4.

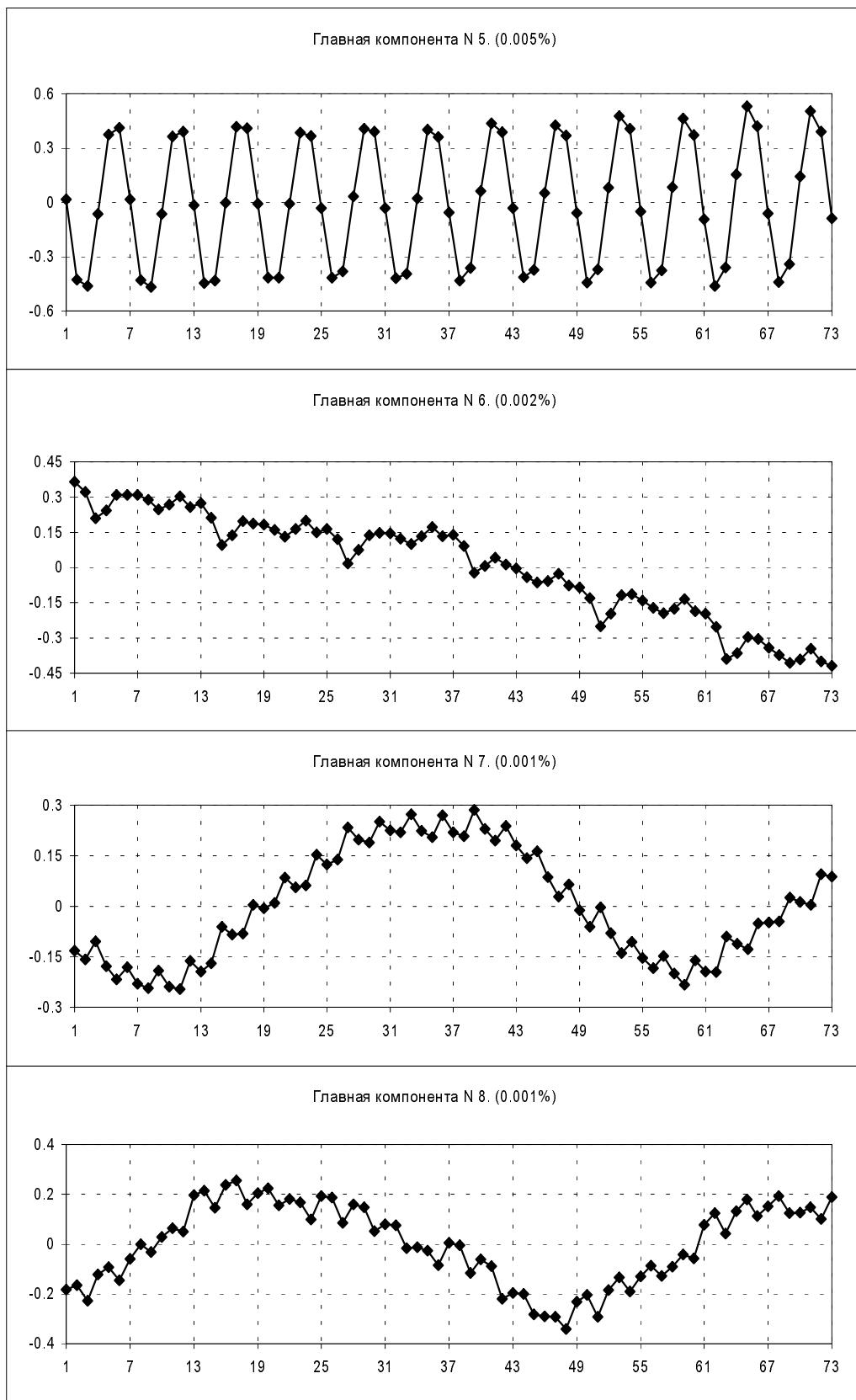


Рис. 3.1. Главные компоненты № 5-8.

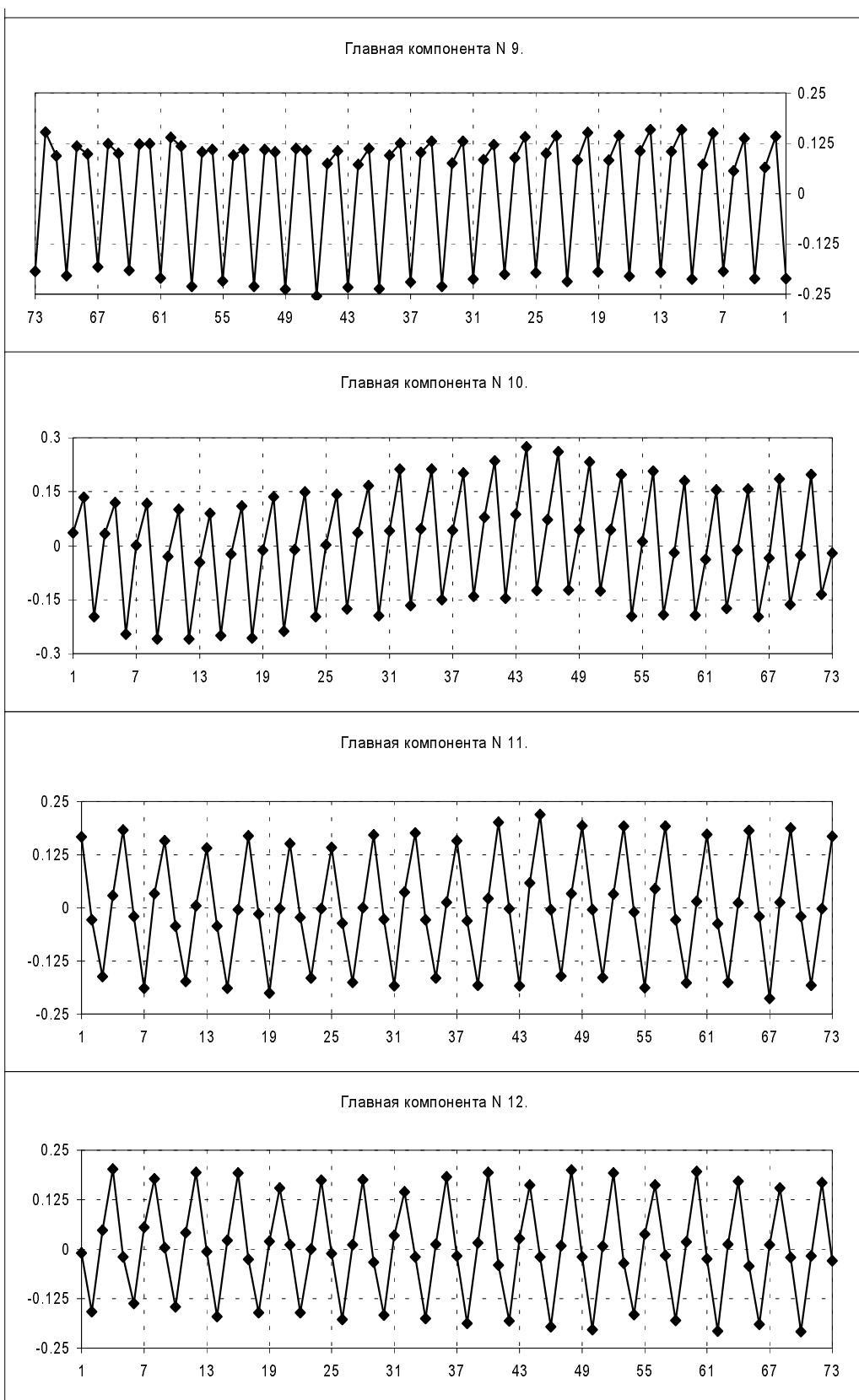


Рис. 4.1. Главные компоненты № 9-12.

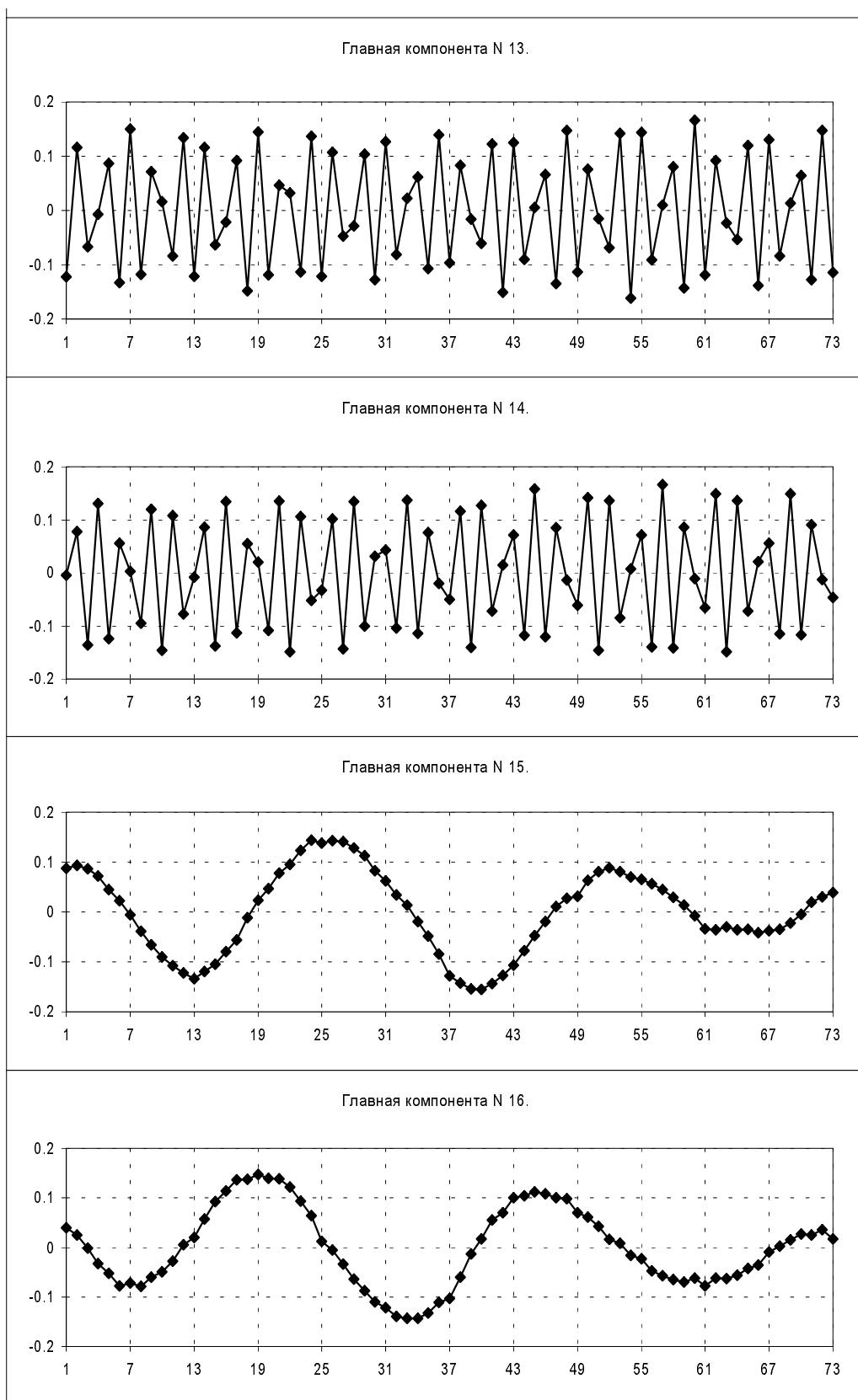


Рис. 5.1. Главные компоненты № 13-16.

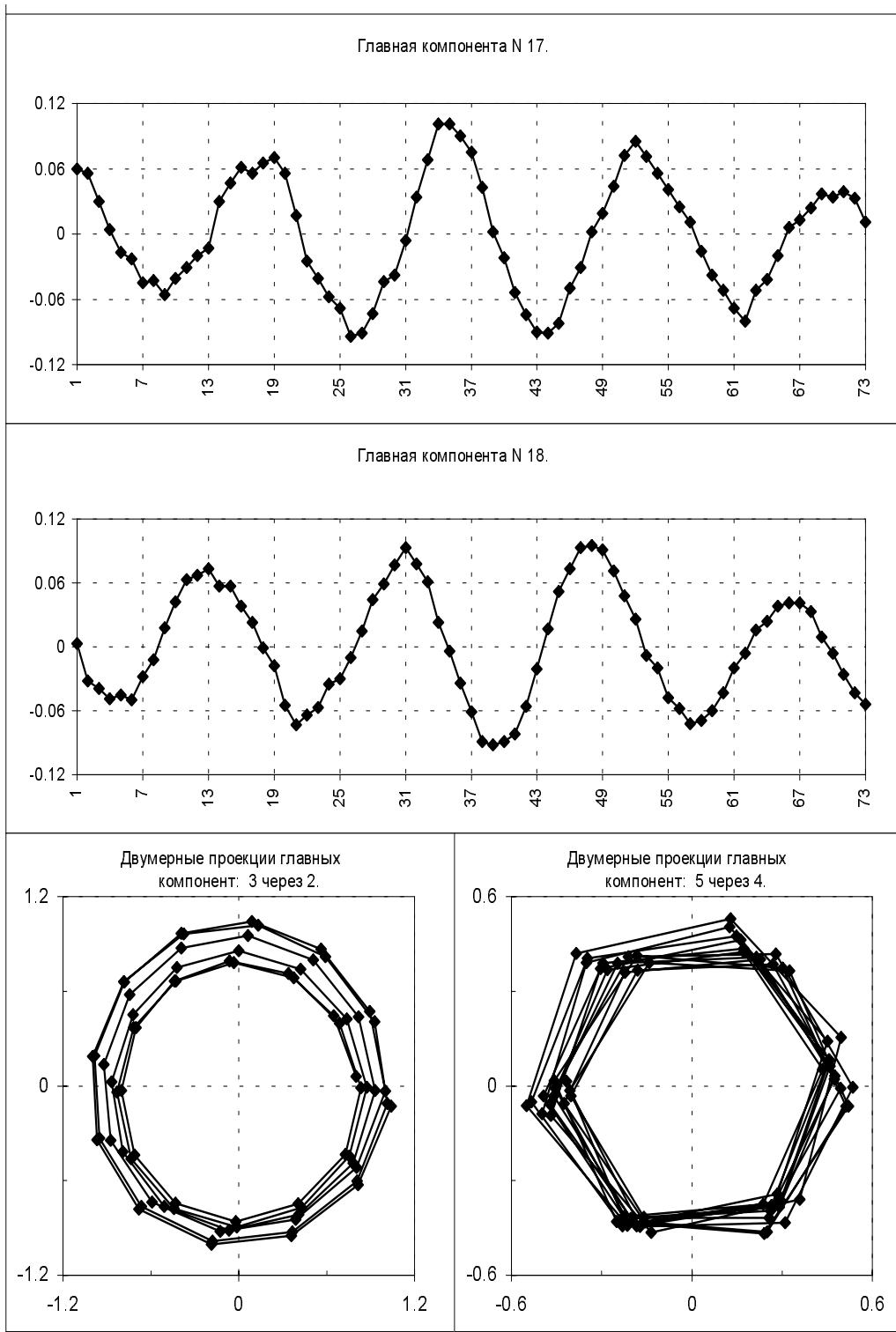


Рис. 6.1. Главные компоненты № 17-18. Двумерные проекции ГК 2-3, 4-5.

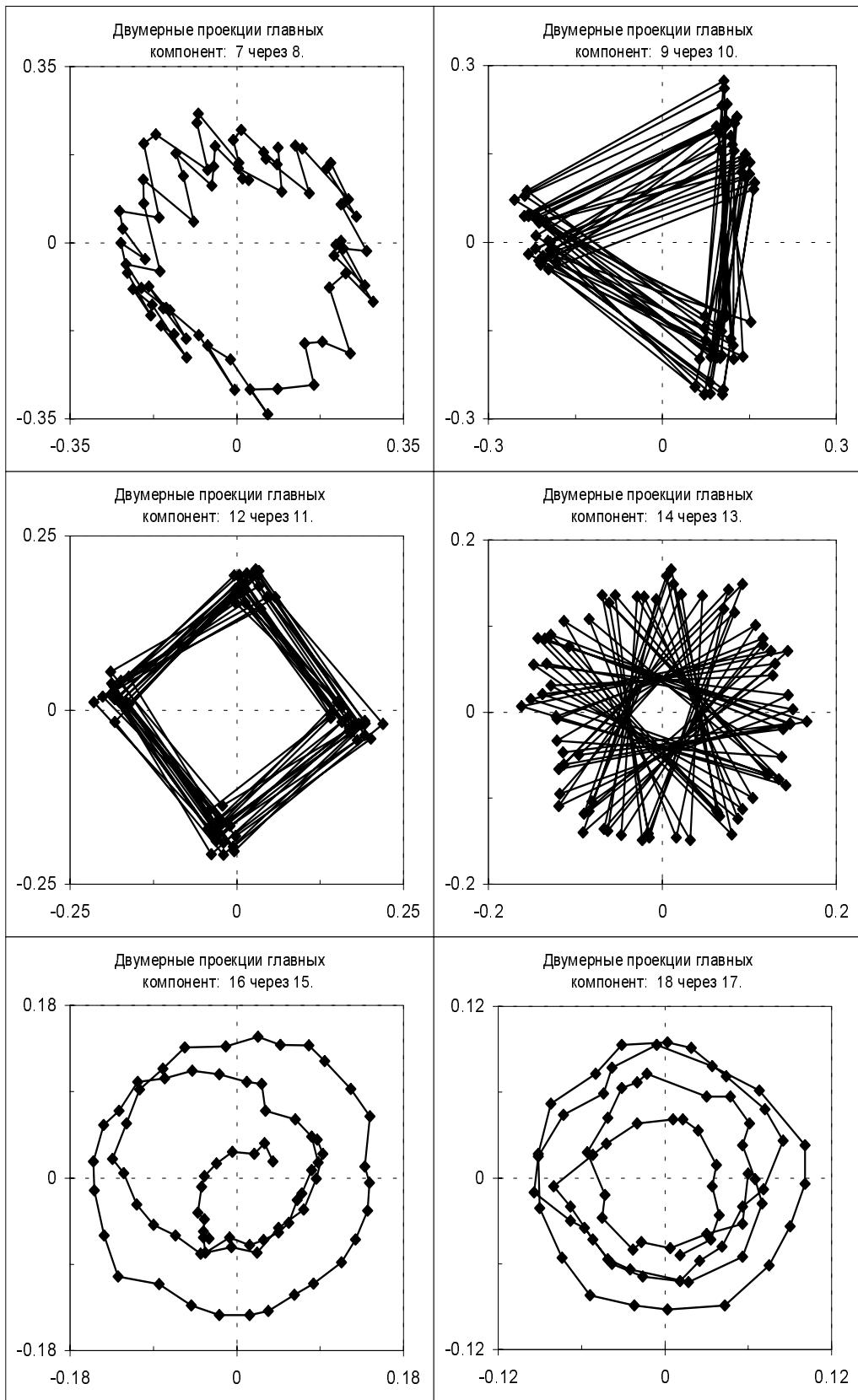


Рис. 7.1. Двумерные проекции ГК 7-8, 9-10, 11-12, 13-14, 15-16, 17-18

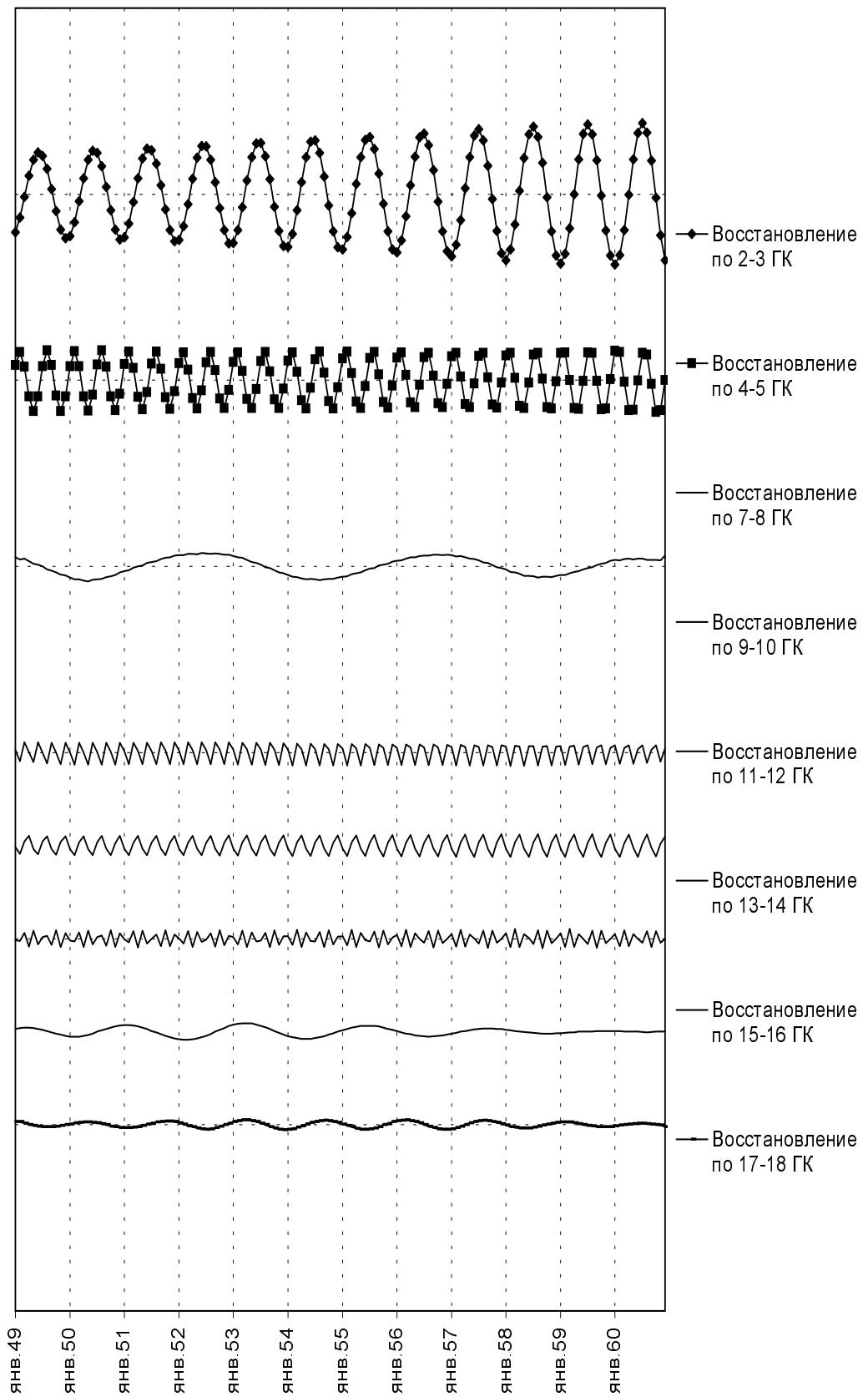


Рис. 8.1. Восстановление основных периодик.

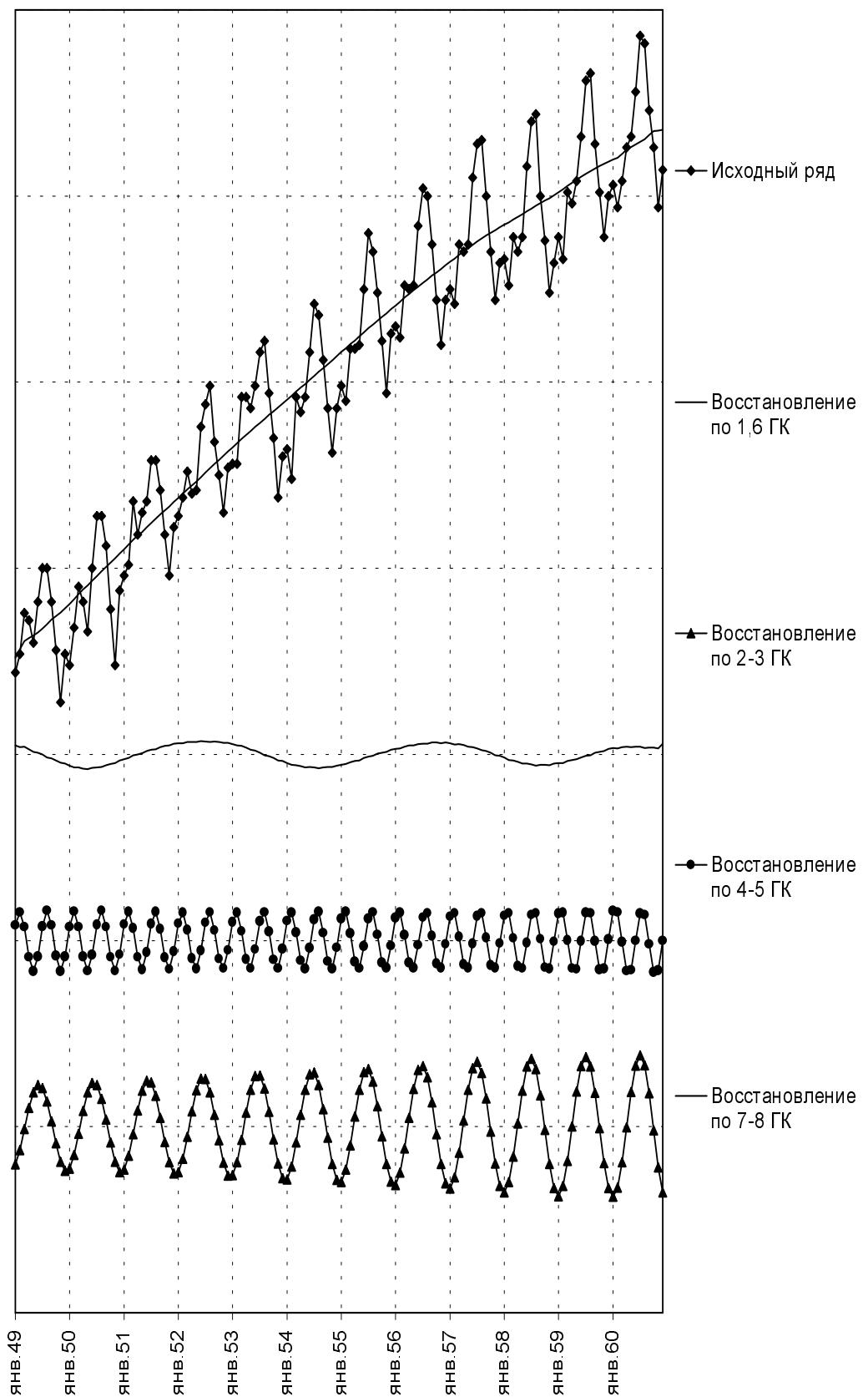


Рис. 9.1. Тренд и основные периодики.

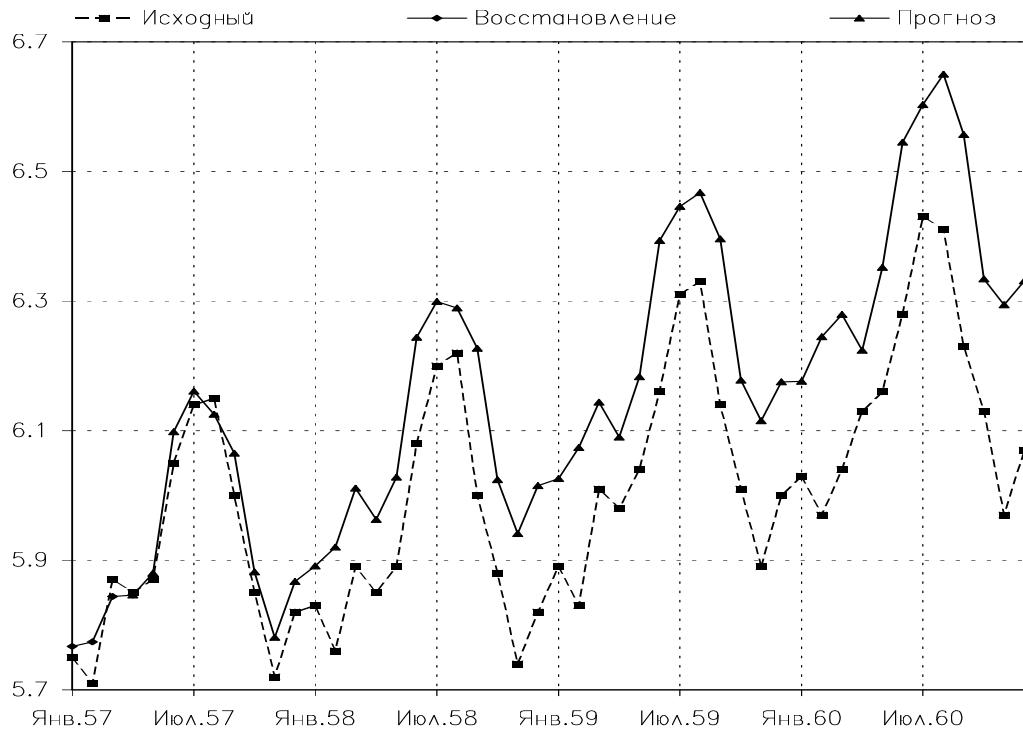
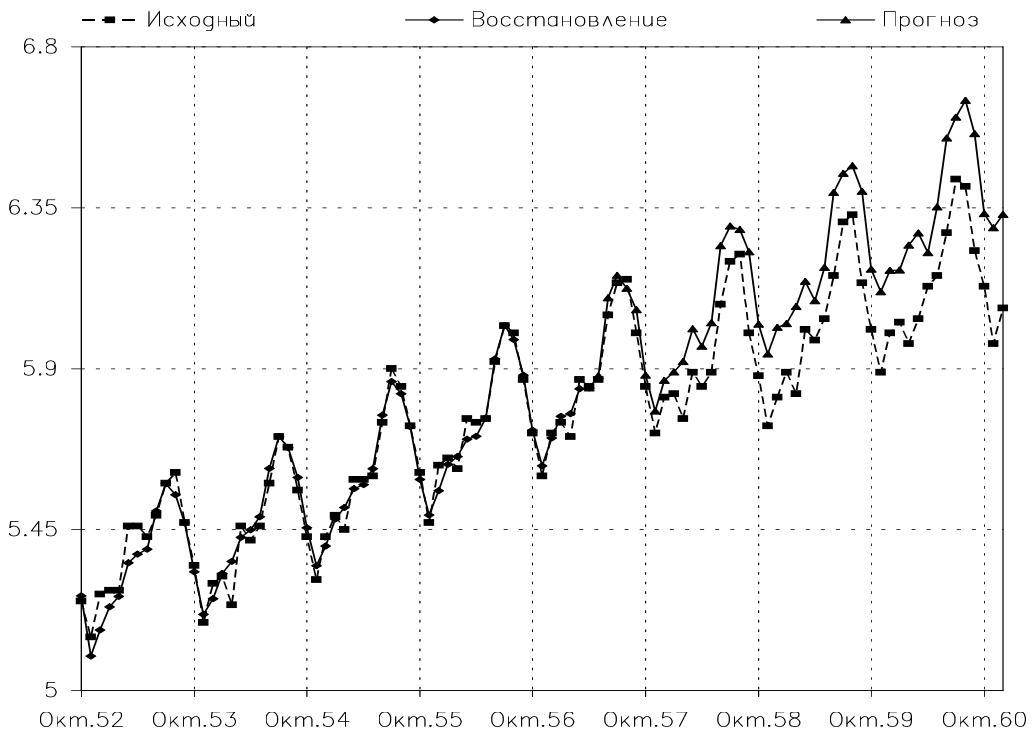


Рис. 10.1. Исходный ряд, восстановление и прогноз. На нижнем рисунке масштаб увеличен.

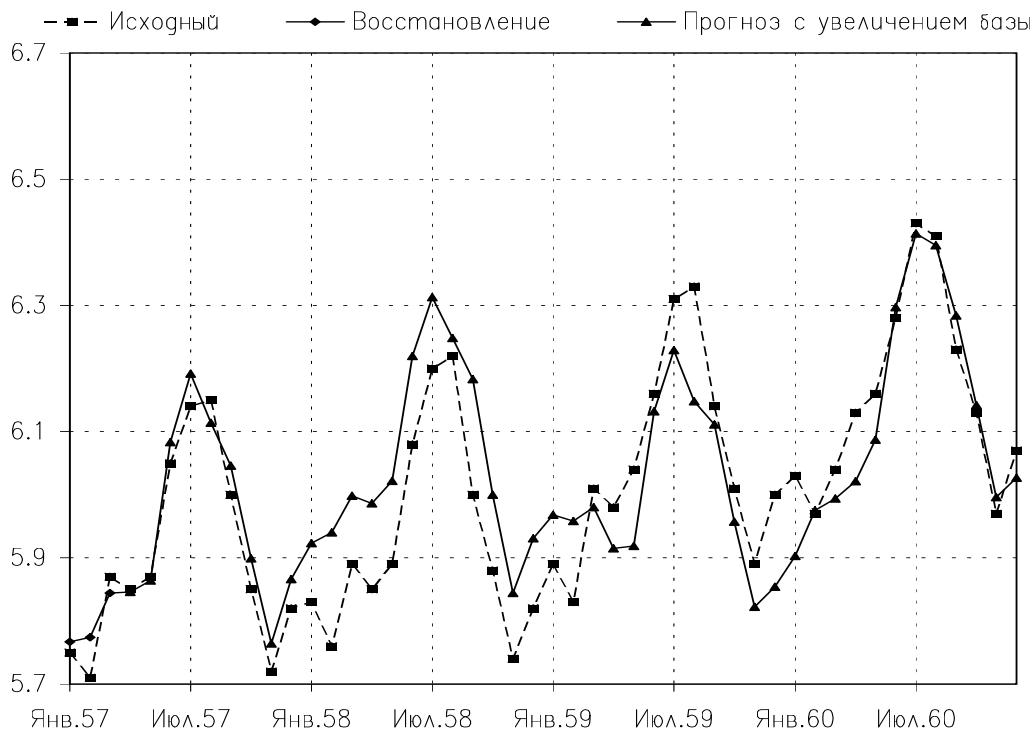


Рис. 11.1. Исходный ряд, восстановление и прогноз с увеличением базы.

Двумерные проекции колебательных главных компонент приведены на Рис. 6.1-7.1. Переходим к описанию результатов восстановления по поднаборам главных компонент. Восстановление по первой ГК дает ряд, близкий к линейно возрастающему, однако добавление шестой ГК позволило несколько лучше учесть форму тренда (слабый изгиб приблизительно после половины ряда), см. Рис. 9.1. Восстановление по колебательным главным компонентам дает нам синусоидальные кривые правильной формы с почти постоянными амплитудами, см. Рис. 8.1. Первые 8 ГК практически полностью описывают ряд (99.992% дисперсии) и согласование восстановленного ряда и исходного исключительно хорошее. Отметим, однако, что некоторые младшие ГК также дают гармоники достаточно правильной формы.

На основе первых восьми ГК производился ретропрогноз по участку с октября 1952 по апрель 1957 на 42 точки вперед (по декабрь 1960). Форма прогнозируемого ряда практически полностью воспроизводит форму исходного, однако имеется некоторое расхождение в тренде (прогноз приподнят.) см. Рис. 10.1. На следующем этапе проводился ретропрогноз "с увеличением базы". Здесь на каждом шаге проводился прогноз на шесть точек вперед, затем на график заносилась последняя, шестая точка, а база прогноза увеличивалась шагом на одну точку. Как видно из сравнения, в начале прогнозирования имеются некоторые расхождения, которые уменьшаются с ростом базы прогноза. Последние спрогнозированные точки практически точно ложатся на исходные данные, см. Рис. 11.1.

1.2 Анализ классическими методами

Классический подход к анализу временного ряда X_t , где t пробегает некоторый набор индексов, подразумевает применение следующей модели

$$X_t = m_t + S_t + Y_t,$$

где

- S_t — сезонная компонента, т.е. периодическая компонента с заранее известным периодом. Чаще всего это период, соответствующий году, месяцу, или неделе (в зависимости от данных).
- m_t — тренд, т.е. достаточно устойчиво и медленно меняющаяся часть ряда. Под трендом обычно понимают компоненту ряда, приблизительно соответствующую некоторому полиному (небольшой степени), часто периодические составляющие с частотами не кратными годовой, также могут быть отнесены к тренду.
- Y_t — стационарный ряд.

Для реализации упомянутой выше схемы проводились следующие классические обработки.

1. Комбинация 12-точечного скользящего среднего (инвариантного относительно сезонной компоненты) и 2-х точечного усреднения (компенсация сдвига по времени) позволяет выделить тренд, в целом близкий к линейному, однако имеющий 3 достаточно ясно видные "ступеньки", см. Рис. 12.1.
2. Следующим шагом является непараметрическое оценивание сезонной компоненты. $\hat{S}_t = \frac{1}{\#\{t=12m\}} \sum_{t=12m} x_t$, $S_t = \frac{1}{12} \sum_{t=1}^{12} \hat{S}_t$. Полученная таким образом оценка (на интервале 1..12) периодически продолжена на весь интервал см. Рис. 12.1.
3. Выделенная описанным образом сезонная компонента исследована с помощью дискретного преобразования Фурье. Выборочная периодограмма показывает наличие колебательных компонент с периодами (в мес.): 12, 6, 4, 3, 2.4, см. Рис. 15.1. Указанные 5 периодов полностью описывают сезонную компоненту, см. Рис. 17.1.
4. Полученный при вычитании тренда и сезонной компоненты остаток выглядит, в целом, стационарным (см. Рис. 12.1) и исследуется классическими методами теории стационарных временных рядов.
5. Выборочная периодограмма остатка показывает явное наличие колебания с периодом 12 мес., а также, возможно, с периодами около 8 и 6 мес. Для анализа проведено сглаживание периодограммы комбинацией взвешенных скользящих средних. Указанное сглаживание выделяет очень хорошо различимый пик на 12 мес., а также подъемы около 6 и 8 мес., см. Рис. 13.1.
6. Выборочная автокорреляционная функция подтверждает наличие значимых автокорреляций (т.е. ряд не является белым шумом). Кроме того,

форма автокорреляционной функции позволяет заподозрить наличие колебания с периодом около 12 мес., что соответствует результатам спектрального анализа, см. Рис. 14.1.

7. Наличие значимых автокорреляций и присутствие скрытых периодик позволяет предположить, что адекватной моделью в данном случае является авторегрессионный процесс. Критерии выбора авторегрессионной модели дают порядок 8 (AR(8)). В качестве критериев использовались AIC (информационный критерий Акаике), BIC (байесовский информационный критерий Акаике) и CAT (критерий Парсена). См. также [3]. Рис. 16.1.
8. Подогнанная автокорреляционная функция в целом очень похожа на выборочную. Спектр Юла Уолкера близок к сглаженной периодограмме в районе 12 мес., но несколько расходится вблизи двух слабых пиков около 6 и 8 мес. /спектр дает максимумы на частотах 4.5 мес. и 3 мес./, см. Рис. 18.1.

Таким образом произведено разложение исходного процесса в сумму $X_t = m_t + S_t + Y_t$, где

m_t — близкий к линейному тренд,

S_t — сезонная компонента: сумма периодик с периодами 12, 6, 4, 3, 2.4 мес.

Y_t — авторегрессионный процесс 8 порядка.

1.3 Сравнительный анализ

Наиболее интересен сравнительный анализ найденных периодик. Простое сравнение частот дает нам следующее соответствие между главными компонентами и Фурье периодиками.

ГК, №	Преобразование Фурье
2-3	12 мес.
4-5	6 мес.
9-10	4 мес.
11-12	3 мес.
13-14	2.4 мес.

Для детального сравнения этих колебаний потребовалось использование обратного преобразования Фурье. Более точно: после определения местоположения пика в Фурье спектре, производилось обратное преобразование Фурье по этой части спектра. Таким образом были получены 5 вещественных периодических временных рядов, образующих полное разложение сезонной компоненты, см. Рис. 17.1.

Построив указанные временные ряды совместно с восстановлениями исходного ряда по главным компонентам, удалось установить практически полную идентичность указанных в таблице Фурье компонент и восстановлений по соответствующим (см. таблицу выше) главным компонентам (Рис. 23.1,24.1). Отметим, однако, слабое увеличение амплитуды годовой периодики с течением времени. Этого не наблюдается у остальных колебательных компонент.

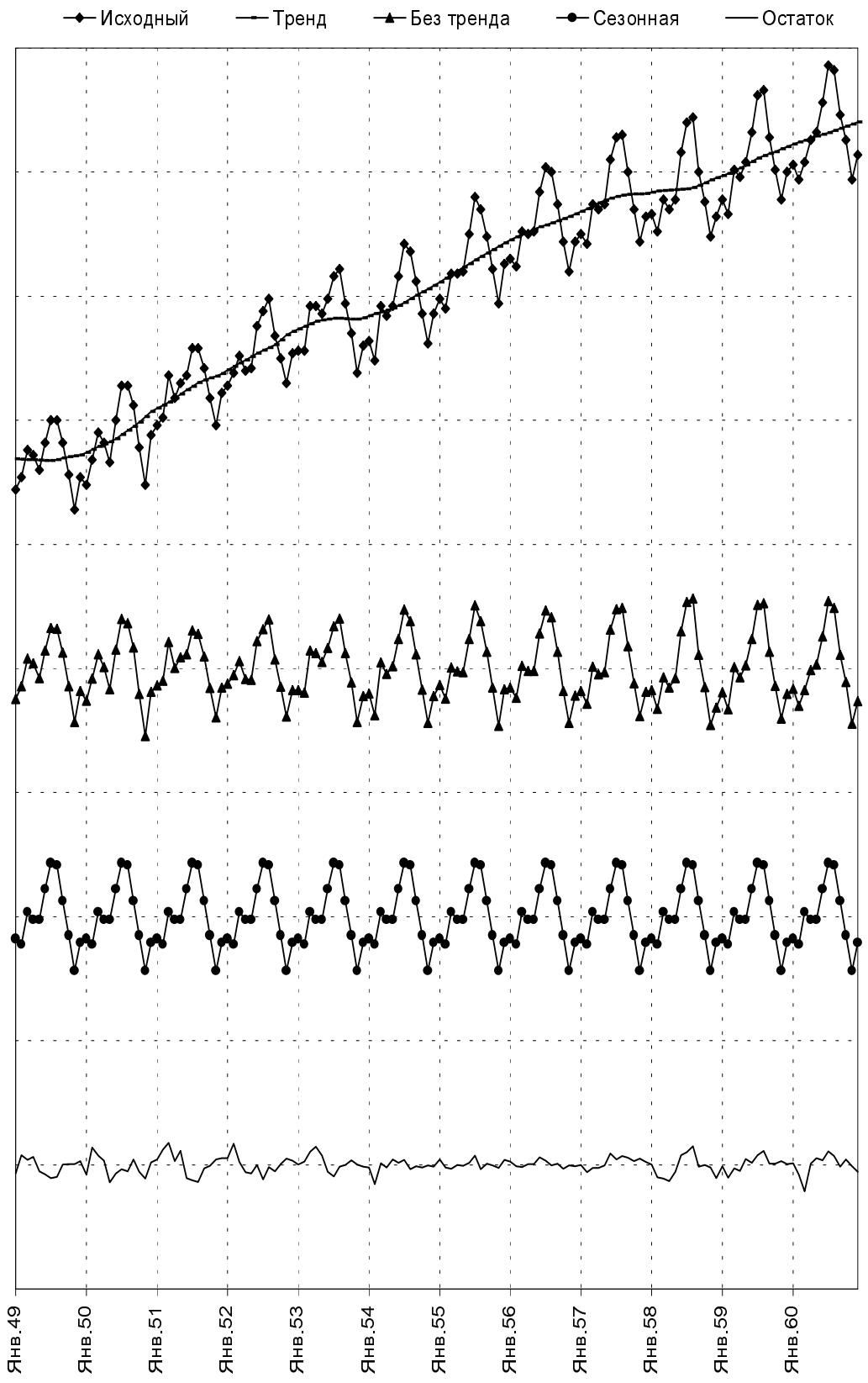


Рис. 12.1. Тренд (скользящее среднее), сезонная компонента, ряд со снятым трендом и остаток.

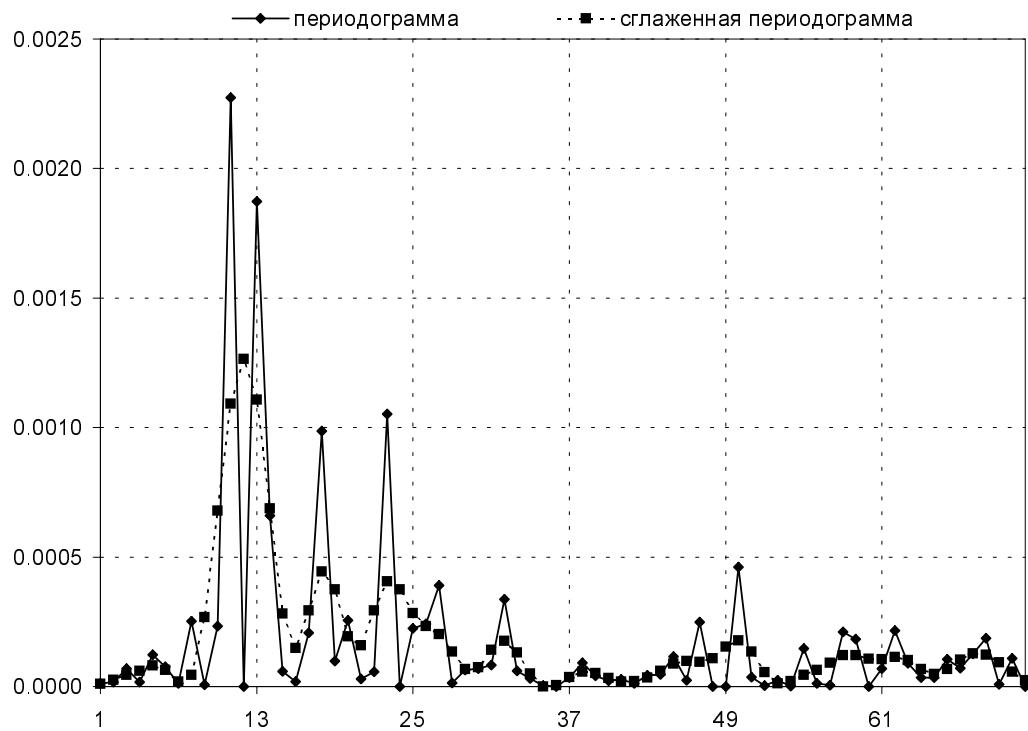


Рис. 13.1. Периодограмма ряда остатков и сглаженная периодограмма.

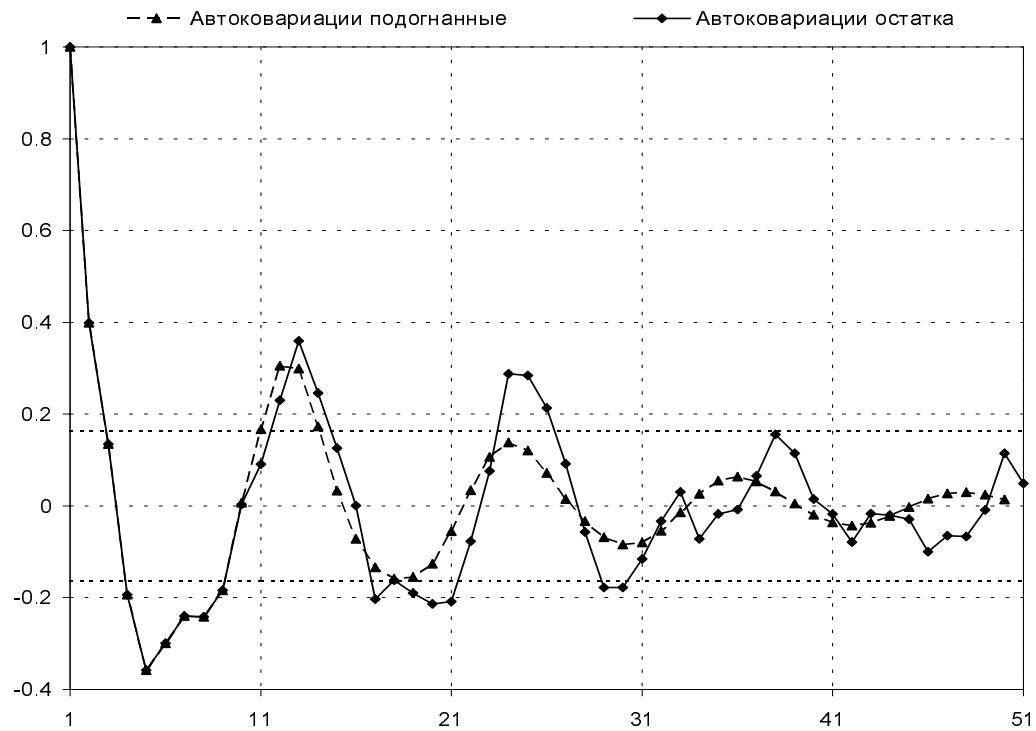


Рис. 14.1. Автоковариационные функции ряда остатков и подогнанного $AR(8)$ процесса.

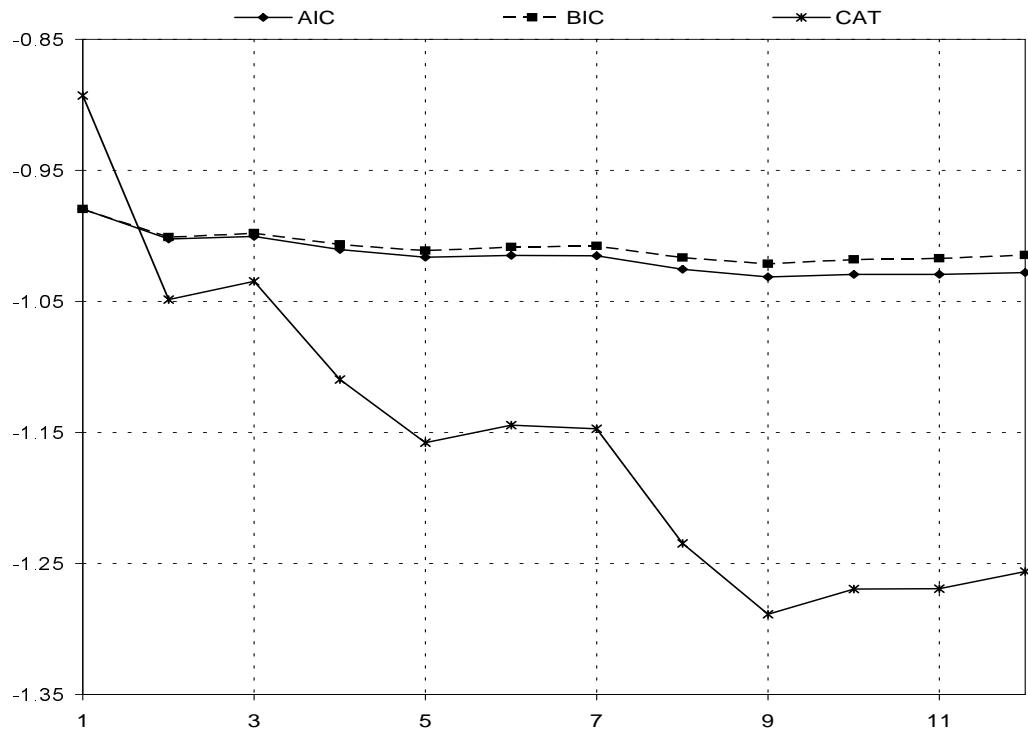


Рис. 15.1. Периодограмма сезонной компоненты.

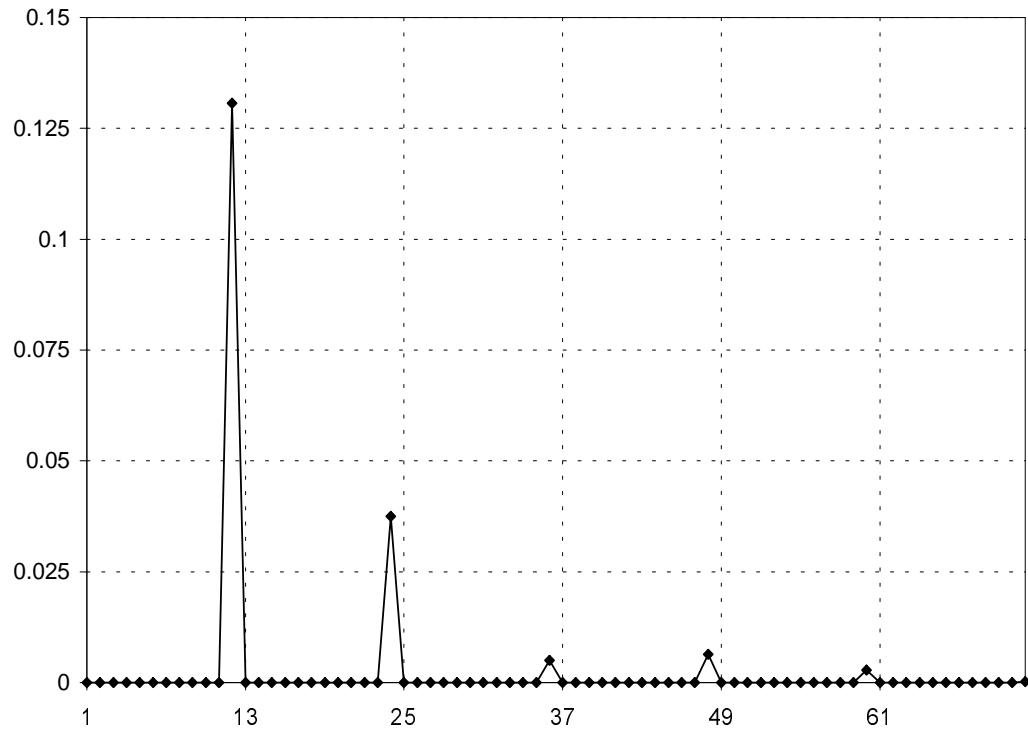


Рис. 16.1. Критерии выбора порядка авторегрессионной модели.

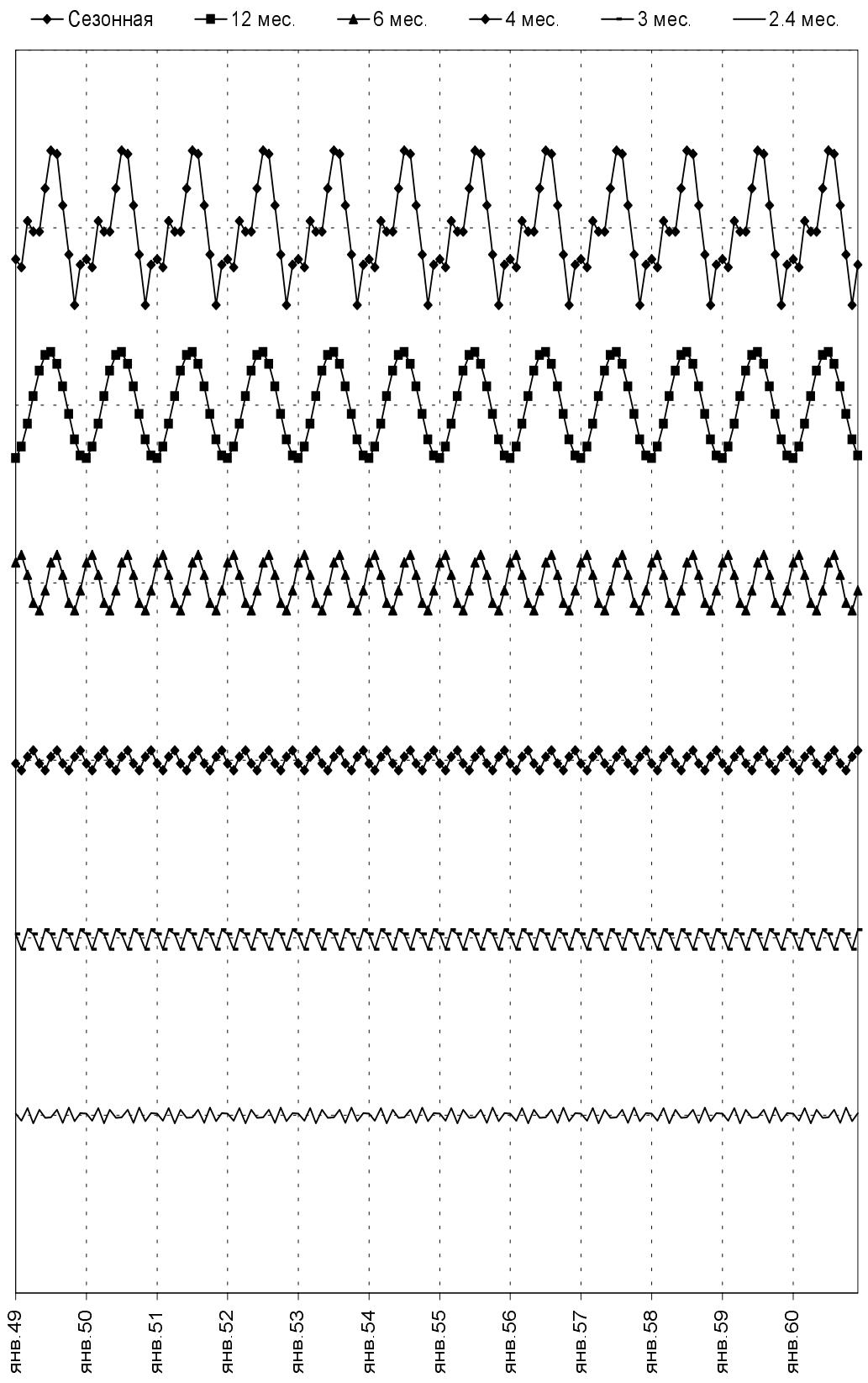


Рис. 17.1. Разложение сезонной компоненты в Фурье гармоники.

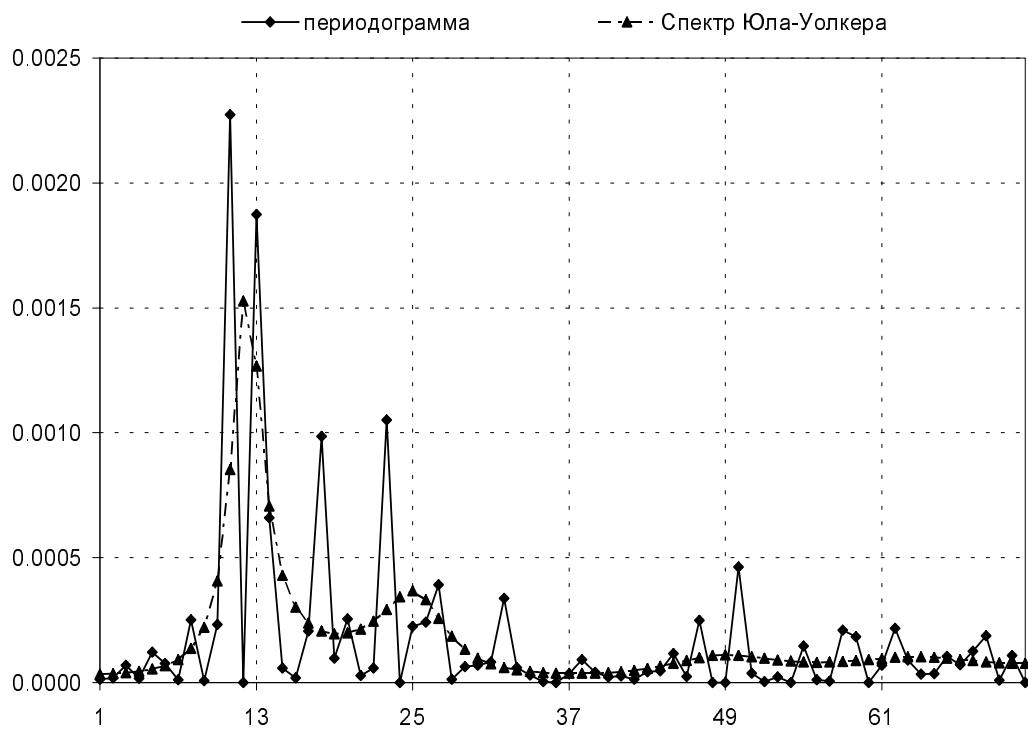


Рис. 18.1. Периодограмма остатка и спектр Юла-Уолкера.

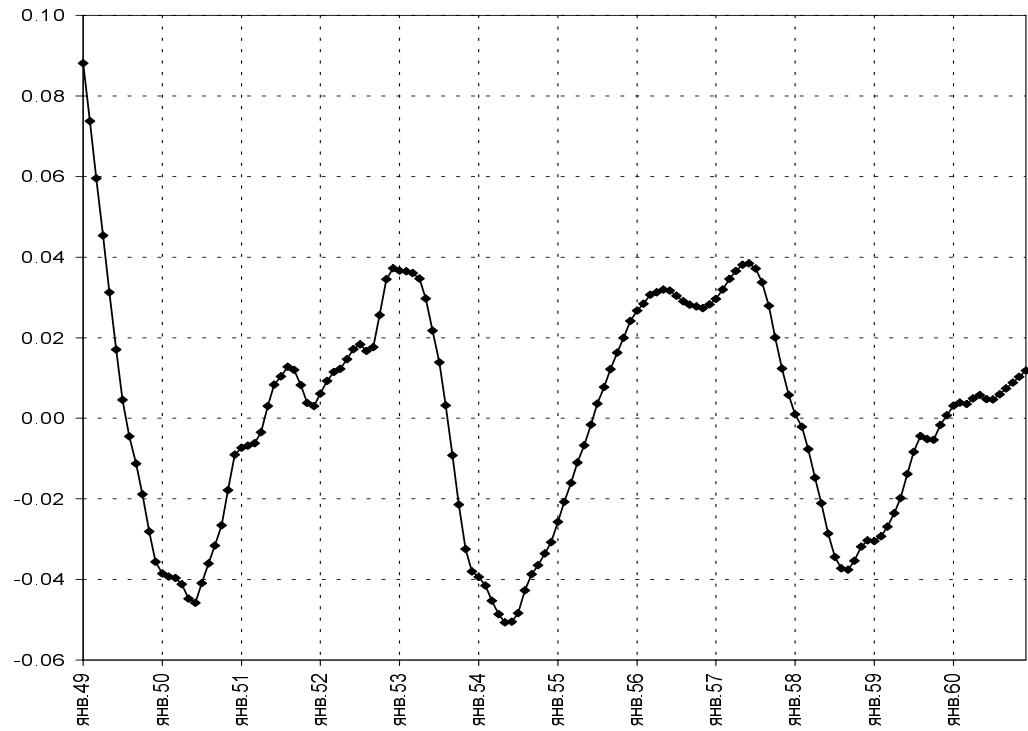


Рис. 19.1. Остаток от подгонки тренда скользящим средним полиномом второй степени.

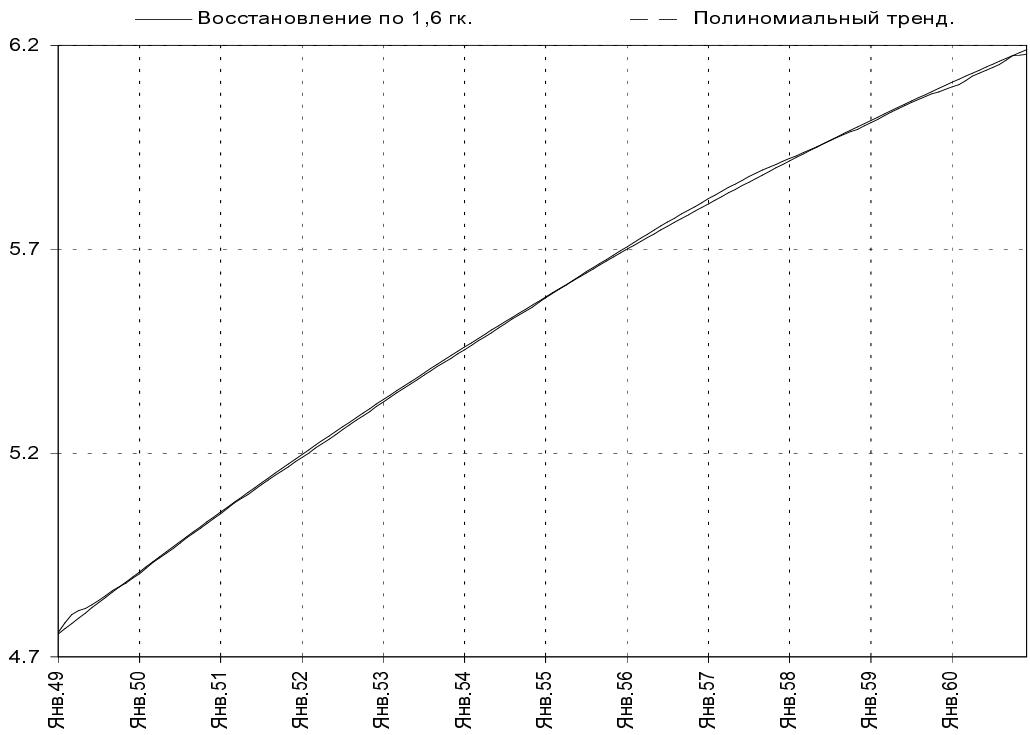


Рис. 20.1. Сравнение восстановления по 1,6 ГК с подгонкой тренда скользящим полиномом второй степени.

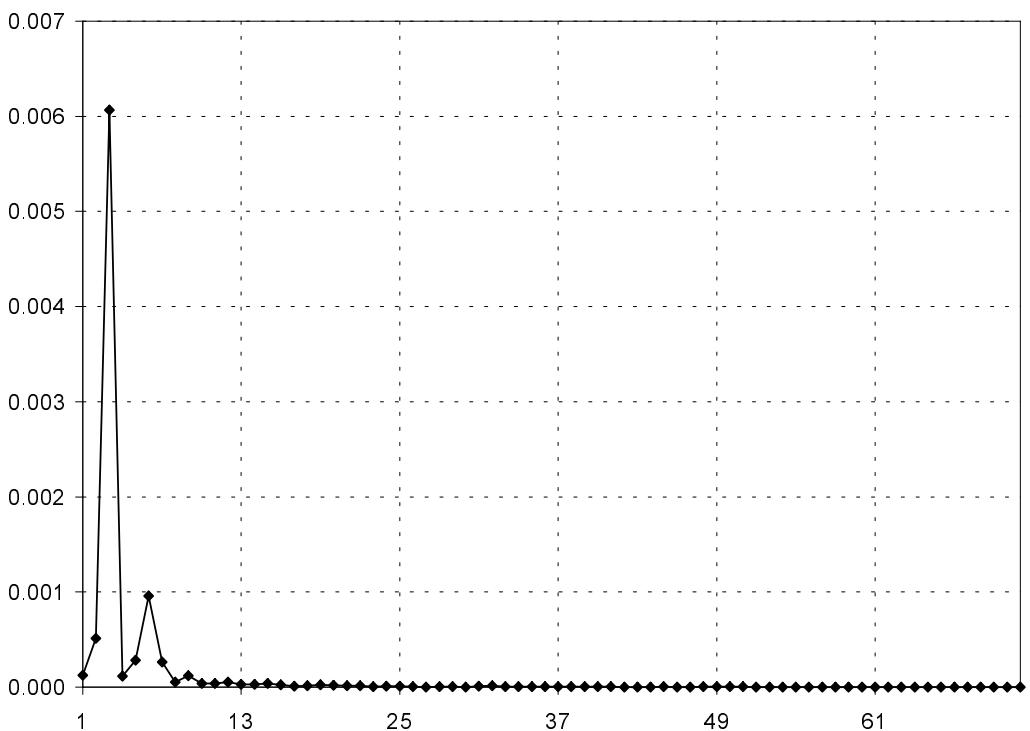


Рис. 21.1. Периодограмма остатка от подгонки тренда.

■ Восстановление по 2-3 гк. ◆ Восстановление по 4-5 гк. ✕ 12 мес. ✖ 6 мес.

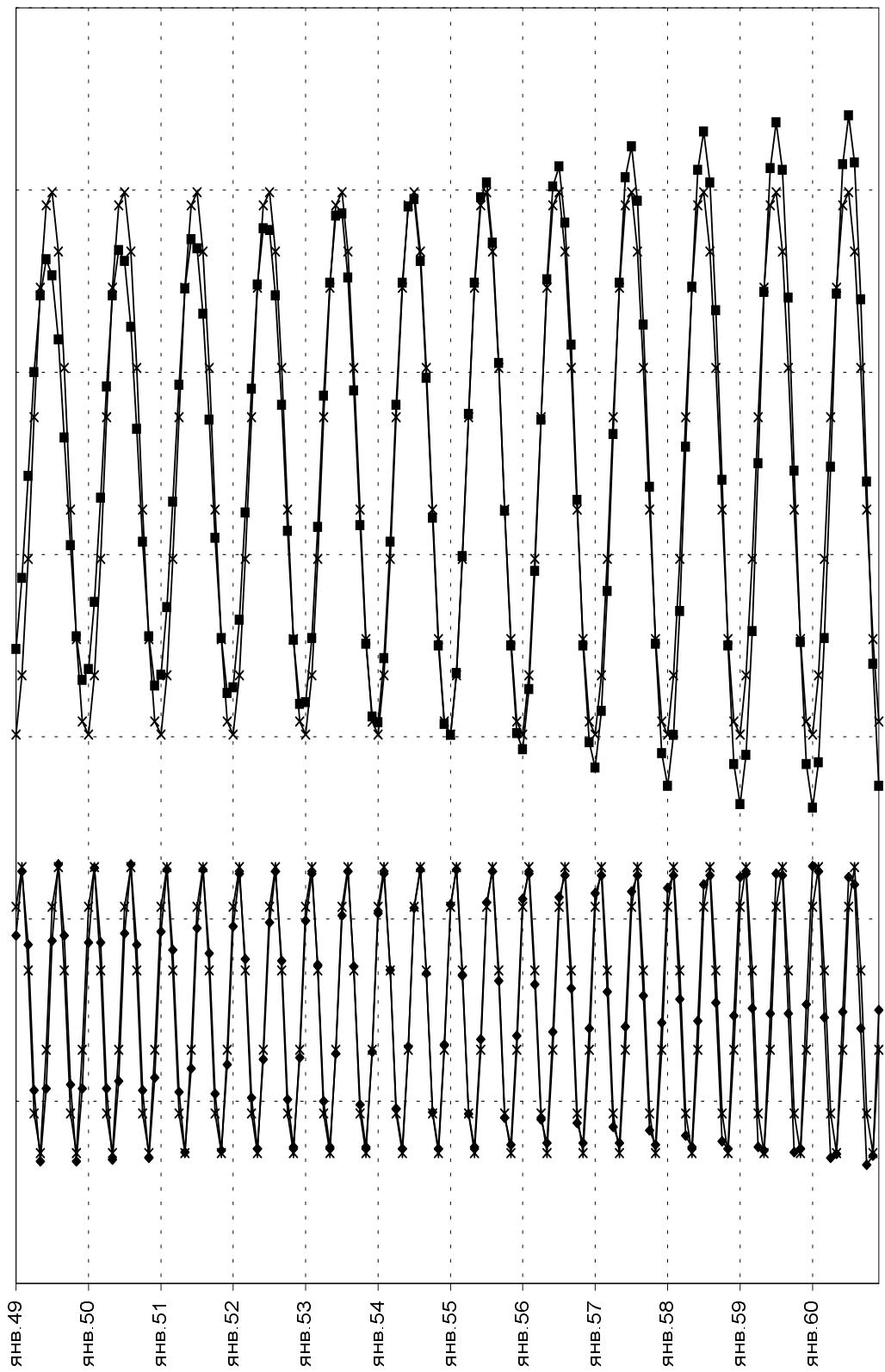


Рис. 22.1. Сравнение низкочастотных периодик, выделенных
Фурье анализом и "Гусеницей".

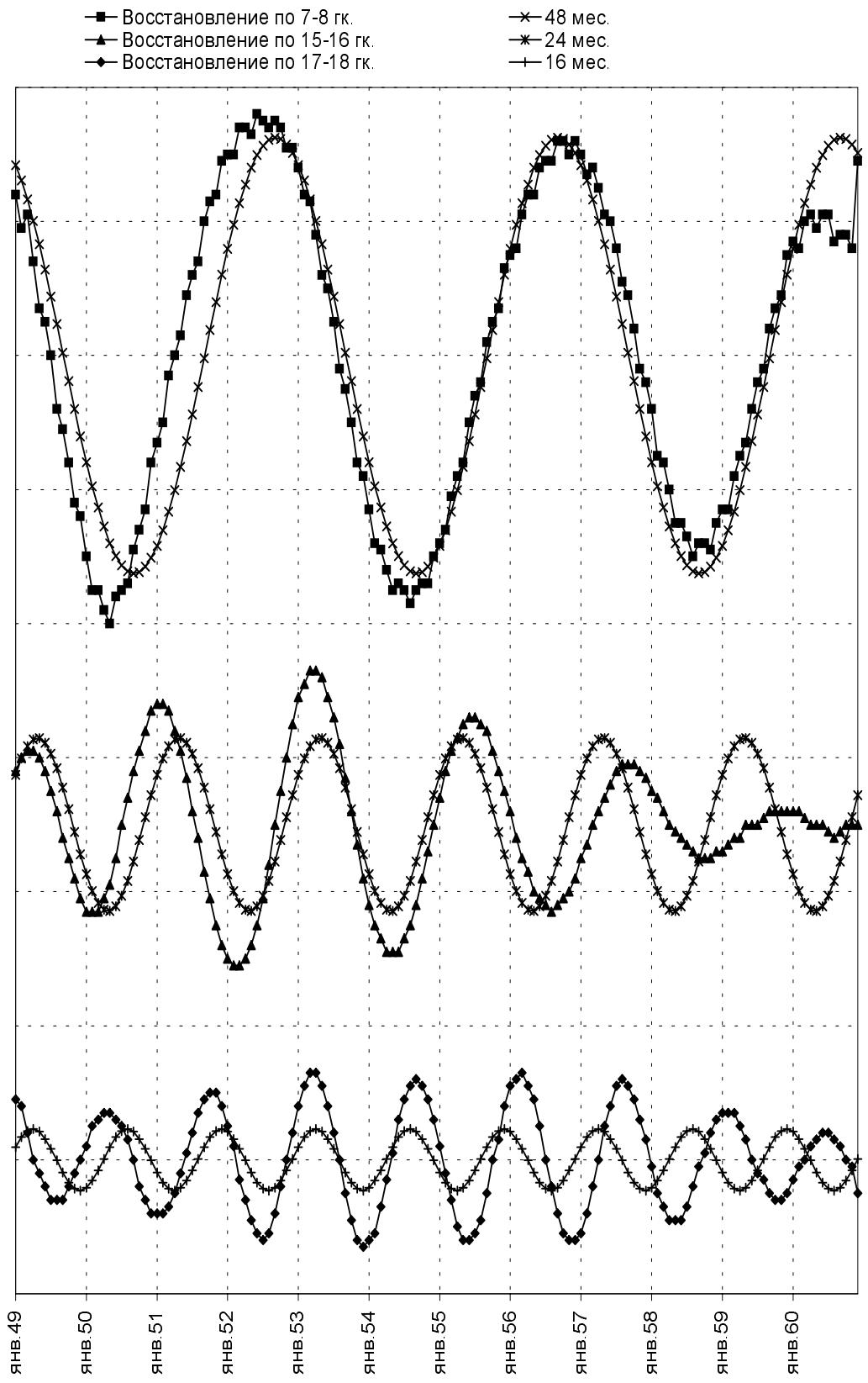


Рис. 23.1. Сравнение гармоник сезонной компоненты, выделенных Фурье анализом и "Гусеницей".

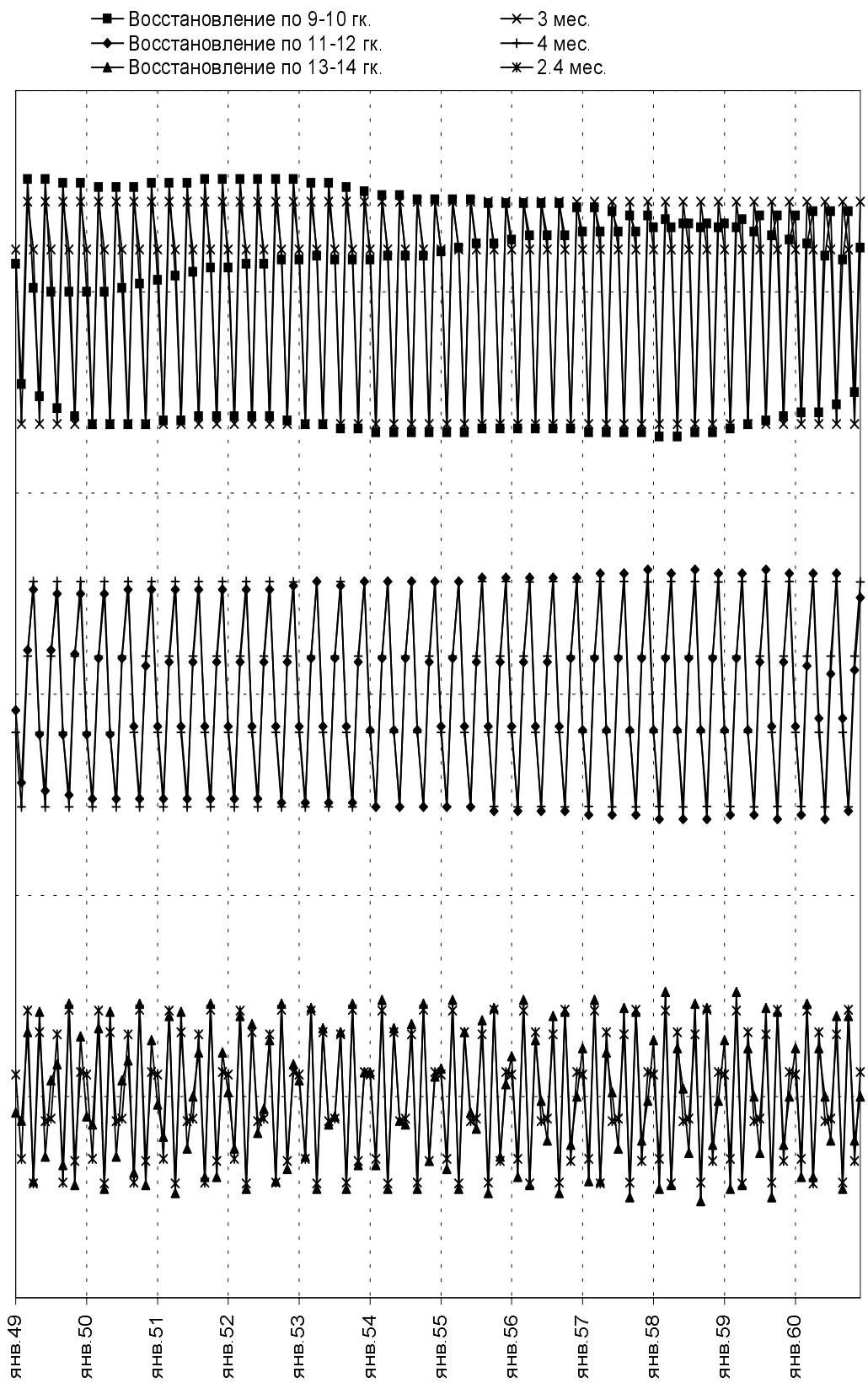


Рис. 24.1. Сравнение гармоник сезонной компоненты, выделенных Фурье анализом и "Гусеницей".

Заметим, что не все виды колебаний отслеживаются классическими методами. Так, колебания с периодами 50 мес. (7-8), 27 мес. (15-16), 18 мес. (17-18), не обнаруживаются преобразованием Фурье, очевидно, войдя в тренд.

Для выделения указанных периодик была произведена следующая обработка тренда, выделенного скользящими средними:

полином второй степени был подогнан методом наименьших квадратов, в качестве критерия выбора порядка использовалось эмпирическое правило: выбрать полином небольшой степени, для которого ряд остатков меняется приблизительно равномерно на всем временном интервале, см. Рис. 19.1, 20.1.

к ряду остатков опять применен анализ Фурье, см. Рис. 21.1. Наблюдаемые пики спектра говорят о наличии колебаний с периодами 48, 24, и около 16 мес.

произведено обратное преобразование Фурье, применение которого позволяет сравнить эти три обнаруженных периодики с восстановлениями по наборам ГК 7-8, 15-16, 17-18., см Рис. 22.1. Сразу видно очень хорошее соответствие 48 мес. периодики, несколько худшее у 24 мес. периодики где ГК дают ряд с заметно большим периодом, амплитуда которого меняется сложным образом, и еще худшее у самой слабой и высокочастотной компоненты, где и частота и амплитуда меняются существенно.

1.4 Общие выводы

Многие существенные черты изученного ряда хорошо отражены как анализом главных компонент, так и классическими методами. Особенно это касается полиномиального (квадратичного) тренда и относительно высокочастотных колебаний с периодами 12, 6, 4, 3, 2.4 мес. Различия начинаются при изучении относительно слабых и низкочастотных колебательных компонент, где адаптивность "гусеницы" начинает играть большую роль. Действительно, колебания, описываемые 15-16 и 17-18 ГК явно имеют периоды не 24 а около 27 мес, и не 16, а 18 мес. соответственно. Это находит косвенное подтверждение в широких спектральных пиках соответствующих колебаний. Похоже, что в этом случае анализ Фурье неадекватен. Аналогичное утверждение можно сделать и относительно годовой периодики. Изменение (хотя и слабое) амплитуды этого (ведущего) колебания может быть замечено даже невооруженным глазом на исходном ряде, но отсутствует, естественно, у Фурье компоненты. Кроме того, следует отметить, что классический анализ требует привлечения очень широкого круга различных методов, часто даже не вполне формализованных, в то время как "гусеница" позволяет произвести указанный анализ быстро, адаптивно и в едином контексте.

2 Заболеваемость СПИДом в США 1982-1988 гг.

Рассматриваемый набор данных представляет заболеваемость СПИДом в США. Наблюдения ежеквартальные, одна точка — 3 месяца, всего приводится 25 точек, с первого квартала 1982 по первый квартал 1988 включительно. Данные опубликованы в статье [6], Table I, и представляют интерес с точки зрения эпидемиологии. Метод гусеница в данном случае работает одинаково хорошо как по исходным, так и по преобразованным данным, однако в целях сравнительного анализа ряд был преобразован с помощью степенного преобразования описанного ниже.

2.1 Анализ методом "Гусеница"

Рассмотрим результаты применения нецентральной гусеницы с длиной $\tau = 12$. Таким выбором длины гусеницы достигается сразу две цели: 12 кратно 4, периоду сезонной (годовой) компоненты, и это достаточно большая длина (половина длины ряда). Исходный ряд и основные выделенные компоненты представлены на Рис. 1.2. Главные компоненты приведены на Рис. 2.2-3.2. Результаты анализа главных компонент кратко могут быть представлены следующей таблицей.

ГК, №	Интерпретация	Рис. №
1	нелинейный тренд	1.2
2	нелинейный тренд	1.2
3-4	годовая периодика (сезонная компонента)	1.2-2.2

Восстановление по 1-2 г.к. дает гладкую возрастающую кривую, близкую к степенной функции. Некоторая немонотонность на конце позволяет предположить снижение скорости роста заболеваемости начиная с конца 1987, начала 1988 гг. Разность между исходным рядом и восстановлением по 1-2 гк. очень хорошо приближается восстановлением по 3-4 гк., см. Рис. 3.2. Интересно отметить, что годовой максимум заболеваемости приходится на второй квартал, что соответствует весенним и летним месяцам, а минимум — на четвертый, соответствующий осени и зиме.

Ретропрогноз по 20 точкам (по 2 квартал 1986 года) на 6 точек (период с 3 квартала 1986 по 4 квартал 1987). Прогноз по 1-2 гк. дает полное согласование с наблюдаемыми данными, см. Рис. 4.2.

Другой вариант обработки тех же исходных данных приведен во введении. В этом случае гусеничной обработке подвергался ряд, составленный из разностей исходного. При этом выделились тренд, годовая периодика, а также полугодовая гармоника, которая совместно годовой образует "сезонную" компоненту.

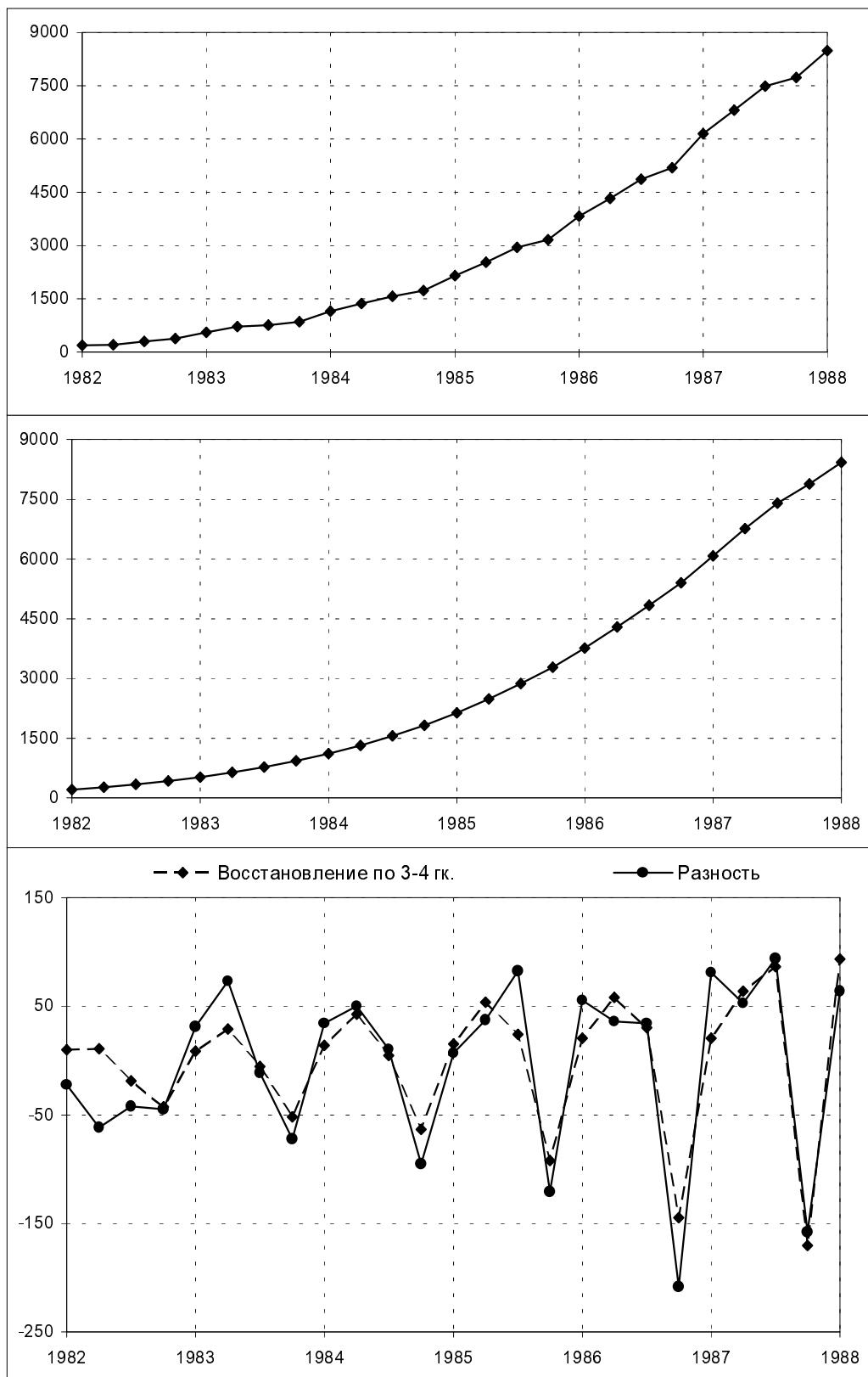


Рис. 1.2. Исходный ряд (верхний), тренд (середина) и сравнение выделенной годовой периодики с остатком после вычета тренда (нижний).

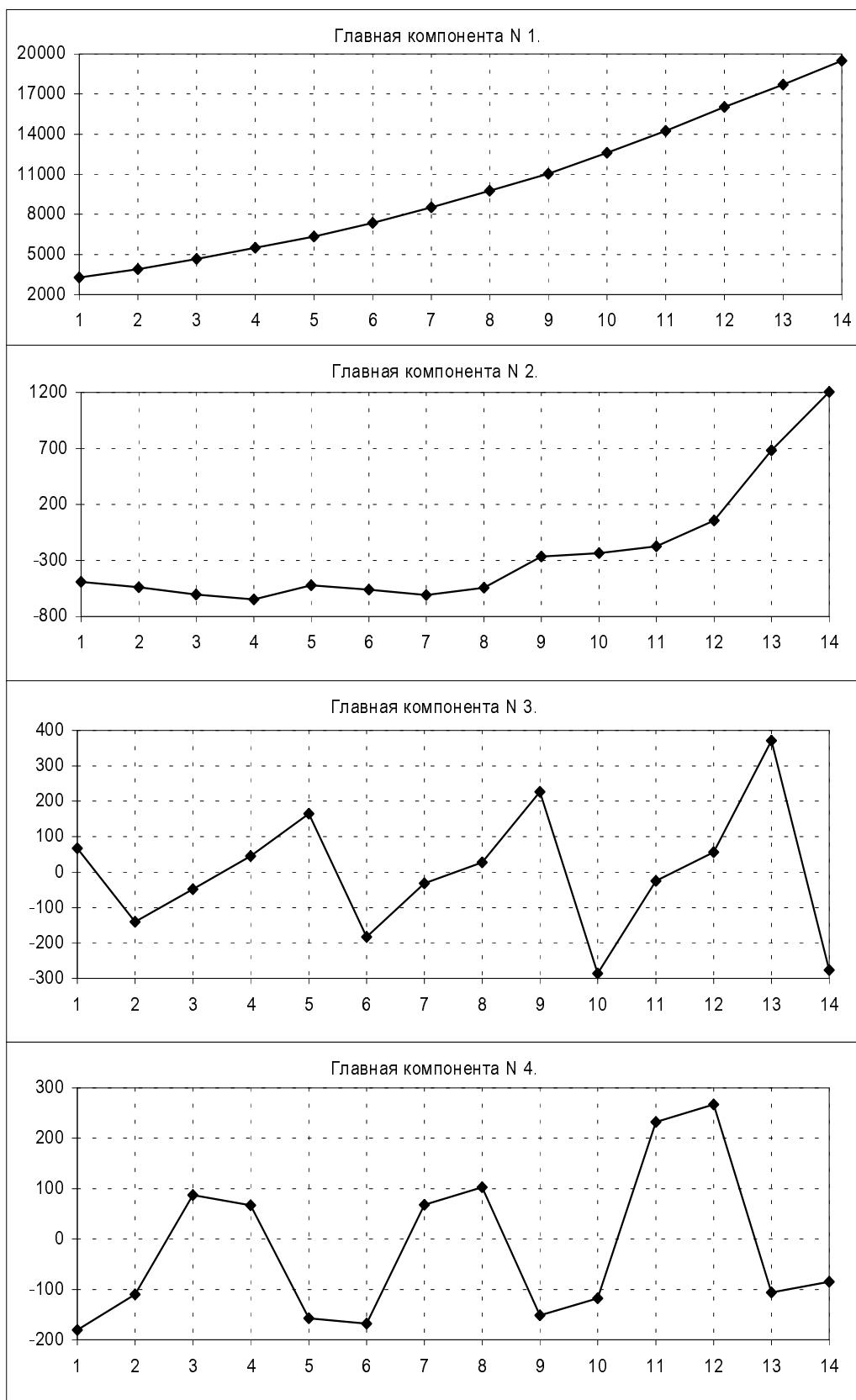


Рис. 2.2. Информативные главные компоненты.

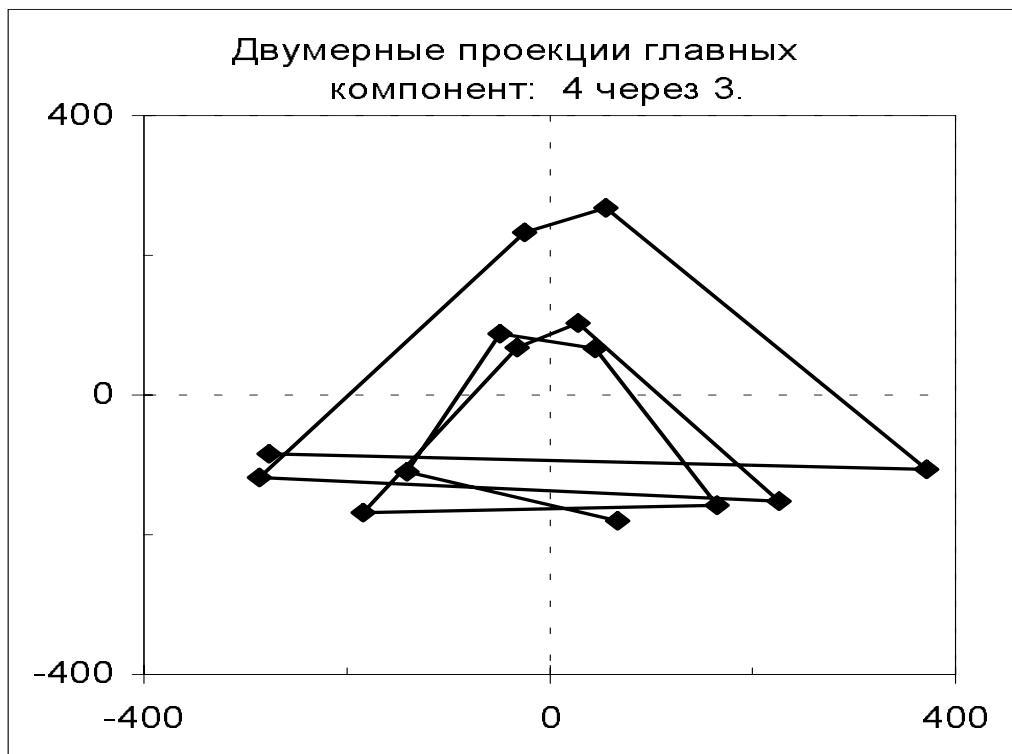


Рис. 3.2. Двумерная проекция ГК 3 и 4.

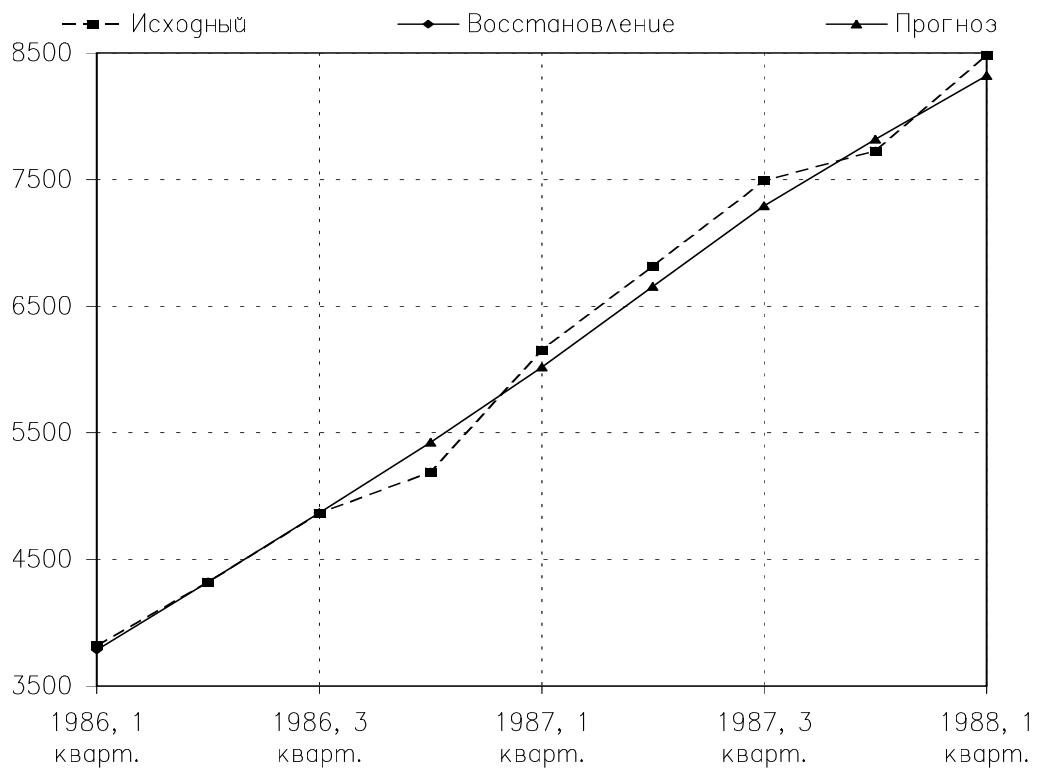


Рис. 4.2. Ретропрогноз и исходный ряд.

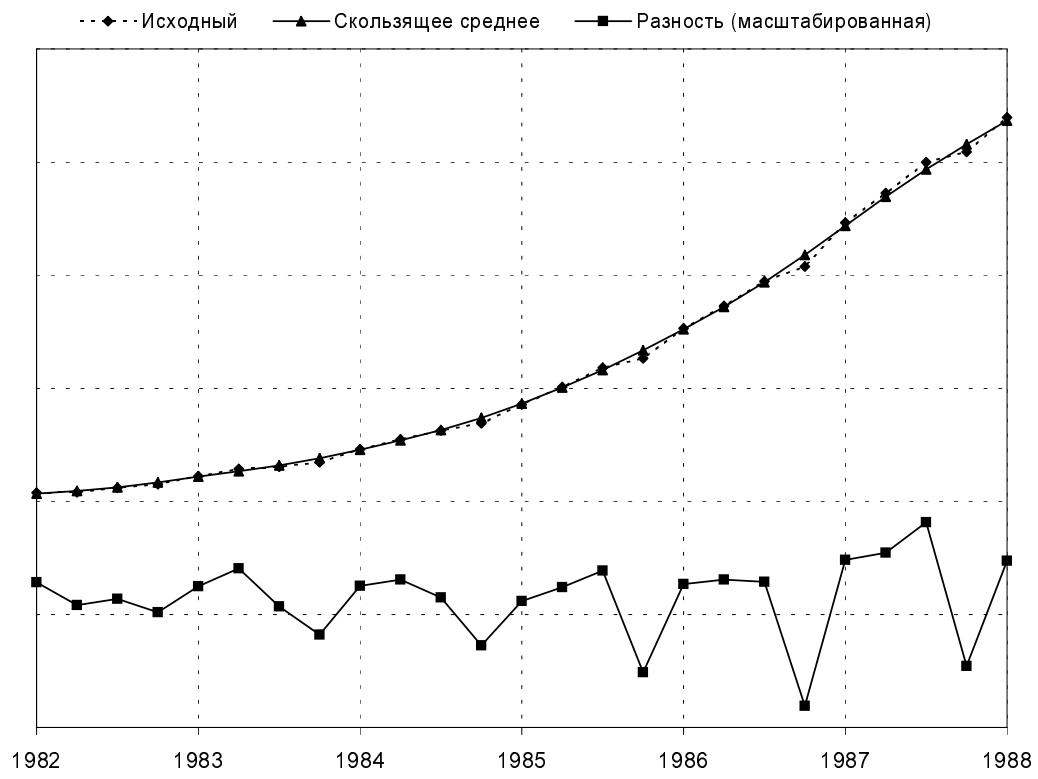


Рис. 5.2. Выделение тренда (скользящее среднее) из исходных данных.

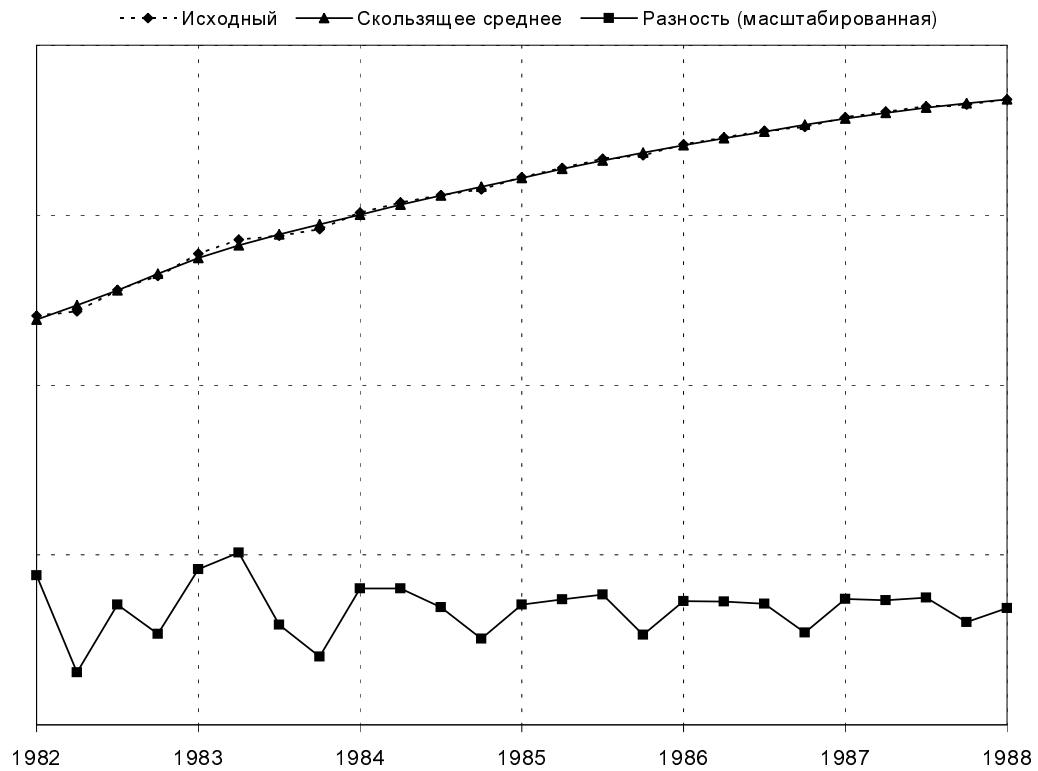


Рис. 6.2. Выделение тренда (скользящее среднее) из логарифмированного ряда.

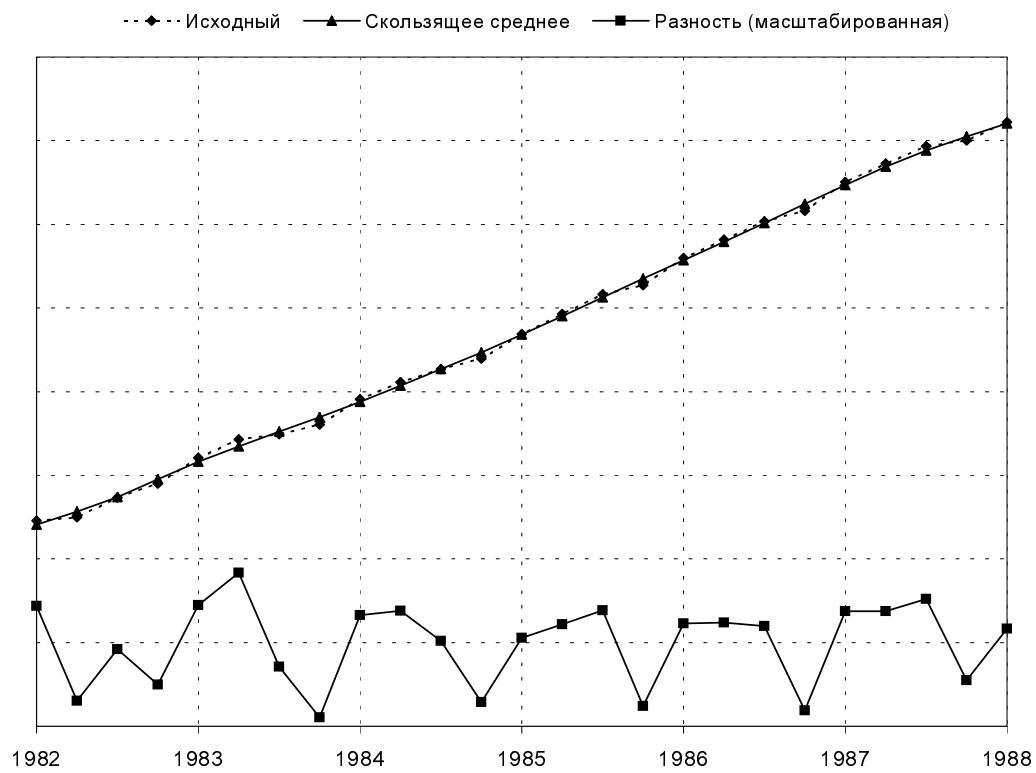


Рис. 7.2. Выделение тренда (скользящее среднее) из ряда, преобразованного согласно формуле $Y = X^{0.38}$.

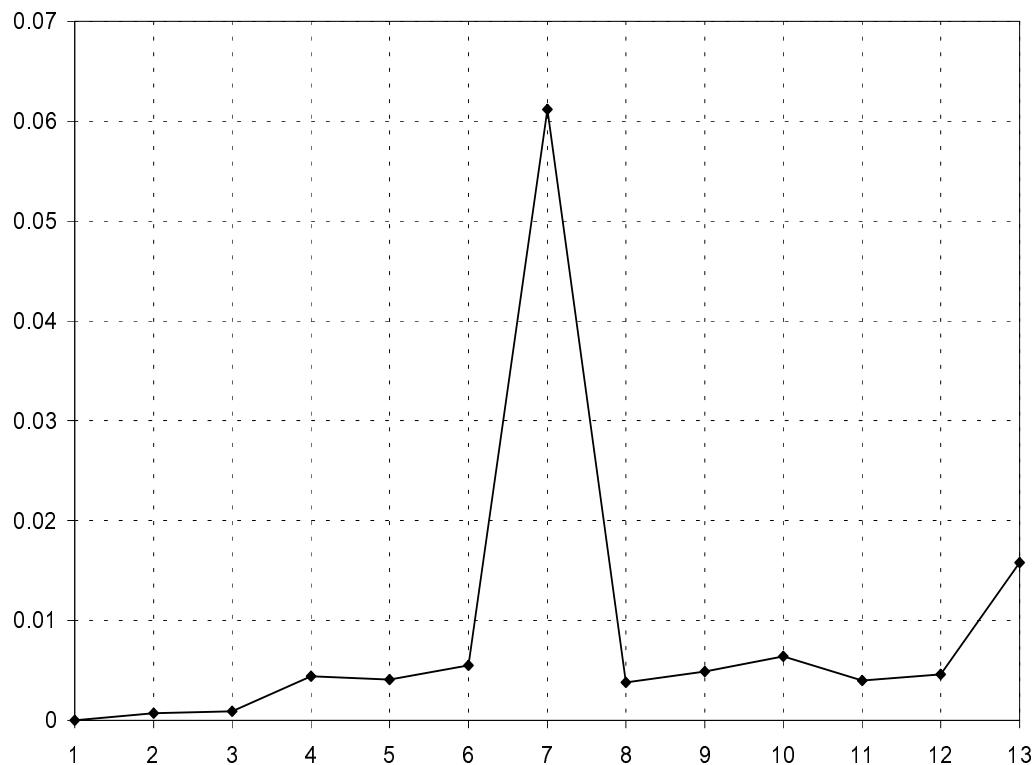


Рис. 8.2. Периодограмма остатка от вычитания тренда из исходного ряда.

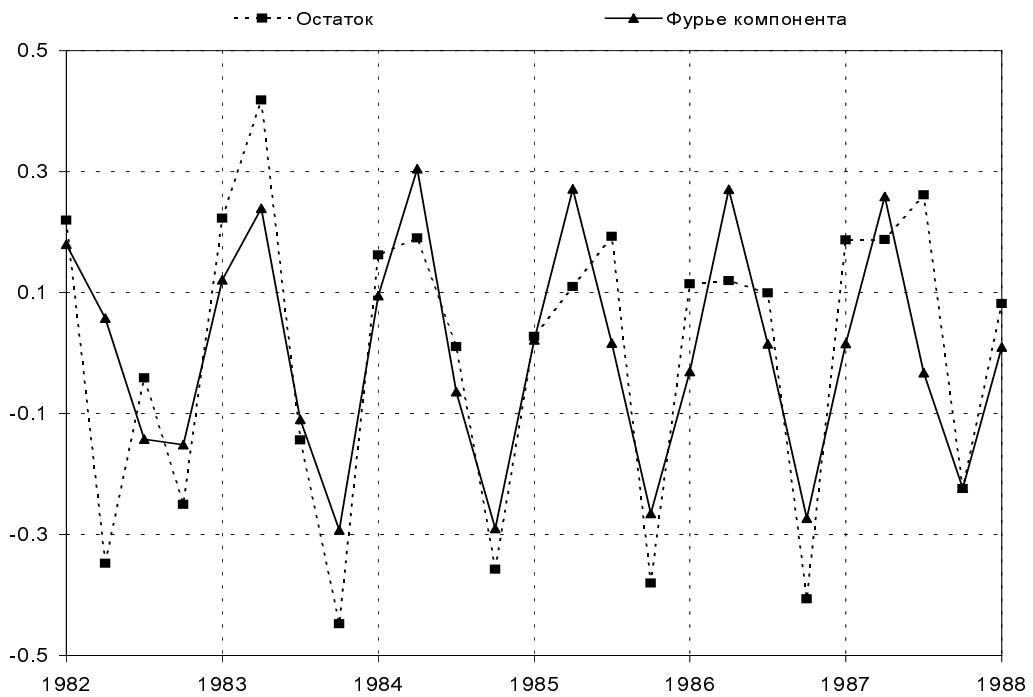


Рис. 9.2. Сравнение Фурье компоненты с рядом остатков.

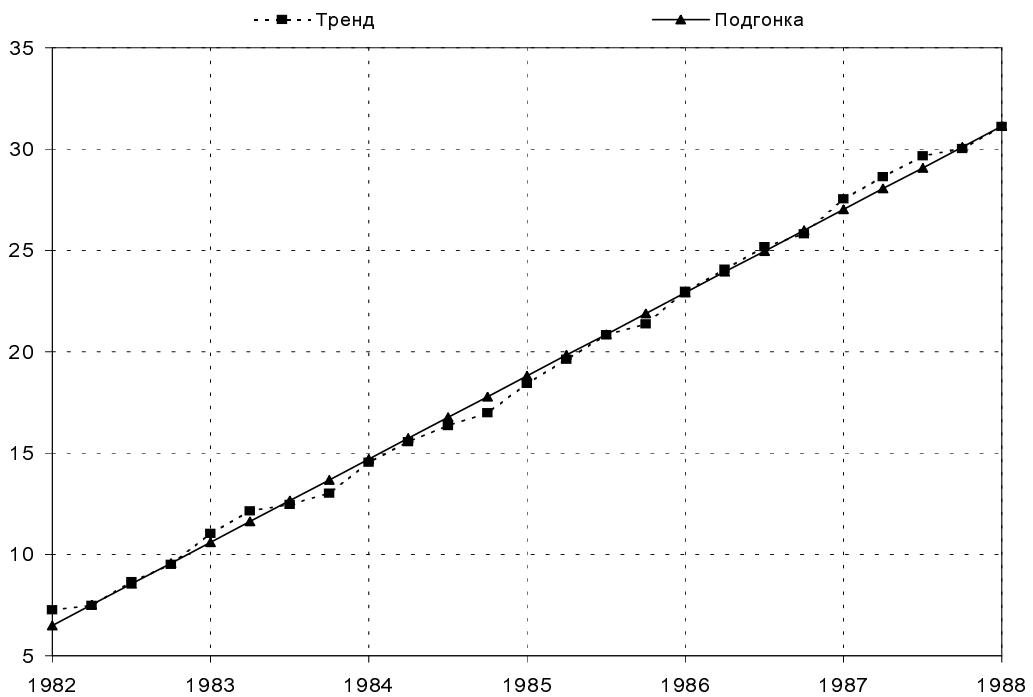


Рис. 10.2. Подгонка тренда (скользящего суммирования) линейной функцией.

2.2 Анализ классическими методами

Учитывая простоту исходного ряда, мы ограничимся выделением тренда и сезонной компоненты, т.е. слагаемых m_t и S_t в общей схеме

$$X_t = m_t + S_t + Y_t,$$

(здесь X_t — исходный ряд).

На первом этапе проведено выделение тренда с помощью комбинации 4-х точечного скользящего среднего (фильтр инвариантен относительно годовой периодики), и 2-х точечного усреднения (коррекция времени).

Однако при выделении тренда из исходных данных получаемый ряд остатков имеет переменную дисперсию, см. Рис. 5.2, поэтому применение преобразования, стабилизирующего дисперсию неизбежно. Однако обычное логарифмическое преобразование недостаточно хорошо работает, см. Рис. 6.2, поскольку некоторое изменение амплитуды ряда остатков все же заметно. Гораздо лучший результат дает степенное преобразование $Y = X^{.38}$, см. Рис. 7.2, которое не только стабилизирует амплитуду остатков, но и линеаризует тренд!

Полученный остаток исследуется с помощью разложения в ряд Фурье, периодограмма приведена на Рис. 8.2. Ясно заметен большой пик — колебание с периодом четыре точки (годовой цикл). Подъем в правой части периодограммы, возможно, означает присутствие гармоники годовой компоненты. Восстановление годовой Фурье периодики очень хорошо приближает ряд остатков, см. Рис. 9.2. В свою очередь, выделенный ранее тренд хорошо подгоняется линейной функцией $m_t = 1.03t + 5.47$, см. Рис. 10.2.

2.3 Общие выводы

К общим выделяемым особенностям рассмотренного ряда следует отнести годовую периодику и нелинейный тренд. Однако "гусеница" выделяет эти компоненты непосредственно из исходных данных, и позволяет непосредственно по ним же производить прогноз. С другой стороны классические методы позволяют аппроксимировать обработанный ряд в аналитической модели, что является компактной и наглядной формой представления результатов анализа.

3 Среднее число яиц на несушку на каждый месяц в США 1938, 1939 и 1940 гг.

Рассматриваемый временной ряд представляет среднее число яиц на несушку в США. Наблюдения ежемесячные, всего приводится 36 точек, что соответствует интервалу 1938-1940 гг. Данные опубликованы в статье [5] и широко использовались во многих классических учебниках по временными рядам как тестовые, см. [2]. Эти данные по характеру являются экономическими и интересны прежде всего как тестовый ряд для сравнения с Фурье анализом. Здесь мы имеем достаточно редкий случай, когда и "Гусеница", и анализ Фурье могут быть применены непосредственно к исходному ряду, который изображен на Рис. 1.3.

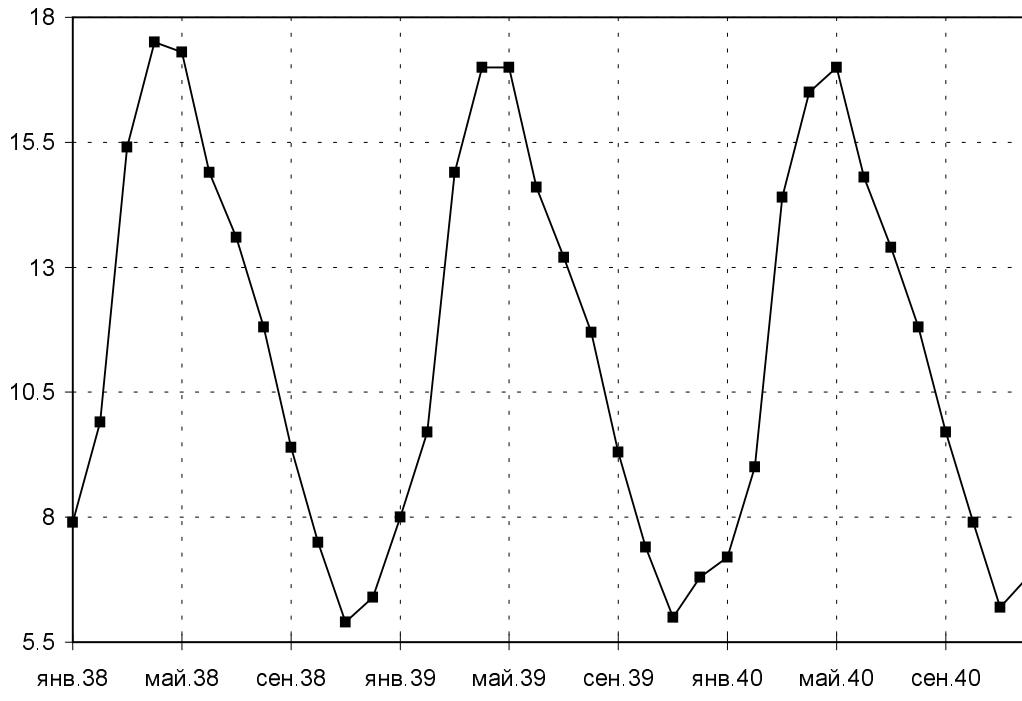


Рис. 1.3. Исходный ряд.

3.1 Анализ методом "Гусеница"

Рассмотрим результаты применения нецентральной гусеницы с длиной $\tau = 18$. Мы выбрали максимально возможную длину гусеницы, для того, чтобы достичь максимального разрешения. Главные компоненты приведены на Рис. 2.3-6.3. Результаты анализа главных компонент кратко могут быть представлены следующей таблицей.

ГК, №	Интерпретация	Рис. №
1	общее среднее	2.3
2-3	годовая периодика (сезонная компонента)	2.3
4-5	полугодовая периодика (гармоника сезонной комп.)	2.3-3.3
6-7	4 мес. колебание (гармоника сезонной компоненты)	3.3
8-9	2.4 мес. периодика (гармоника сезонной компоненты)	3.3-4.3
10-11	3 мес. периодика (гармоника сезонной компоненты)	4.3
12	2 мес. периодика (гармоника сезонной компоненты)	4.3

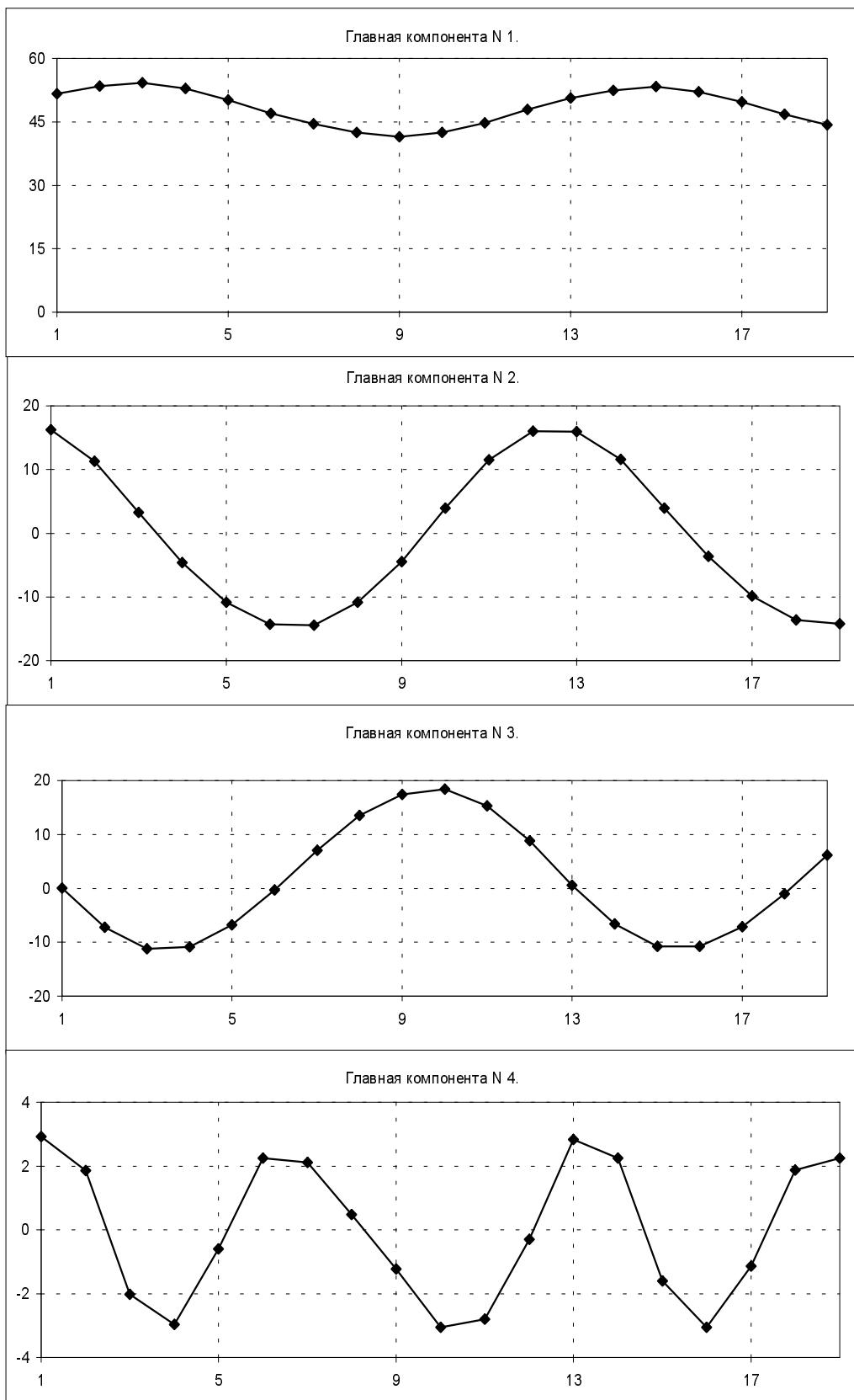


Рис. 2.3. Главные компоненты № 1-4.

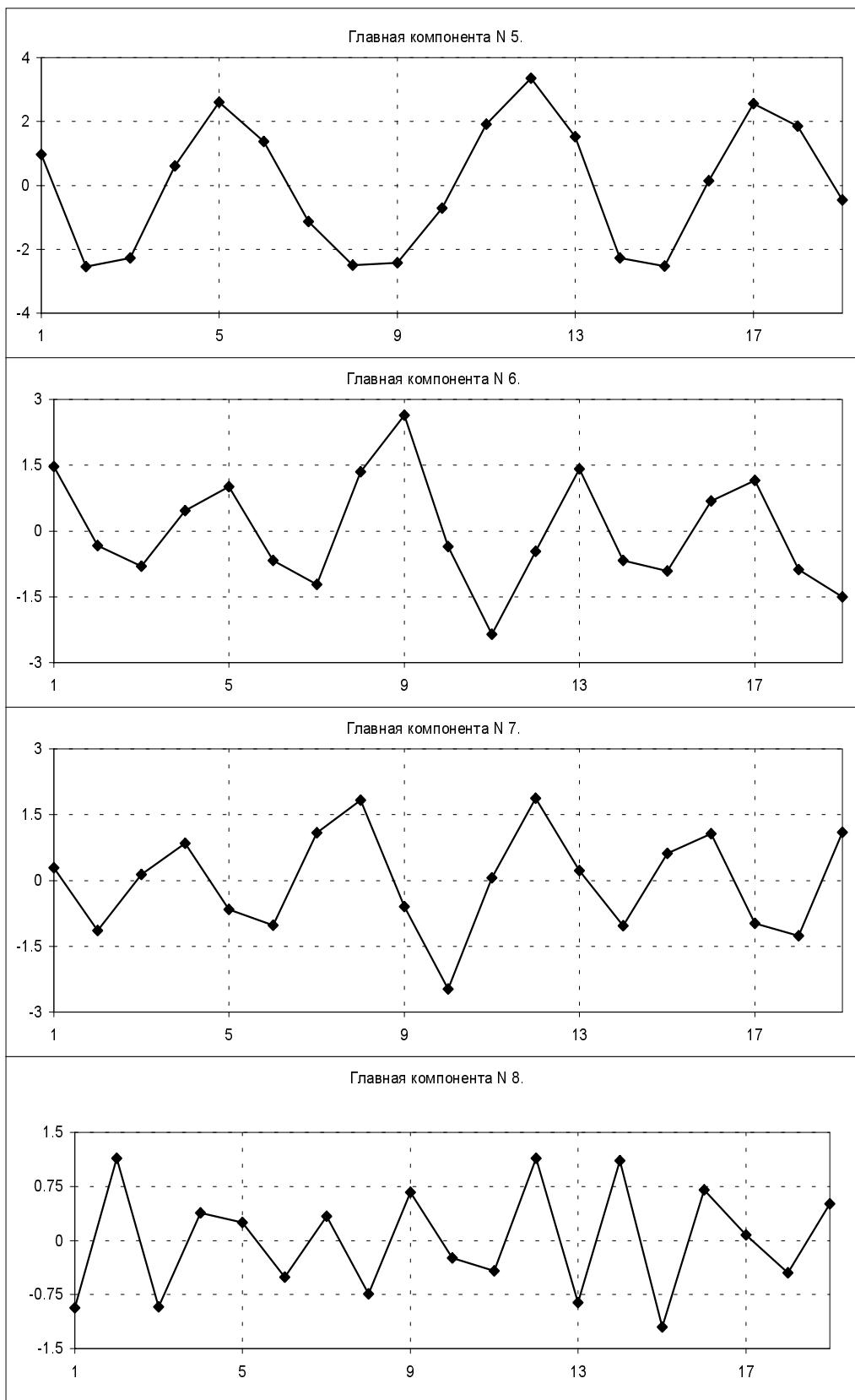


Рис. 3.3. Главные компоненты № 5-8.

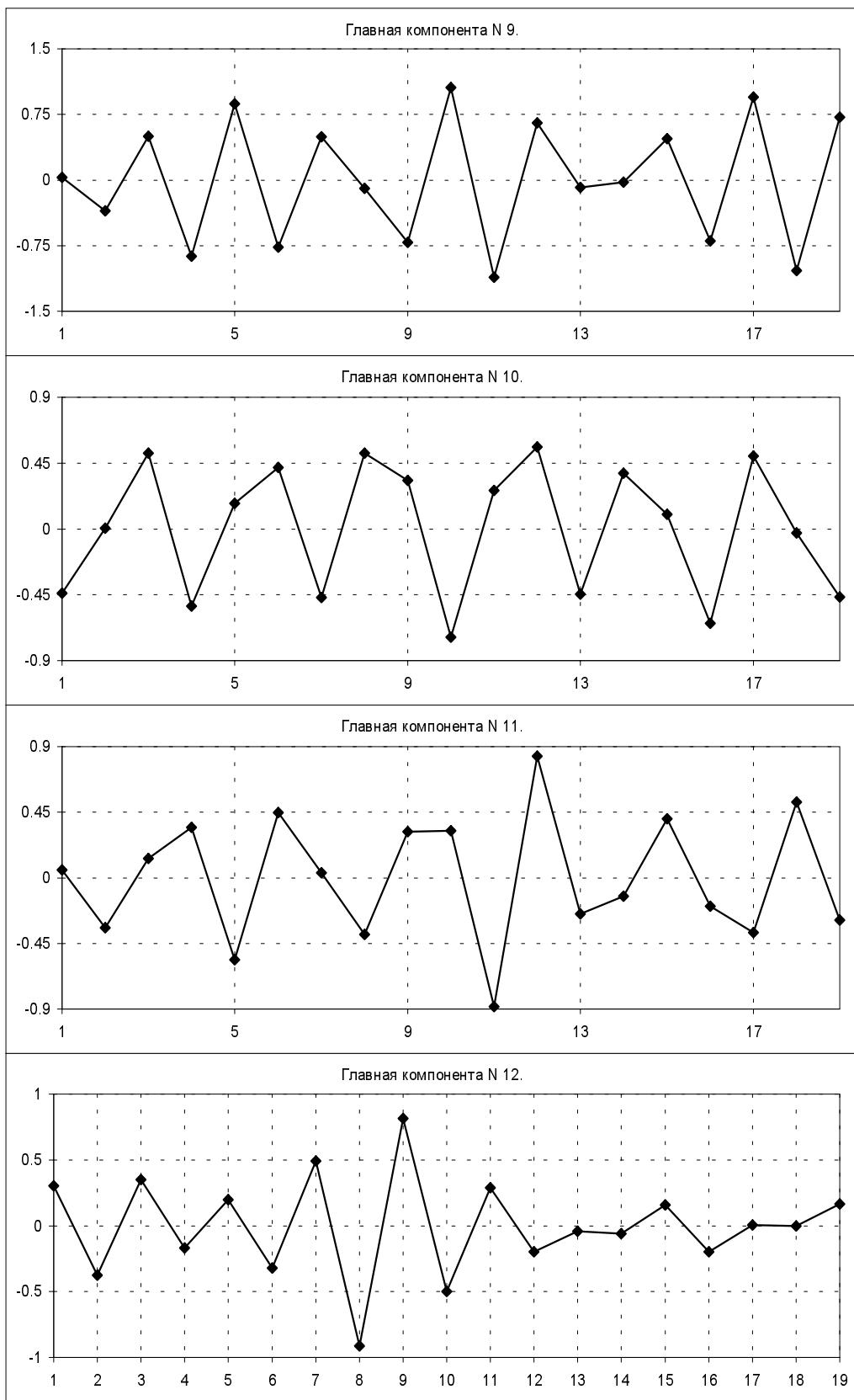


Рис. 4.3. Главные компоненты № 9-12.

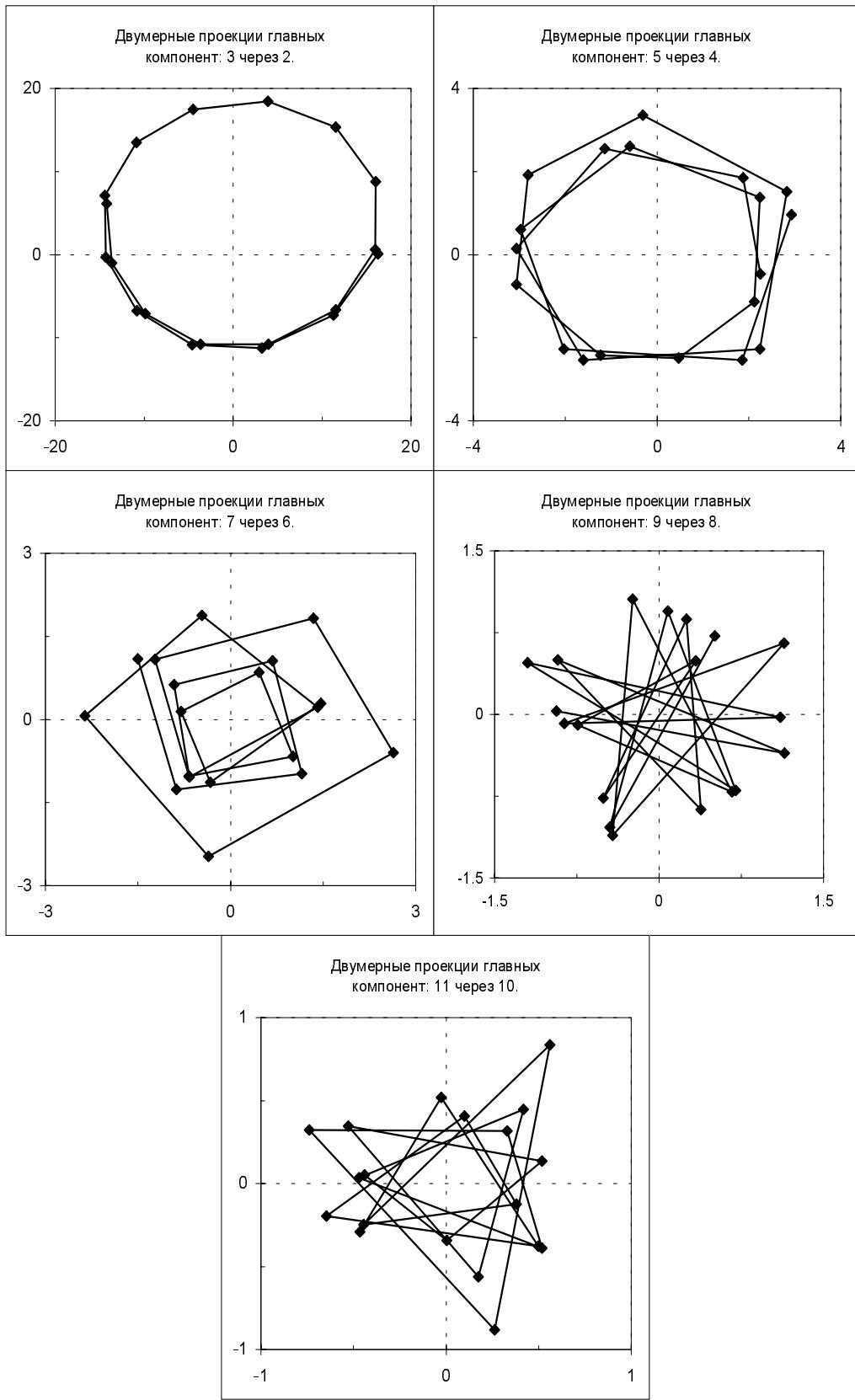


Рис. 5.3. Двумерные проекции ГК 2-3, 4-5, 6-7, 8-9, 10-11.

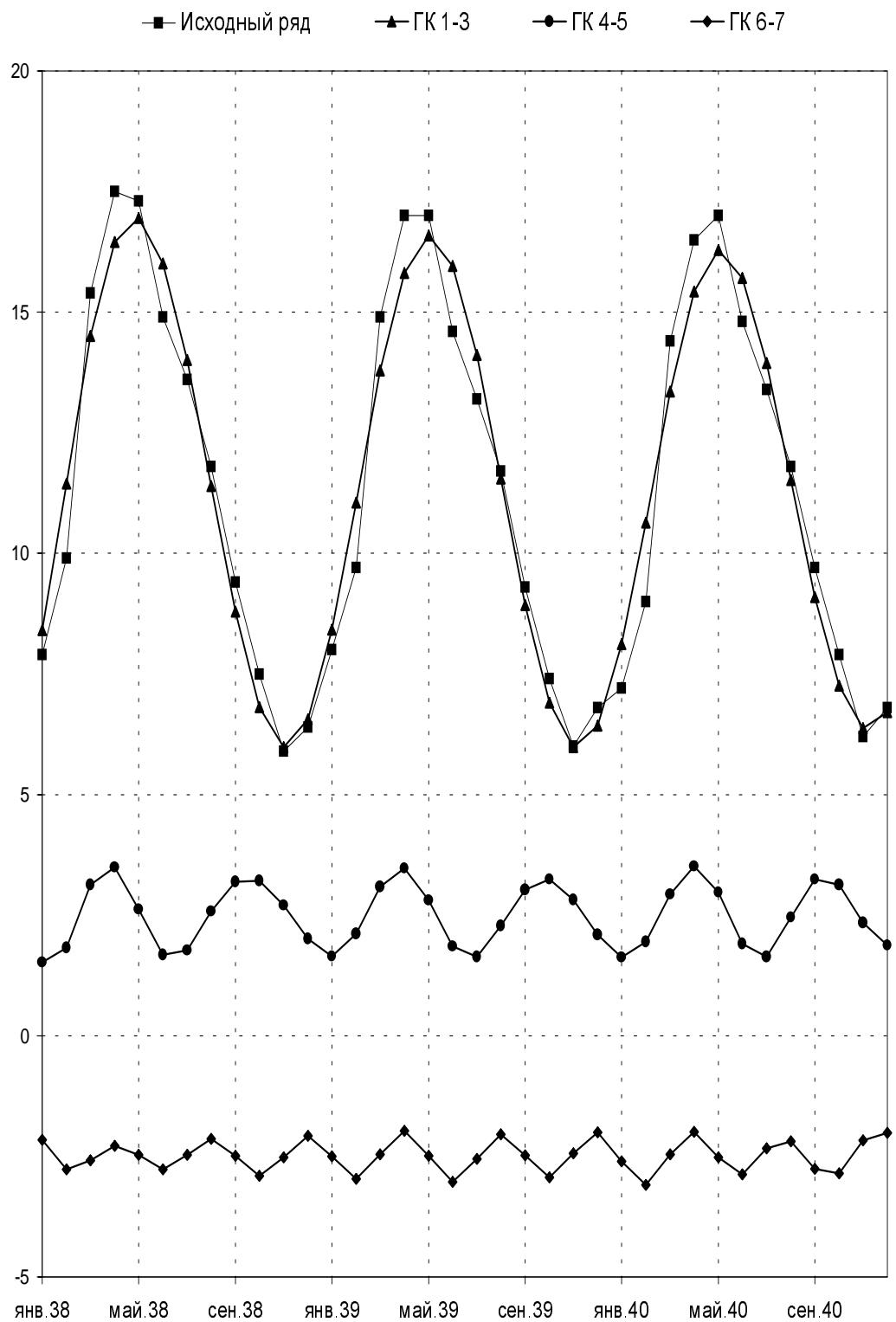


Рис. 6.3. Восстановление основных периодик.

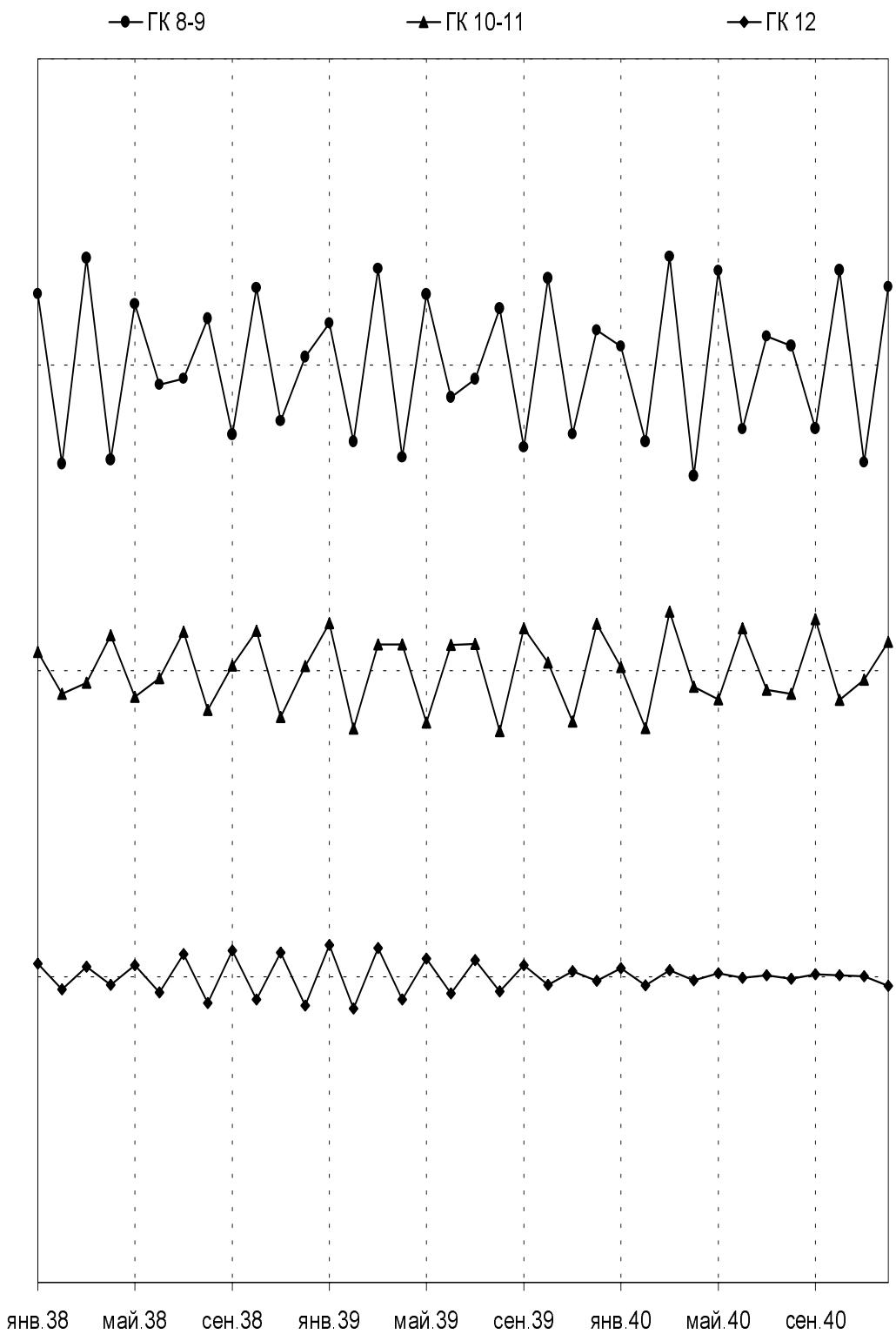


Рис. 7.3. Восстановление основных периодик.

Двумерные проекции главных компонент приведены на Рис. 5.3, а соответствующие восстановления на Рис. 6.3-7.3 (полное разложение исходного ряда разделено на два рисунка, чтобы обеспечить сравнимость амплитуд). На Рис. 8.3 представлен прогноз исходного ряда на основе двухлетнего периода на год вперед и сравнение с исходными данными.

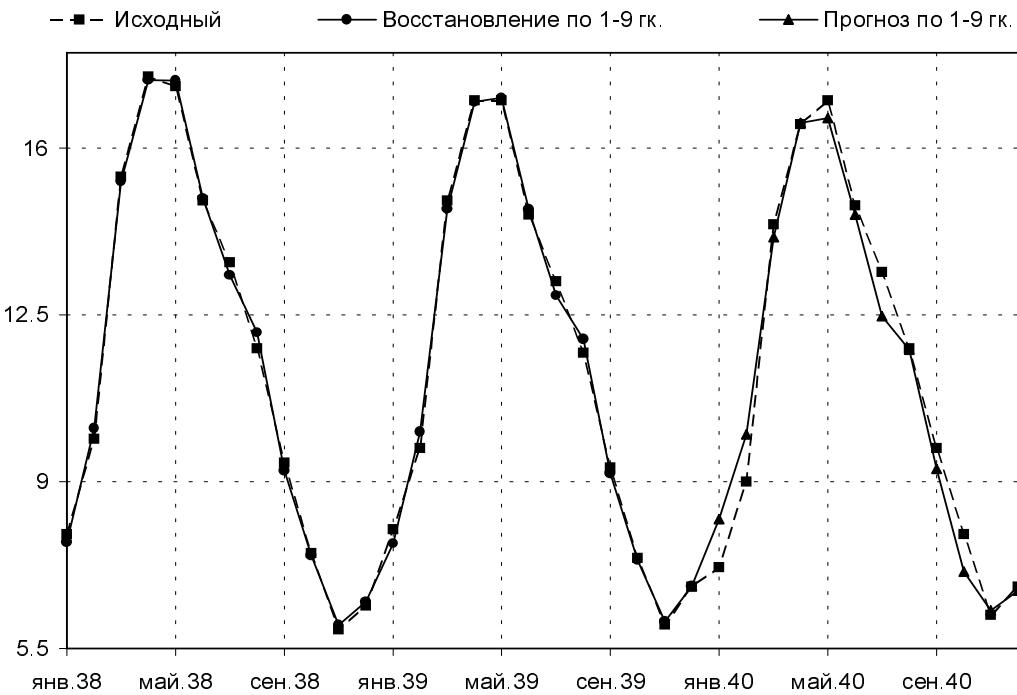


Рис. 8.3. Исходный ряд и прогноз.

Очевидно, в данном случае Гусеница производит разложение исходного ряда по Фурье компонентам. В связи с этим, адекватным классическим методом для сравнения будет обычный Фурье анализ.

3.2 Фурье анализ и сравнение.

Выборочная периодограмма показывает, прежде всего, наличие годовой периодики и еще двух более слабых периодик, см. Рис. 9.3. Увеличив масштаб ряда мы способны определить присутствие еще трех периодических компонент, см. Рис. 10.3. В этом смысле ряд интересен практически полным отсутствием шума: простое увеличение масштаба увеличивает разрешение метода!

На Рис. 11.3 изображено разложение исходного ряда по найденным Фурье компонентам. Легко видеть, что в целом, компоненты, выделенные с помощью Фурье разложения, близки к гусеничным восстановлениям. Для более аккуратного сравнения компоненты, выделенные Фурье анализом и гусеничные компоненты построены в одинаковых координатах, см. Рис. 12.3-13.3. Хорошо видно, что за исключением самой слабой, двухмесячной периодики серьезных расхождений между "Гусеницей" и Фурье анализом не наблюдается (отметим лишь некоторое уменьшение амплитуды годовой и четырехмесячной периодик в начале и

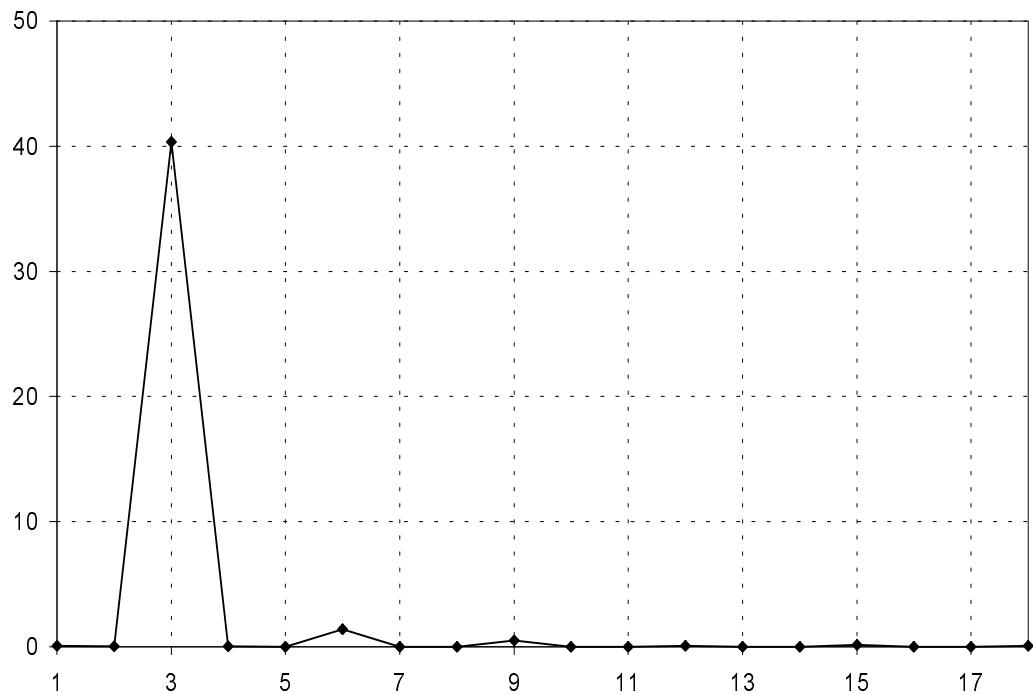


Рис. 9.3. Выборочная периодограмма.

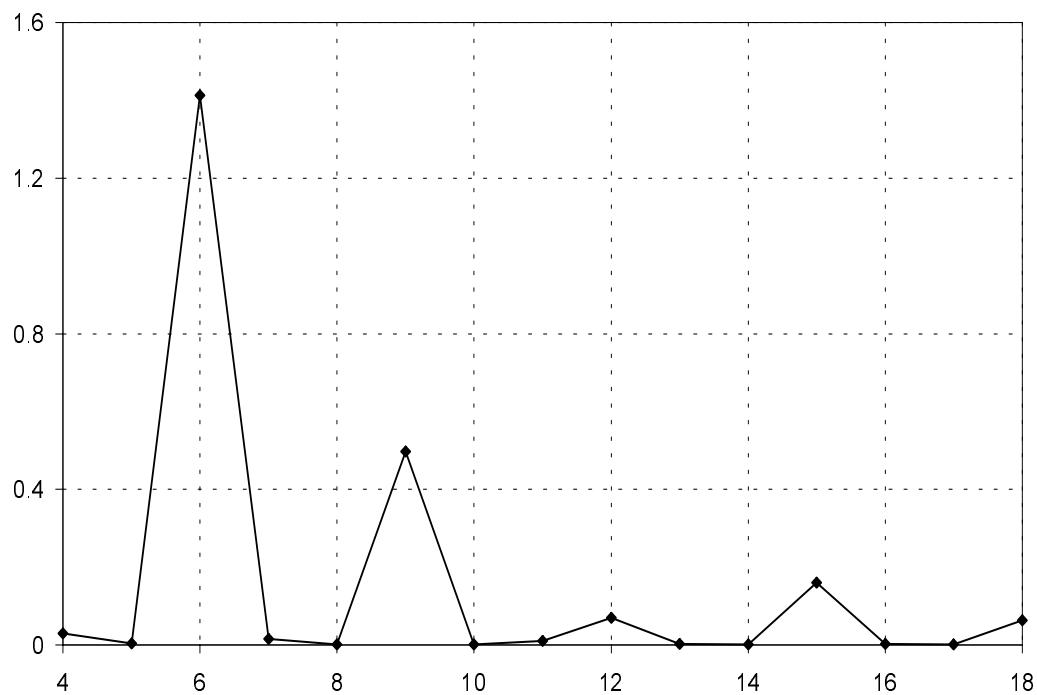


Рис. 10.3. Выборочная периодограмма, без годовой компоненты.

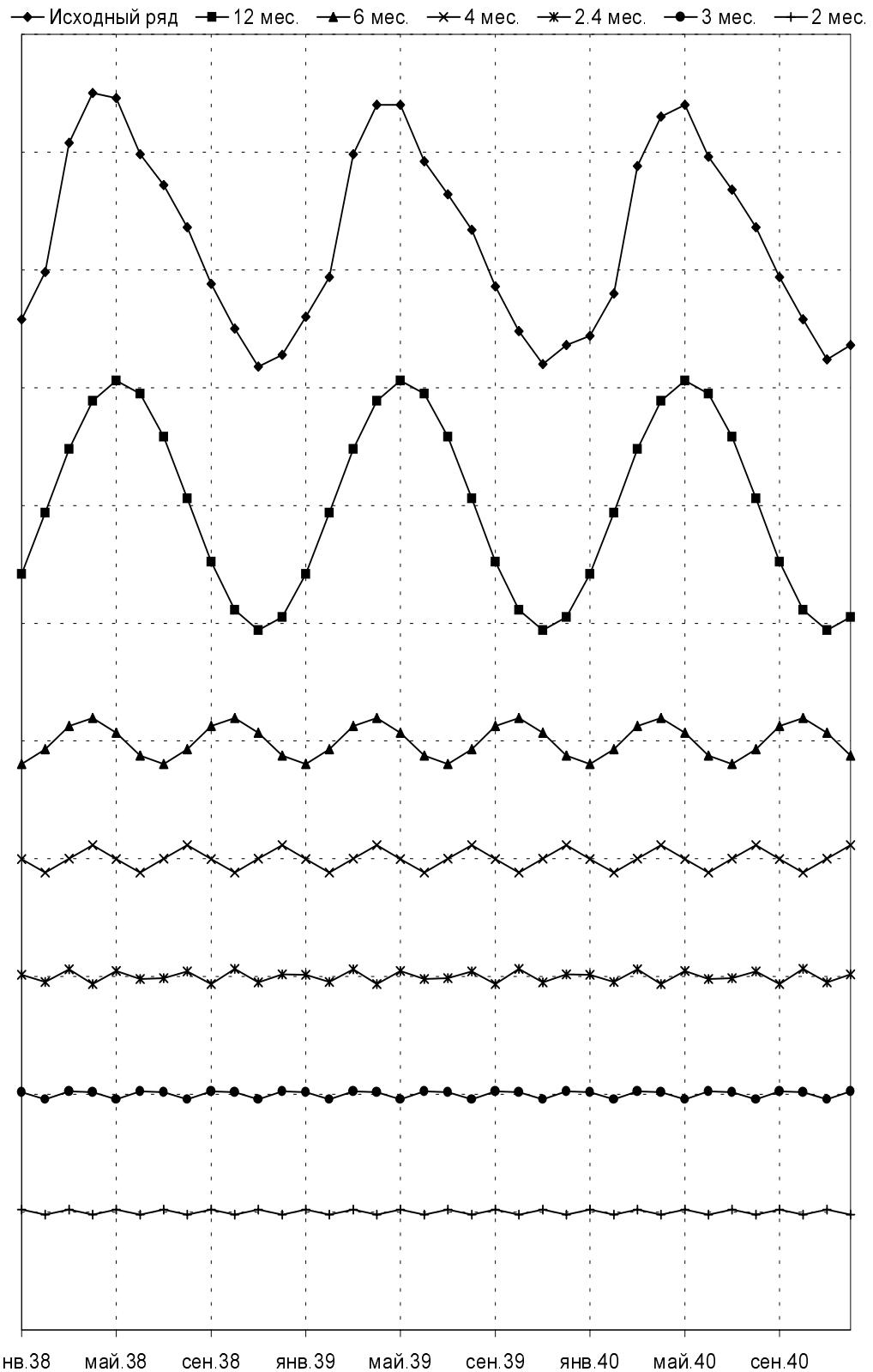


Рис. 11.3. Фурье разложение исходного ряда.

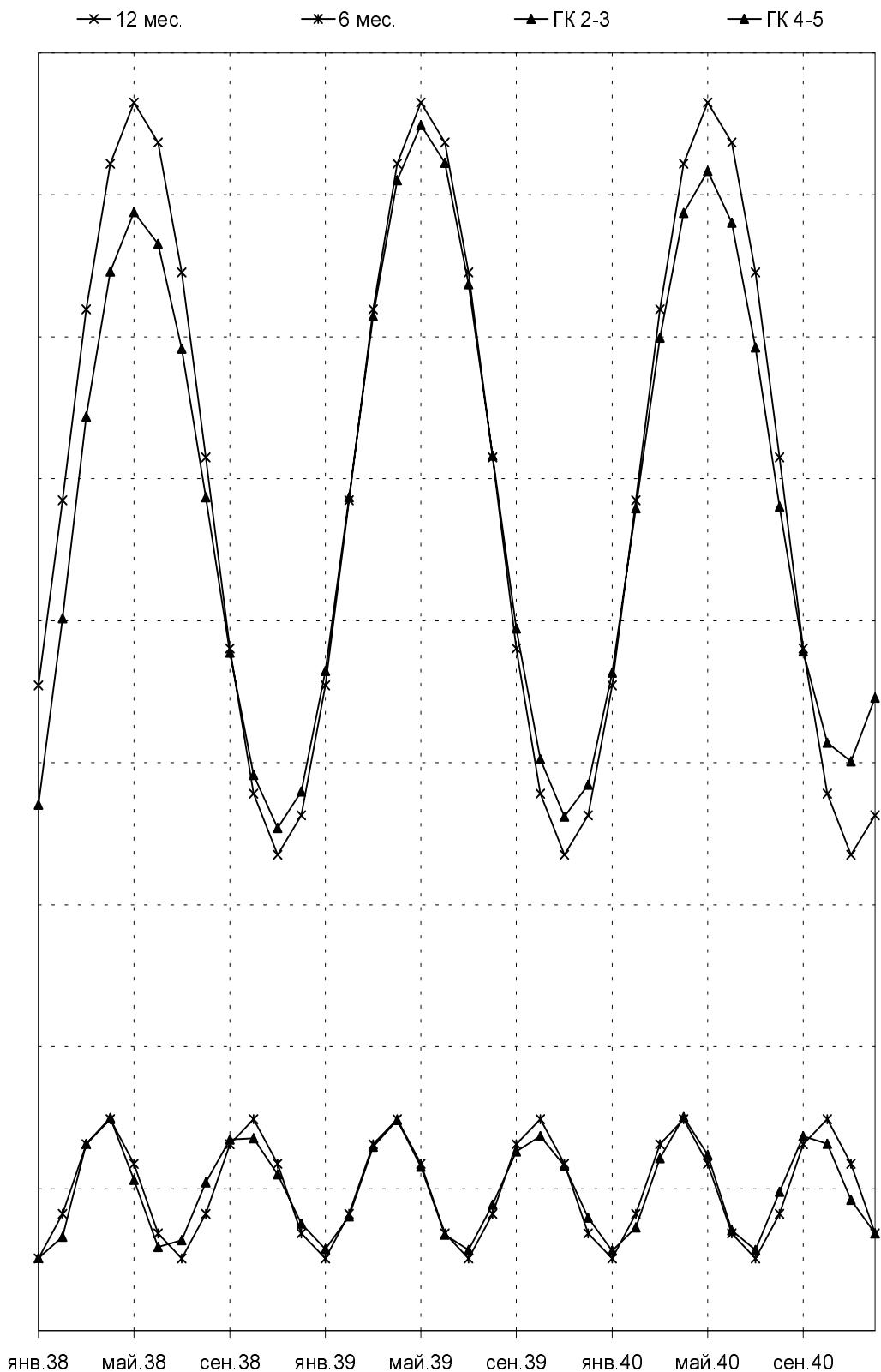


Рис. 12.3. Сравнение Фурье компонент и гусеничных восстановлений.
Среднее, годовая и полугодовая периодики.

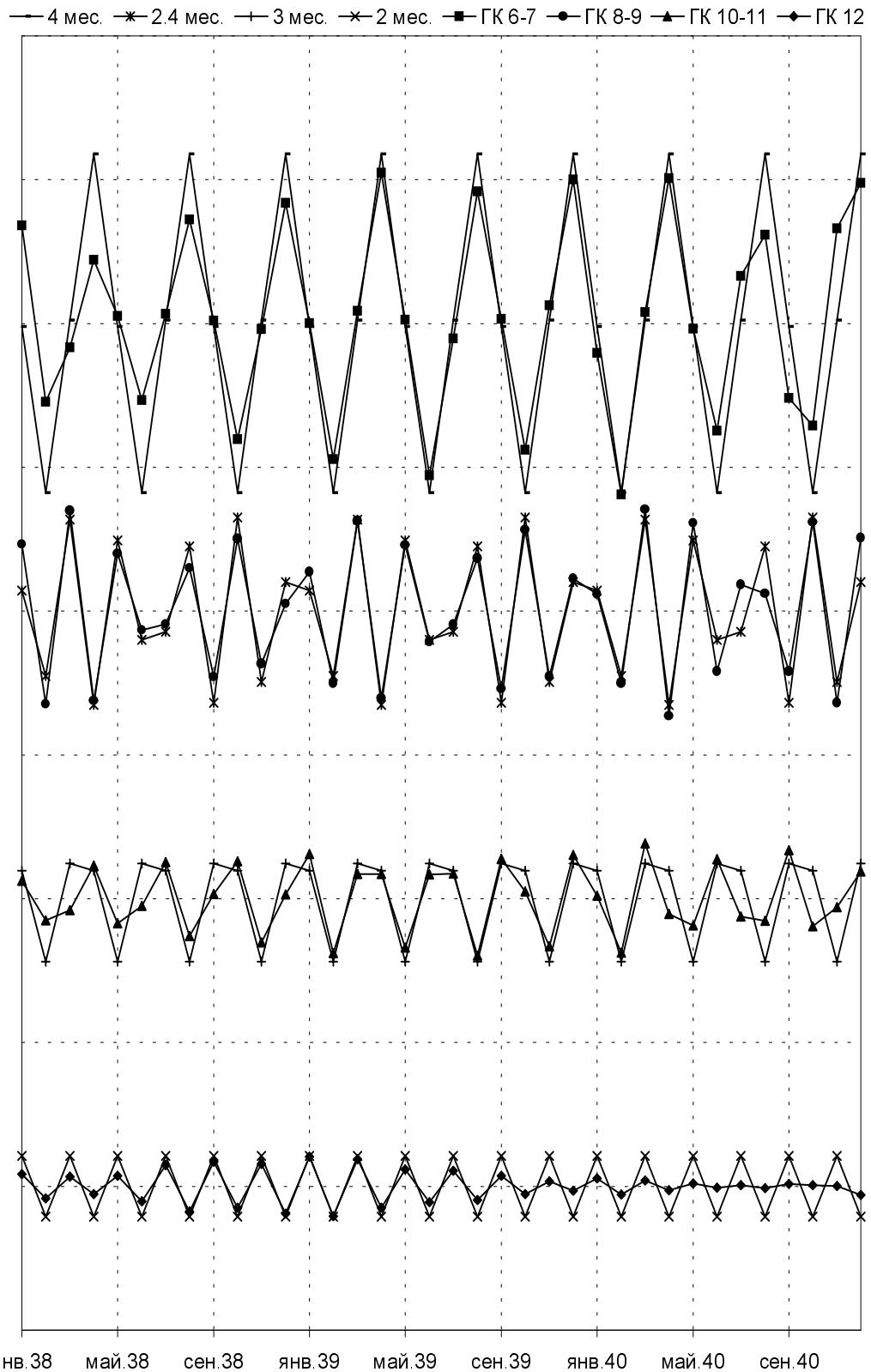


Рис. 13.3. Сравнение Фурье компонент и гусеничных восстановлений.
Высокочастотные периодики.

в конце ряда). Двухмесячная периодика выделенная "Гусеницей", совершенно затухает к концу 1939 г., в то время как Фурье компонента, естественно, так себя не ведет.

В целом можно сказать, что в этом, очень простом случае, Фурье анализ и метод гусеница дают практически полностью эквивалентные описания исходного ряда.

Литература

- [1] Д. Бокс, Д. Дженкинс, (1970), *Анализ временных рядов, прогноз и управление*. М., Наука.
- [2] Кендалл, М.Дж., Стьюарт, А., (1976), *Многомерный статистический анализ и временные ряды*. М., Наука.
- [3] Марпл-мл., С.П., (1990), *Цифровой спектральный анализ и его приложения*. М., Мир.
- [4] Brown, R.G. (1963) *Smoothing, forecasting and prediction of discrete time series*, New Jersey, Prentice-Hall.
- [5] Report of the Bureau of Agricultural Economics, U.S. Dept. of Agriculture, on the Poultry and Egg Situation, March 1941.
- [6] Rosenberg, P.S. and Gail, M.H. (1991) Backcalculation of flexible linear models of the Human Immunodeficiency Virus infection curve. *Applied Statistics*, 40, 269-282.

Глава V Теоретические аспекты

V.1 Разложения временных рядов

В.В. Некруткин

Введение

Целью настоящей статьи является математическое обоснование метода "Гусеница" (в другой терминологии — метод *SSA* или *Singular Spectrum Analisys*), — весьма продуктивного способа анализа временных рядов. Алгоритм метода, его история, особенности, возможности, варианты, способы применения и многочисленные примеры приведены в главе I и поэтому здесь не описываются.

На сегодняшний день имеется достаточно полный список работ, прямо или косвенно связанных с "Гусеницей" (см. библиографию в главе II); в их число входит и теоретическая статья [5], непосредственно относящаяся к математической теории метода.

"Гусеница", однако, допускает содержательное описание и изучение с многих точек зрения и с использованием различной техники. Поэтому не следует удивляться, что настоящая работа отличается от цитированных как математическим аппаратом, так и общей направленностью. Кратко остановимся на упомянутых отличиях.

В [5] рассматриваются, как правило, временные ряды, и основное внимание уделяется размерности их конечномерных разверток. Это оказывается весьма плодотворным для определения "реальной размерности" временного ряда. Математически такой подход приводит к языку геометрии конечномерных пространств. Он оказывается также связан с техникой линейных конечно-разностных уравнений.

В то же время в других работах (см. [6]) метод *SSA* описывается как метод анализа стационарных временных рядов (с этим связаны и особенности соответствующих алгоритмов).

Изложенный ниже материал имеет несколько другую направленность. Он появился как попытка математически формализовать и обосновать реальную практику работы с системой "Гусеница", которая часто дает красивые и интерпретируемые, но "необъяснимые" без адекватной математической теории результаты. В частности, одной из главных проблем метода является выбор длины "Гусеницы" — параметра, который играет совершенно различную роль при анализе коротких и длинных временных рядов. Другая проблема состоит в группировке главных компонент, которая математически описывается как задача "разделения слагаемых в сумме" и соответствует "идеальному" варианту работы метода.

Кроме того, исследуемый метод не является абсолютно жестким и допускает различные варианты и модификации. Поэтому естественно рассмотреть все эти модификации в едином ключе, то есть выработать общий взгляд на "Гусеницу" и подобрать адекватный математический аппарат.

Таким аппаратом оказывается функциональный анализ, точнее — теория операторов Гильберта-Шмидта. Тем самым вместо дискретного времени появляется непрерывное, то есть вместо временных рядов — функции одной переменной. Что касается дискретизации времени, то его можно формально получить, рассматривая считающие меры в соответствующих измеримых пространствах.

Обсудим теперь основные идеи и содержание настоящей работы. Рассмотрим вещественнозначную функцию f , заданную на промежутке $[0, t]$. Будем считать, что эта функция является реализацией некоторого (вообще говоря, случайного) процесса, характеристики которого *a priori* нам не известны. Мы предполагаем, что реализация f может быть разложена в сумму функций f_n , каждая из которых соответствует некоторой существенной стороне исследуемого явления (или является "шумом"). Тем самым наша цель — получить и исследовать разложение

$$f = \sum_n f_n, \quad (1)$$

где слагаемые f_n были бы "интерпретируемые" и "независимы".

Стандартным способом для этого является введение подходящей регрессионной модели, параметры которой оцениваются по реализации f . Существует, однако, большой произвол в выборе таких моделей, что зачастую приводит к чисто механической подгонке "под результат". Такая подгонка не имеет реального смысла и всегда возможна при большом числе параметров. Поэтому будем считать, что у нас параметрическая модель процесса отсутствует.

"Независимость" слагаемых f_n в (1) может быть достигнута разложением f по (ортогональному) базису в некотором гильбертовом пространстве \mathbf{H} (это также является разновидностью параметрической модели), однако встает вопрос о выборе конкретного базиса, от которого будет зависеть "содержательность" и "интерпретируемость" разложения.

Таким образом, мы предполагаем, что у нас нет априорной информации о виде функций f_n , то есть нет модели, "внешней" по отношению к f .

В этой ситуации возникает задача разложения функции f по базису, порожденному самой f . Несмотря на кажущуюся противоречивость формулировки такой задачи, она не так уж абсурдна: в многомерной статистике, теории случайных процессов и линейной алгебре есть целый ряд подобных примеров. Пожалуй, наиболее известным из них является метод главных компонент, где центрированный случайный вектор $\vec{\xi}$ порождает ортонормированный базис (состоящий из собственных векторов ковариационной матрицы $\vec{\xi}$), по которому сам и раскладывается. Несомненная продуктивность метода главных компонент, как нам кажется, основана в первую очередь именно на "внутреннем" происхождении этого базиса. Отметим, что другие методы многомерной статистики (например, факторный анализ или метод канонических переменных) также связаны с разложением случайного вектора по "внутреннему" базису, только этот базис определяется из других соображений.

Аналогом метода главных компонент в теории случайных процессов является разложение Карунена-Лоэва, а в линейной алгебре — сингулярное разложение матриц.

Заметим, однако, что во всех перечисленных случаях мы имели дело с функциями двух переменных (для случайного вектора и случайного процесса роль

второй переменной играет элементарное событие), в то время как у нас есть функция $f(z)$ только одной переменной $z \in [0, t]$. Поэтому нужно "переделать" отрезок $[0, t]$ в прямоугольник π и "перенести" функцию f с $[0, t]$ на π . Формально эта операция задается с помощью отображения $\theta : \pi \rightarrow [0, t]$ и вместо функции f рассматривается функция двух переменных $g = f \circ \theta$. В стандартном алгоритме аддитивной "Гусеницы" $\theta(x, s) = x + s$, $(x, s) \in \pi$ и $g(x, s) = f(x + s)$ (см. главу I).

Функция g при соответствующих условиях может рассматриваться как ядро интегрального оператора, который оказывается оператором Гильберта-Шмидта и, следовательно, обладает рядом привлекательных свойств. В частности, g раскладывается по двум ортогональным системам функций (аналоги собственных векторов и главных компонент в анализе главных компонент). Это разложение называется ниже разложением Шмидта. Выбирая некоторые члены разложения и "возвращаясь" на отрезок $[0, t]$ при помощи операции реконструкции (то есть проектирования, см. [5] в дискретном случае, общая схема рассмотрена в разделе 2), мы получаем один или несколько членов разложения (1), которые затем необходимо интерпретировать и анализировать.

Такова схема применения аппарата функционального анализа для описания "Гусеницы". Кратко остановимся теперь на содержании статьи по разделам.

Раздел 1 посвящен изложению необходимых результатов теории оператора Гильберта-Шмидта. В нем также приведены доказательства менее традиционных, но простых утверждений о разложении Шмидта, необходимых в дальнейшем.

В разделе 2 рассматривается общая схема метода "Гусеница". В нем формально описывается операция реконструкции, рассматриваются различные варианты центрирования (некентральный, с однократным центрированием и бицентрический), а также обсуждается "идеальный" вариант процедуры метода, когда после реконструкции члены ряда (1) оказываются ортогональными в соответствии с L^2 , а "искажения" информации не происходит. Последнее обстоятельство оказывается тесно связанным с задачей "разделения слагаемых в сумме", описанной в разделе 1.

Материал раздела 3 посвящен "аддитивному", наиболее изученному и распространенному на практике варианту "Гусеницы". Здесь конкретизируются и уточняются результаты раздела 2. Особенное внимание уделяется двум вопросам — функциям, для которых разложение Шмидта конечно (эта часть раздела имеет большие пересечения с [5]), и условиям выделения слагаемых в сумме (то есть "идеальному" случаю).

Рассматриваются не только общие условия выделения слагаемых в сумме, но и многочисленные примеры. Эти примеры построены по единому образцу: выбирается некоторая "общепотребительная" функция (скажем, линейная или экспонента) и выясняется, от какого класса функций она отделяется разными модификациями "Гусеницы". При этом оказывается, что эти классы достаточно бедны (и так или иначе связаны с некоторой периодичностью), а условия на длину "Гусеницы" и величину t являются весьма жесткими. Это, конечно, является следствием практически нереального требования "идеальности".

Другая картина возникает при приближенном выделении слагаемых (раздел 4). Оказывается, что процедура метода "Гусеница" (рассматривается для простоты только аддитивный вариант) является устойчивой по отношению к

малым изменениям функции f и параметров.

Более интересные результаты получаются в случае асимптотической разделимости (то есть длинных рядов). Условия разделимости в этом случае существенно облегчаются. В частности, жесткие допредельные условия на длину "Гусеницы" τ (нужно "попасть в период" некоторой функции) превращаются в реально наблюдаемые на практике $\tau \rightarrow \infty$, $t - \tau \rightarrow \infty$.

Еще более выразителен случай "выделения сигнала из шума". Оказывается (точные формулировки см. ниже), что практически любой "сигнал" (детерминированная функция) всегда асимптотически отделяется методом "Гусеница" от случайного "шума", обладающего естественными эргодическими свойствами. Это также наблюдаемый на практике результат.

Особый случай представляет линейный сигнал (тренд). При применении бицентальной "Гусеницы" линейная функция целиком переходит в "среднее", а разложение Шмидта действует только на "шум", фактически фильтруя его. В результате, как показано на примере дискретного времени и белого шума, однократным применением метода "Гусеница" можно уменьшить дисперсию помехи в $O(n)$ раз (n — длина ряда).

Раздел 5 содержит пример неаддитивной "Гусеницы". Этот пример не имеет практического значения, но интересен с точки зрения расширения возможностей метода.

Наконец, сделаем одно замечание. Оно относится к предположению о вещественных значениях функции f . Это предположение не принципиально, и введено по той причине, что большинство практических примеров применения метода "Гусеница" относится к одномерным времененным рядам. Вся изложенная ниже теория в равной степени относится и к комплексному случаю, однако, с точки зрения практики, анализ структуры одномерных и двумерных временных рядов — это существенно различные задачи.

1 Разложение Шмидта и его свойства

Прежде чем перейти к теории метода "Гусеница", кратко остановимся на необходимых сведениях, относящихся к теории операторов Гильберта-Шмидта (см., например, [1], гл. 6). Почти все утверждения этого раздела приводятся без доказательств, так как они либо совпадают с известными фактами функционального анализа, либо легко из них следуют.

Пусть $(D_1, \mathcal{U}_1, \mu_1)$ и $(D_2, \mathcal{U}_2, \mu_2)$ — два пространства с мерой. Обозначим $\mathbf{L}_{1,2}^2$ — гильбертово пространство вещественнозначных функций, суммируемых с квадратом на множестве $D_1 \times D_2$ по мере $\mu_1 \otimes \mu_2$. Аналогичным образом введем сокращенные обозначения \mathbf{L}_1^2 для $L^2(D_1, \mu_1)$ и \mathbf{L}_2^2 для $L^2(D_2, \mu_2)$. Скалярные произведения в этих пространствах обозначим $\langle \cdot, \cdot \rangle_{1,2}$, $\langle \cdot, \cdot \rangle_1$ и $\langle \cdot, \cdot \rangle_2$, а нормы — $\|\cdot\|_{1,2}$, $\|\cdot\|_1$ и $\|\cdot\|_2$. Меры μ_1 и μ_2 будут считаться конечными.

Рассмотрим ненулевую функцию $g \in \mathbf{L}_{1,2}^2$. Эта функция порождает оператор $\mathbf{G} : \mathbf{L}_2^2 \mapsto \mathbf{L}_1^2$, задаваемый равенством

$$\mathbf{G}h = \int_{D_2} g(\cdot, s)h(s)\mu_2(ds), \quad (1)$$

где $h \in \mathbf{L}_2^2$.

Оператор \mathbf{G} является оператором Гильберта-Шмидта, и поэтому его свойства (как и свойства сопряженного к нему оператора \mathbf{G}^*) тесно связаны со свойствами самосопряженных операторов $\mathbf{G}^*\mathbf{G} : \mathbf{L}_2^2 \rightarrow \mathbf{L}_2^2$ и $\mathbf{G}\mathbf{G}^* : \mathbf{L}_1^2 \rightarrow \mathbf{L}_1^2$. Перечислим некоторые из них.

1. Оператор $\mathbf{G}^*\mathbf{G}$ обладает не более чем счетным набором ненулевых (то есть ортогональных ядру $\ker \mathbf{G}^*\mathbf{G}$) ортонормированных в \mathbf{L}_2^2 собственных функций φ_n , причем соответствующие собственные числа λ_n являются положительными.
2. Имеет место равенство

$$\|g\|_{1,2}^2 = \sum_n \lambda_n.$$

В дальнейшем мы будем считать, что числа λ_n упорядочены по убыванию.

3. Функции

$$\psi_n = \mathbf{G}\varphi_n / \sqrt{\lambda_n} \quad (2)$$

являются ортонормированными в \mathbf{L}_1^2 и представляют собой полную систему ненулевых собственных функций оператора $\mathbf{G}\mathbf{G}^*$ с теми же самыми собственными числами λ_n . Аналогичным образом

$$\varphi_n = \mathbf{G}^*\psi_n / \sqrt{\lambda_n}.$$

4. Операторы \mathbf{G} и \mathbf{G}^* допускают представления (разложения Шмидта)

$$\mathbf{G}h_2 = \sum_n \sqrt{\lambda_n} \langle h_2, \varphi_n \rangle_2 \psi_n \quad (3)$$

и

$$\mathbf{G}^*h_1 = \sum_n \sqrt{\lambda_n} \langle h_1, \psi_n \rangle_1 \varphi_n.$$

Таким образом, функции φ_n и ψ_n находятся в естественном отношении двойственности.

Отметим также, что операторы $\mathbf{G}^*\mathbf{G}$ и $\mathbf{G}\mathbf{G}^*$ являются интегральными с ядрами

$$\int_{D_1} g(x, u)g(x, v)\mu_1(dx) \quad \text{и} \quad \int_{D_2} g(x, s)g(y, s)\mu_2(ds)$$

соответственно.

5. Наконец, разложению Шмидта (1.3) оператора \mathbf{G} соответствует разложение в $\mathbf{L}_{1,2}^2$ его ядра,

$$g = \sum_n \sqrt{\lambda_n} \psi_n \otimes \varphi_n. \quad (4)$$

Равенство (1.4), являющееся основой всех дальнейших построений, будем также называть *разложением Шмидта* функции g .

Различные варианты разложения Шмидта хорошо известны и широко применяются. Приведем несколько примеров.

Разложение Карунена-Лоэва ([2], 34.5).

Пусть $\mu_2(D_2) = 1$ (то есть мы можем трактовать $(D_2, \mathcal{U}_2, \mu_2)$ как вероятностное пространство). Пусть, кроме того,

$$\int_{D_2} g(x, s) \mu_2(ds) = 0$$

при любом x , $D_1 = [0, T]$, \mathcal{U}_1 — борелевская σ -алгебра подмножеств $[0, T]$ и μ_1 — мера Лебега на \mathcal{U}_1 . Предположим, наконец, что функция $g(x, \cdot)$ является непрерывной по x в среднеквадратическом относительно меры μ_2 . Тогда переменная x играет роль времени, функция $g(x, s)$ является случайным процессом с нулевым средним и непрерывной ковариационной функцией (которая является ядром интегрального оператора $\mathbf{G}\mathbf{G}^*$), функции φ_n представляют собой "белый шум с дискретным временем", а разложение Шмидта называется (в теории случайных процессов) *разложением Карунена-Лоэва* и является равномерным по x в среднеквадратическом относительно меры μ_2 .

Анализ главных компонент ([3], гл. 13).

Этот случай отличается от предыдущего только тем, что $D_1 = \{1, \dots, n\}$, а мера μ_1 приписывает каждой точке $i \in D_1$ нагрузку, равную $1/n$. Теперь g является (центрированным) случайным вектором, ψ_n — это (ортонормированная) система собственных векторов его ковариационной матрицы, а φ_n — соответствующие (нормированные) главные компоненты случного вектора g . Таким образом, здесь разложение Шмидта — это "разложение случного вектора по (ковариационным) главным компонентам".

Сингулярное разложение матриц ([4], гл. 9).

Пусть $D_1 = \{1, \dots, m\}$, $D_2 = \{1, \dots, k\}$, а меры μ_1 и μ_2 снова являются считающими. Тем самым функция g превращается в прямоугольную матрицу G с элементами $g(i, j)$, $1 \leq i \leq m$, $1 \leq j \leq k$, а функции φ_n и ψ_n — в (ортонормированные) ненулевые собственные вектора матриц $G^T G$ и $G G^T$. Их количество, очевидно, равно рангу матрицы G , $r = \text{rank } G$. Обозначим через F матрицу размером $n \times r$ со столбцами ψ_n , H — матрицу размером $k \times r$ со столбцами φ_n и $\Lambda^{1/2}$ — диагональную матрицу размером $r \times r$ с диагональными элементами, равными $\sqrt{\lambda_n}$. Тогда разложение Шмидта приобретет вид

$$G = F^T \Lambda^{1/2} H,$$

то есть превратится в сингулярное разложение матрицы G .

Хотя следующие утверждения вполне элементарны с точки зрения функционального анализа, мы приведем их доказательства из-за важности этих фактов для дальнейшего и ввиду их очевидной параллельности основным свойствам главных компонент.

Прежде всего, отметим две характеристики разложения Шмидта. Первая из них относится к единственности представления (1.4).

Лемма 1.1 Пусть $\{\varphi_n\}$ и $\{\psi_n\}$ — две системы функций, ортонормированных соответственно в L_2^2 и L_1^2 , c_n — некоторые ненулевые константы и сумма

$\sum_n c_n^2$ конечна. Пусть $g \in \mathbf{L}_{1,2}^2$ и

$$g = \sum_n c_n \psi_n \otimes \varphi_n, \quad (5)$$

причем последний ряд сходится в $\mathbf{L}_{1,2}^2$. Пусть, наконец, функция g определяет оператор \mathbf{G} согласно (1.1). Тогда (1.5) является разложением Шмидта функции g .

Доказательство. Как нетрудно видеть, ядро интегрального оператора $\mathbf{G}^* \mathbf{G}$ имеет вид $\sum_k c_k^2 \varphi_k \otimes \varphi_k$. Отсюда немедленно следует, что φ_k является собственной функцией оператора $\mathbf{G}^* \mathbf{G}$ и соответствует собственному числу c_k^2 . Кроме того, любая функция, ортогональная φ_k при всех k , принадлежит ядру оператора $\mathbf{G}^* \mathbf{G}$. \square

Другая характеристика разложения Шмидта связана с его экстремальными свойствами. Обозначим H_k подмножество $\mathbf{L}_{1,2}^2$, состоящее из функций вида $\sum_{i=1}^k \eta_i \otimes \phi_i$, где $\eta_i \in \mathbf{L}_1^2$, а $\phi_i \in \mathbf{L}_2^2$.

Лемма 1.2 Имеют место следующие утверждения.

1. При $h \in H_k$

$$\|g - h\|_{1,2}^2 \geq \sum_{i>k} \lambda_i, \quad (6)$$

причем равенство в (1.6) достигается при

$$h = \sum_{i=1}^k \sqrt{\lambda_i} \psi_i \otimes \varphi_i.$$

2. Пусть $I = \{i_1, \dots, i_k\}$ — некоторое подмножество множества $N = \{1, 2, \dots\}$, $\max i_j = m$, $J = \{1, \dots, m\} \setminus I$. Пусть $g \in \mathbf{L}_{1,2}^2$ и $H_k^I \subset H_k$ состоит из таких элементов $h = \sum_{i=1}^k \eta_i \otimes \phi_i$, что η_i ортогональна каждой функции ψ_l , $l \in J$, из разложения (1.4). Тогда при $h \in H_k^I$

$$\|g - h\|_{1,2}^2 \geq \sum_{i \in N \setminus I} \lambda_i, \quad (7)$$

причем равенство в (1.7) достигается при

$$h = \sum_{i \in I} \sqrt{\lambda_i} \psi_i \otimes \varphi_i.$$

Доказательство. Будем доказывать оба пункта леммы одновременно. Пусть η_1, \dots, η_k — фиксированное конечное семейство ортонормированных функций в \mathbf{L}_1^2 , ϕ_i , $1 \leq i \leq k$ — некоторые элементы \mathbf{L}_2^2 и $h = \sum_{i=1}^k \eta_i \otimes \phi_i$. Так как

$$\langle g - \sum_{i=1}^k \eta_i \otimes \mathbf{G}^* \eta_i, h \rangle_{1,2} = 0,$$

при любых ϕ_i , то

$$\|g - h\|_{1,2}^2 \geq \|g - \sum_{i=1}^k \eta_i \otimes \mathbf{G}^* \eta_i\|_{1,2}^2 = \|g\|_{1,2}^2 - \sum_{i=1}^k \langle \eta_i, \mathbf{G} \mathbf{G}^* \eta_i \rangle_1. \quad (8)$$

Тем самым утверждение леммы сводится к минимизации правой части (1.8) без всяких условий (первый пункт леммы) или при условиях $\langle \eta_i, \psi_l \rangle_1 = 0$, $1 \leq i \leq k$, $l \in J$ (второй пункт).

Результат безусловной минимизации хорошо известен — минимум достигается при $\eta_i = \psi_i$ и равен правой части (1.6). Совершенно аналогично доказывается и второе утверждение. \square

Следующий вопрос, который нам нужно решить — это вопрос о выделении слагаемых в сумме с помощью разложений (1.4).

Пусть $g_1, g_2 \in \mathbf{L}_{1,2}^2$ и $g = g_1 + g_2$. Как уже отмечалось, каждая из этих трех функций задает по интегральному оператору $\mathbf{L}_2^2 \rightarrow \mathbf{L}_1^2$ (обозначим их соответственно \mathbf{G}_1 , \mathbf{G}_2 и \mathbf{G}). Тем самым каждая из функций g_1, g_2 и g порождает по две системы отонормированных функций; мы обозначим их $\{\varphi_{1n}, \psi_{1n}\}$, $\{\varphi_{2n}, \psi_{2n}\}$ и $\{\varphi_n, \psi_n\}$. Обозначим $\mathcal{L}_{1,\varphi}$, $\mathcal{L}_{2,\varphi}$ замкнутые линейные подпространства \mathbf{L}_2^2 с базисами $\{\varphi_{1n}\}$ и $\{\varphi_{2n}\}$. Аналогичные обозначения $\mathcal{L}_{1,\psi}$ и $\mathcal{L}_{2,\psi}$ будут применяться для соответствующих подпространств \mathbf{L}_1^2 .

Если $\mathcal{L}_{1,\varphi} \perp \mathcal{L}_{2,\varphi}$ и $\mathcal{L}_{1,\psi} \perp \mathcal{L}_{2,\psi}$ то будем говорить, что *разложение Шмидта* (1.4) *разделяет функции* g_1 и g_2 в сумме $g = g_1 + g_2$.

Поясним сначала это определение. Пусть функции g_1 и g_2 разделяются в сумме $g = g_1 + g_2$. Тогда

$$g = \sum_k \sqrt{\lambda_{1k}} \psi_{1k} \otimes \varphi_{1k} + \sum_l \sqrt{\lambda_{2l}} \psi_{2l} \otimes \varphi_{2l}, \quad (9)$$

причем $\langle \psi_{1k}, \psi_{2l} \rangle_1 = \langle \varphi_{1k}, \varphi_{2l} \rangle_2 = 0$ при любых k и l . Согласно лемме 1.1 это означает, что (1.9) является разложением Шмидта функции g , то есть ψ_{1k} и ψ_{2l} являются собственными функциями оператора $\mathbf{G}_1 \mathbf{G}_2^*$, а φ_{1k} и φ_{2l} — $\mathbf{G}_1^* \mathbf{G}_2$, причем других ненулевых собственных функций у этих операторов нет. Тем самым члены разложения Шмидта функции g можно сгруппировать в две суммы, одна из которых равна g_1 , а вторая — g_2 . Поэтому представление $g = g_1 + g_2$ является естественным с точки зрения разложения Шмидта.

Лемма 1.3 *Следующие условия эквивалентны:*

1. *Разложение Шмидта разделяет* g_1 и g_2 в сумме $g = g_1 + g_2$.
2. *Операторы* $\mathbf{G}_1^* \mathbf{G}_2$ и $\mathbf{G}_1 \mathbf{G}_2^*$ *являются нулевыми.*

Доказательство. Пусть g_1 и g_2 разделяются в сумме. Тогда

$$\mathbf{G}_1^* \mathbf{G}_2 h = \sum_k \sqrt{\lambda_{1k}} \langle \mathbf{G}_2 h, \psi_{1k} \rangle_1 \varphi_{1k}.$$

В свою очередь

$$\langle \mathbf{G}_2 h, \psi_{1k} \rangle_1 = \sum_l \sqrt{\lambda_{2l}} \langle h, \varphi_{1l} \rangle_2 \langle \psi_{2l}, \psi_{1k} \rangle_1 = 0,$$

так что $\mathbf{G}_1^* \mathbf{G}_2 h = 0$ при любом $h \in \mathbf{L}_2^2$. Точно такое же доказательство проходит и для оператора $\mathbf{G}_1 \mathbf{G}_2^*$.

Докажем теперь обратное утверждение. Поскольку $\mathbf{G}_2 \varphi_{2j} = \sqrt{\lambda_{2j}} \psi_{2j}$, то

$$\mathbf{G}_1^* \mathbf{G}_2 \varphi_{2j} = \sqrt{\lambda_{2j}} \sum_k \sqrt{\lambda_{1k}} \langle \psi_{2j}, \psi_{1k} \rangle_1 \varphi_{1k} = 0.$$

Отсюда немедленно следует, что ψ_{2j} и ψ_{1k} ортогональны. Аналогично доказывается ортогональность φ_{2j} и φ_{1k} . \square

Поскольку операторы \mathbf{G}_1 и \mathbf{G}_2 являются интегральными, то условие разделимости функций g_1 и g_2 сводится к выполнению двух условий:

a) при $\mu_1^{\otimes 2}$ -почти всех $(x, y) \in D_1^2$

$$\int_{D_2} g_1(x, s) g_2(y, s) \mu_2(ds) = 0; \quad (10)$$

b) при $\mu_2^{\otimes 2}$ -почти всех $(u, v) \in D_2^2$

$$\int_{D_1} g_1(x, u) g_2(x, v) \mu_1(dx) = 0. \quad (11)$$

Замечание 1.1 По аналогии с методом главных компонент многомерного статистического анализа, мы будем называть оператор $\mathbf{G}\mathbf{G}^*$ (*нецентральным*) *корреляционным оператором*, функции φ_n — *факторными функциями*, а функции ψ_n — *собственными функциями*.

Здесь уместно остановиться на этой аналогии более подробно. Функция $g(x, s)$ играет роль матрицы наблюдений, причем переменная x соответствует набору признаков, а переменная s — индивидам. Таким образом, D_1 является пространством признаков, а D_2 — пространством индивидов. Сразу же виден некий произвол в "назначении" индивидов и признаков. Однако этот произвол естественен: если в стандартном методе главных компонент признаки предполагаются "связанными" (коррелированными), а индивиды — "независимыми" (повторная выборка), то при применении этого метода, скажем, для анализа нескольких коррелированных временных рядов разница между "индивидуами" и "признаками" не столь очевидна: мы можем в качестве "индивидуов" рассматривать последовательные моменты времени, тогда "признаками" будут сами временные ряды и мы будем изучать связь между ними на всем промежутке наблюдения сразу. Можно действовать и противоположным образом — считать "индивидуами" временные ряды, а "признаками" — моменты времени. Тогда, применив метод главных компонент, мы сможем исследовать различные "компоненты" динамики всех рядов одновременно.

Если подходить к методу главных компонент как методу анализа структуры данных, основная черта которого — разложение двумерного массива данных по ортонормированным системам, порожденным самими данными, то оба эти приема одинаково приемлемы.

В этом контексте понятна двойственность ортогональных системам $\{\varphi_n\}$ и $\{\psi_n\}$ — она соответствует двойственности индивидов и признаков.

2 Общая схема метода "Гусеница"

Начнем с описания общей схемы метода "Гусеница" с точки зрения материала, изложенного в предыдущем разделе.

Пусть $t > 0$ и f — измеримая функция $[0, t] \mapsto \mathbb{R}$. Зафиксировав числа $a, b > 0$, рассмотрим прямоугольник $\pi_{a,b} = [0, a] \times [0, b]$ и измеримую сюръекцию $\theta : \pi_{a,b} \mapsto [0, t]$, которая стандартным образом порождает в $\pi_{a,b}$ σ -алгебру \mathcal{B}_θ . Кроме того, рассмотрим конечные меры μ_1 и μ_2 на борелевских подмножествах $[0, a]$ и $[0, b]$ соответственно. Положим $D_1 = [0, a]$ и $D_2 = [0, b]$.

Тогда мы можем определить функцию $g : \pi_{a,b} \mapsto \mathbb{R}$ равенством

$$g(x, s) = f(\theta(x, s)). \quad (1)$$

Кроме того, отображение θ порождает на отрезке $[0, t]$ с борелевской σ -алгеброй подмножества меру $\nu = (\mu_1 \otimes \mu_2) \theta^{-1}$, так что

$$\nu(A) = \mu_1 \otimes \mu_2(\theta^{-1}A).$$

Таким образом, мы получаем еще одно гильбертово пространство $L^2([0, t], d\nu)$, или, сокращенно, $L^2(d\nu)$. Предположим, что $g \in L^2_{1,2}$ (то есть $f \in L^2(d\nu)$), и поэтому мы можем рассматривать разложение (1.4), которое в данном случае приобретает вид

$$g(x, s) = f(\theta(x, s)) = \sum_n \sqrt{\lambda_n} \psi_n(x) \varphi_n(s). \quad (2)$$

Переход от разложения (2.2) к разложению исходной функции f осуществляется с помощью операции реконструкции (см. [5], где фактически рассматривается случай считающих мер).

А именно, обозначим L^2_θ замкнутое линейное подпространство пространства $L^2_{1,2}$, состоящее из \mathcal{B}_θ -измеримых функций, то есть из функций, зависящих от переменных x и s через отображение θ . Тогда проекция n -го члена ряда (2.2) на подпространство L^2_θ имеет вид $f_n(\theta(x, s))$ и

$$f(\theta(x, s)) = \sum_n f_n(\theta(x, s))$$

$\mu_1 \otimes \mu_2$ -почти всюду. Функцию $f_n : [0, t] \mapsto \mathbb{R}$ (определенную ν -почти всюду) естественно называть *результатом реконструкции функции f по n -й главной компоненте φ_n* . Аналогичным образом определяется реконструкция f по некоторому подмножеству слагаемых в правой части (2.2). Очевидно,

$$f = \sum_n f_n, \quad (3)$$

причем ряд в правой части (2.3) сходится в $L^2(d\nu)$. Отметим, что норма функции f_n в $L_2(d\nu)$ не превосходит $\sqrt{\lambda_n}$.

Часто операцию проектирования $L^2_{1,2} \hookrightarrow L^2_\theta$ можно провести явно. Действительно, не умаляя общности, можно считать, что меры μ_1 и μ_2 являются нормированными, и, следовательно, мера $\mathbf{P}_{1,2} = \mu_1 \otimes \mu_2$ тоже является вероятностной. Тогда оператор проектирования — это оператор условного математического ожидания. Поэтому нам достаточно уметь вычислять условное распределение

$$P_{1,2}(\cdot; u) = \mathbf{P}_{1,2}\left((x, s) \in \cdot \mid \theta(x, s) = u\right).$$

Тем самым член f_n ряда (2.3) приобретает вид

$$f_n = \sqrt{\lambda_n} \iint \psi_n(x) \varphi_n(s) P_{1,2}(dx ds; \cdot).$$

Вычисление интеграла

$$Pr_\theta h(u) = \iint h(x, s) P_{1,2}(dx ds; u) \quad (4)$$

(здесь Pr_θ — оператор проектирования на \mathbf{L}_θ^2) не представляет сложности, если меры μ_1 и μ_2 являются дискретными ([5], см. также главу I). Приведем пример абсолютно непрерывной ситуации, когда можно явно выписать вид интеграла (2.4).

Рассмотрим отображение $\Theta : (x, s) \mapsto (\theta(x, s), s)$, обозначим $B_\theta = \Theta(\pi_{a,b})$ и предположим, что отображение $\Theta : \pi_{a,b} \mapsto B_\theta$ является биекцией, так что $\Theta^{-1} : (z, s) \mapsto (\eta(z, s), s)$. Будем считать функцию η достаточно гладкой. Тогда, как нетрудно видеть, (2.4) имеет вид

$$Pr_\theta h(u) = \int_0^b h(\eta(u, s), s) A(u, s) \mu_2(ds) / \int_0^b A(u, s) \mu_2(ds), \quad (5)$$

где

$$A(u, s) = p_1(\eta(u, s)) |\partial_u \eta(u, s)| \mathbf{1}_{B_\theta}(u, s),$$

p_1 — плотность распределения μ_1 и ∂_u — знак частной производной.

Смысл операции восстановления достаточно ясен. Функция $g = f \circ \theta$ раскладывается в (2.2) по двум ортогональным системам, порожденным самой этой функцией. Это (как и в методе главных компонент) позволяет надеяться, что функции φ_n и ψ_n отражают существенные черты поведения функции g и поэтому, выделяя и суммируя некоторое конечное подмножество слагаемых в правой части (2.2) (то есть "восстанавливая" функцию g по нескольким главным компонентам), можно получить достаточно точные и интерпретируемые результаты.

Нас, однако, интересует не функция $f \circ \theta$, а функция f . Ввиду того, что каждое слагаемое (и любая конечная сумма слагаемых) в (2.2) не является, вообще говоря, функцией от $\theta(x, s)$, необходим специальный переход от восстановления функции g к реконструкции функции f . Этот переход осуществляется операцией проектирования.

Таким образом, реконструкция функции f по нескольким главным компонентам сводится к суммированию функций f_n по заранее выбранному набору индексов n . Это соответствует решению двух оптимизационных задач: сначала (лемма 1.2) отысканию в H_k (или в соответствующем H_k^I) функции, ближайшей к $f \circ \theta$, а потом проектировании этой функции на \mathbf{L}_θ^2 .

Следует отметить, что функции f_n не являются, вообще говоря, ортогональными в $\mathbf{L}^2(d\nu)$. С другой стороны, полное восстановление функции f по всем главным компонентам является, конечно же, тождественной операцией.

Отвлекаясь от вычислительных проблем, практический успех описанной конструкции (при фиксированной функции f) зависит от задания отображения θ и выбора мер μ_1 и μ_2 . Алгоритм аддитивной "Гусеницы" ([5] и глава I настоящего сборника, в данный момент мы не обращаем внимания на меры μ_1 и μ_2) соответствует (при некотором фиксированном числе $\tau \in (0, t)$, называемом *длиной "Гусеницы"*) выбору $a = \tau$, $b = t - \tau$ и $\theta(x, s) = x + s$. Этот случай является наиболее исследованным как с теоретической, так и с практической точки зрения. Возможны, конечно, и другие варианты отображения θ , например, мультиплексивная "Гусеница" ($\theta(x, s) = xs$, $x \in [0, \tau]$, $s \in [0, t/\tau]$).

Меры μ_1 и μ_2 , как уже говорилось, участвуют в определении операторов \mathbf{G} и \mathbf{G}^* , и, следовательно, существенным образом влияют на разложение (2.2). Кроме того, эти меры вместе с функцией θ определяют меру ν , которая не является заранее заданной. Например, при $\theta(x, s) = x + s$ и мерах μ_1 и μ_2 ,

являющихся нормированными мерами Лебега, мера ν уже не будет мерой Лебега (она будет равна свертке двух равномерных распределений), и, следовательно, в разложении (2.3) "маленькие" и "большие" значения аргумента будут играть меньшую роль, чем "средние".

Эту ситуацию можно пытаться исправить, вводя в качестве μ_1 и μ_2 "неравномерные" распределения. При этом, однако, основное разложение (2.2) несколько изменится.

В принципе велик произвол и в выборе функции g (то есть при выбранном отображении θ функцию g можно задавать и другими, отличными от (2.1), способами). Укажем несколько возможностей, имеющих прямые аналогии в методах многомерной статистики. Во-первых, это переход от "некентрального" случая к "центральному". Этот переход осуществляется преобразованием

$$g(x, s) = f(\theta(x, s)) - \mathcal{E}_2 f(x), \quad (6)$$

где

$$\mathcal{E}_2 f(x) = \int_{D_2} f(\theta(x, s)) \mu_2(ds) / \mu_2(D_2). \quad (7)$$

Оператор $\mathbf{G}\mathbf{G}^*$ естественно называть в этом случае *центрационным оператором*, а соответствующие главные компоненты φ_n приобретают дополнительное свойство центрирования (ортогональности единице)

$$\int_{D_2} \varphi_n(s) \mu_2(ds) = 0. \quad (8)$$

Еще один "статистический" вариант задания функции g — это "корреляционный" вариант, когда, определив функцию

$$\sigma_1^2(x) = \|f(\theta(x, \cdot)) - \mathcal{E}_2 f(x)\|_2^2$$

и считая ее отличной от нуля μ_1 -почти всюду, мы полагаем

$$g(x, s) = (f(\theta(x, s)) - \mathcal{E}_2 f(x)) / \sigma_1(x). \quad (9)$$

Здесь оператор $\mathbf{G}\mathbf{G}^*$ будет называться *корреляционным*, а равенство (2.8), естественно, будет сохраняться.

Восстановление функции f в обоих вариантах очевидно. В центральном ковариационном случае аналогом (2.2) является

$$f(\theta(x, s)) = \mathcal{E}_2 f(x) + \sum_n \sqrt{\lambda_n} \psi_n(x) \varphi_n(s). \quad (10)$$

Тем самым, кроме членов разложения (1.4) функции (2.6), проектированию в \mathbf{L}_θ^2 подлежит еще и функция $\mathcal{E}_2 f(x)$ ("среднее"), рассматриваемая как функция двух переменных x и s . Это дает возможность рассматривать реконструкцию функции f как "со средним", так и просто "по выбранным главным компонентам".

В "корреляционном" случае вместо (2.2) мы будем иметь

$$f(\theta(x, s)) = \mathcal{E}_2 f(x) + \sum_n \sqrt{\lambda_n} \sigma_1(x) \psi_n(x) \varphi_n(s) \quad (11)$$

с необходимостью проектировать на \mathbf{L}_θ^2 каждый член правой части (2.11).

С формальной точки зрения различие между (2.10) и (2.11) незначительно. Действительно, положим $\psi_n^\sigma = \sigma_1 \psi_n$ и $d\mu_1^\sigma = \sigma_1^{-2} d\mu_1$. Тогда функции ψ_n^σ будут ортонормированы в $L^2(\mu_1^\sigma)$ и по лемме 1.1 (2.11) является разложением Шмидта той же функции $f(\theta(x, s)) - \mathcal{E}_2 f(x)$, но не в L^2_{12} , а в $L^2(\mu_1^\sigma \otimes \mu_2)$. Таким образом, операция нормировки в (2.9) соответствует замене меры в пространстве признаков.

Основной особенностью разложения (2.10) состоит в том, что функция $\mathcal{E}_2 f$, вообще говоря, не является ортогональной собственным функциям ψ_n (см. лемму 1.1). В тоже время, если добавить к системе $\{\varphi_n\}_{n>0}$ константу $\varphi_0 = 1/\mu_2(D_2)$, то эта "расширенная" система главных компонент останется ортогональной.

Такая несимметричность признаков и индивидов, порожденная центрированием "по признакам" (2.6) и обычная в методе главных компонент, здесь выглядит довольно неестественно. Можно применять двойное центрирование

$$g(x, s) = f(\theta(x, s)) + \mathcal{E}_{1,2} f - \mathcal{E}_2 f(x) - \mathcal{E}_1 f(s), \quad (12)$$

где

$$\mathcal{E}_{1,2} f = \int_{D_1} \int_{D_2} f(\theta(x, s)) \mu_1(dx) \mu_2(ds) / \mu_1(D_1) \mu_2(D_2),$$

а $\mathcal{E}_1 f(s)$ определяется аналогично (2.7). Тогда вместо (2.2) получим

$$f \circ \theta = -\mathcal{E}_{1,2} f + \mathcal{E}_2 f + \mathcal{E}_1 f + \sum_n \sqrt{\lambda_n} \psi_n \otimes \varphi_n. \quad (13)$$

В многомерном статистическом анализе (2.13) соответствует (ковариационному) анализу главных компонент остатков ξ_{ij} в схеме дисперсионного анализа $x_{ij} = x.. + x_{i.} + x_{.j} + \xi_{ij}$. Оператор $\mathbf{G}\mathbf{G}^*$ будет называться здесь *бицентральным ковариационным оператором*. Отметим, что в бицентральном случае собственные функции и функции главных компонент являются ортогональными единице.

Конечно, нарушение симметрии признаков и индивидов при нормировке "по признакам" в (2.11) исправляется соответствующей заменой меры μ_2 в пространстве индивидов.

Рассмотрим теперь "идеальный" вариант описанной выше процедуры метода "Гусеница". Остановимся сначала на нецентральном случае. Предположим, что мы сгруппировали члены в правой части (2.2) соответственно трем непересекающимся группам индексов I_1 , I_2 и I_3 . Обозначив

$$g_i = \sum_{n \in I_i} \sqrt{\lambda_n} \psi_n \otimes \varphi_n, \quad (14)$$

мы получим разложение

$$f \circ \theta = g_1 + g_2 + g_3. \quad (15)$$

Будем считать, что g_3 представляет собой "шум" и интересоваться только g_1 и g_2 . Проведя проектирование g_1 и g_2 на L^2_θ , мы получим функции $f_1 \circ \theta$ и $f_2 \circ \theta$, и, по смыслу метода, должны относиться к найденным функциям f_1 и f_2 как к "существенным" и "принципиально различным" составляющим функции f . В частности, это означает, что в идеальном варианте f_1 и f_2 должны быть ортогональны в $L^2(d\nu)$. Кроме того, проектирование приводит к некоторой

нежелательной потере информации. Поэтому здесь естественно потребовать, чтобы $g_i \in \mathbf{L}_\theta^2$, то есть $g_i = f_i \circ \theta$ и, следовательно,

$$f \circ \theta = f_1 \circ \theta + f_2 \circ \theta + f_3 \circ \theta. \quad (16)$$

Равенство (2.16) обеспечивает и ортогональность функций f_i в $\mathbf{L}^2(d\nu)$. Тем самым мы пришли к условиям разложимости функции $f \circ \theta$ на слагаемые g_i , $i = 1, 2, 3$ (лемма 1.3). Кроме того, из практических соображений множества I_1 и I_2 должны быть конечными.

Таким образом, возникают следующие вопросы:

1. Как велик и сколь "естественен" класс функций f , для которых разложение (2.2) конечно?
2. Сколько велик класс пар "естественных" функций f_1 и f_2 , для которых слагаемые $g_1 = f_1 \circ \theta$ и $g_2 = f_2 \circ \theta$ разделяются разложением (2.2) в сумме $g = g_1 + g_2$?

Второй вопрос можно сформулировать по-другому. Пусть f_1 — общеупотребительная функция (линейная, вещественная или мнимая экспонента и т. д.). Сколько велик класс функций f_2 , для которых разделяются слагаемые в сумме $f_1 \circ \theta + f_2 \circ \theta$?

От ответа на эти вопросы зависит практический успех метода "Гусеница". Конечно, на практике идеальных вариантов не встречается, но небольшие их искажения приводят к малым искажениям результатов, и получившиеся функции, как показывает опыт, хорошо идентифицируются даже зрительно (см. главу I).

Отметим, что аналитические условия разделимости слагаемых приведены в (1.10) и (1.11).

Рассмотрим теперь центральный случай. Его особенность состоит в том, что "среднее" $\mathcal{E}_2 f(x)$ выступает как единое целое и не может быть распределено между слагаемыми. Поэтому в сумме (2.15) одно из слагаемых (скажем, g_1) должно иметь вид

$$g_1 = \mathcal{E}_2 f + \sum_{n \in I_1} \sqrt{\lambda_n} \psi_n \otimes \varphi_n,$$

а остальные — по-прежнему удовлетворять (2.14). При этом функции g_i должны, как и раньше, совпадать со своими проекциями на \mathbf{L}_θ^2 . Условия разделимости (1.10) и (1.11) сохраняются с заменой g_1 на $g_1 - \mathcal{E}_2 g_1$. Кроме того, функции g_i , $i > 1$, должны удовлетворять соотношениям

$$\mathcal{E}_2 g_i = 0. \quad (17)$$

Здесь, однако, появилась некоторая несимметрия между функциями g_1 и g_i при $i > 1$ (g_i , $i > 1$ всегда является центрированной, а g_1 , вообще говоря, нет). Поэтому особую роль играют функции, инвариантные относительно центрирования (то есть удовлетворяющие (2.17)). Если g_1 инвариантна относительно центрирования, то g_1 и g_i "меняются местами" — g_i уже не обязана удовлетворять (2.17).

В бицентральном случае все происходит аналогично. Отметим только, что условие (2.17) переходит в

$$\mathcal{E}_2 g_i = 0, \quad \mathcal{E}_1 g_i = 0. \quad (18)$$

Поскольку мы предполагаем, что $g_i = f_i \circ \theta$, то отсюда (как и при однократном центрировании) следует равенство

$$\int_{[0,t]} f_i d\nu = 0.$$

Замечание о несимметрии g_1 и $g_i, i > 1$, сохраняется с заменой (2.17) на (2.18). Отметим также, что некоторые функции (константы при однократном центрировании и линейные функции — при двукратном) переходят в ноль в результате центрирования. Это означает, что они отделимы от любой функции g_i , удовлетворяющей (2.17) или (2.18).

3 Аддитивная "Гусеница". Примеры и варианты

В этом разделе мы обсудим свойства аддитивного варианта метода "Гусеница", определяемого равенством $\theta(x, s) = x + s$, где $x \in [0, \tau]$, $s \in [0, t - \tau]$ и $\tau \in (0, t)$. В основном нас будут интересовать два вопроса — описание класса функций f , порождающих конечные ортогональные системы и выделение "простых" слагаемых в сумме. Кроме того, мы обсудим роль различных вариантов центрирования.

Следует отметить, что для аддитивной "Гусеницы" важную роль играет число τ — длина "Гусеницы". Поэтому мы будем подчеркивать эту роль небольшим изменением обозначений (писать g_τ вместо g , \mathbf{G}_τ вместо \mathbf{G} и тому подобное). Кроме того, мы ограничимся случаем непрерывных функций f , что с практической точки зрения является достаточно реалистическим предположением, а функции φ_n и ϕ_n оказываются при этом непрерывными.

Итак, пусть у нас есть число $\tau \in (0, t)$ — длина гусеницы. В нецентральном случае функция g_τ определяется равенством

$$g_\tau(x, s) = f(x + s), \quad (1)$$

где $x \in [0, \tau]$, $s \in [0, t - \tau]$. Пусть μ^λ — семейство конечных мер на борелевских подмножествах отрезков $[0, \lambda]$, $\lambda \leq t$, причем $\mu_1 = \mu^\tau$ и $\mu_2 = \mu^{t-\tau}$. Определив оператор \mathbf{G}_τ равенством (1.1), обозначим для краткости $\mathbf{R}_\tau = \mathbf{G}_\tau \mathbf{G}_\tau^*$.

Отметим прежде всего следующее обстоятельство. Так как, очевидно, выполняется $g_\tau(x, s) = g_{t-\tau}(s, x)$, то операторы \mathbf{G}_τ^* и $\mathbf{G}_{t-\tau}$ совпадают. Поэтому $\psi_n^\tau = \varphi_n^{t-\tau}$, что с практической точки зрения означает, что нам достаточно рассматривать длины гусеницы, не превосходящие $t/2$. Это замечание остается верным и в случае двойного центрирования. При однократном центрировании оно, естественно, уже не верно.

3.1 Конечномерные собственные подпространства

Для начала заметим, что некоторые функции f всегда порождают конечномерные собственные подпространства вне зависимости от длины гусеницы τ и мер μ^λ . Легко доказываются следующие утверждения.

1. Если f — полином степени t , то для любого τ и любых конечных мер μ^λ все ненулевые собственные функции ψ_n^τ (и функции главных компонент φ_n^τ) оператора \mathbf{R}_τ являются полиномами степени не большей, чем t . Тем самым

размерность ненулевого собственного подпространства \mathbf{R}_τ не превосходит $m + 1$.

Этот факт следует из того, что ядро $R_\tau(x, z)$ оператора \mathbf{R}_τ является полиномом по каждой из переменных x и z и, следовательно, $\mathbf{R}_\tau h$ также является полиномом степени не большей, чем m , для любой функции $h \in \mathbf{L}_\tau^2$,

2. *Вещественная экспонента $f(y) = \exp(\beta y)$ порождает (не более чем) одномерное ненулевое собственное подпространство оператора \mathbf{R}_τ , при этом $\psi_1^\tau(x) = c_1(\tau) \exp(\beta x)$ и $\varphi_1^\tau(s) = c_2(\tau) \exp(\beta s)$ соответственно μ^τ и $\mu^{t-\tau}$ почти всюду.*

3. *Если $f(y) = \sin(\omega y + \varphi)$, то ненулевое собственное подпространство \mathbf{R}_τ будет не более чем двумерным.*

Действительно, как нетрудно видеть, в этом случае при любых τ и μ^λ ненулевые собственные функции оператора \mathbf{R}_τ являются линейными комбинациями $\sin(\omega x)$ и $\cos(\omega x)$. Это же верно, конечно, и для функций главных компонент.

Очевидно, линейные комбинации произведений этих трех функций (то есть полиномов и вещественных и мнимых экспонент) также будут порождать конечномерные ненулевые собственные подпространства.

Во всех этих случаях, очевидно, центрирование (однократное или двукратное) не меняет сам факт конечномерности собственных подпространств. Более того, так как нормировка есть просто замена меры в пространстве признаков, то и в этом случае описанный класс функций сохраняет свои свойства. Размерность собственного пространства при этом, конечно, может измениться.

Ситуация, когда меры μ^λ являются взаимно абсолютно непрерывными с соответствующими мерами Лебега, а функция f — достаточно гладкой, является хорошо изученной. В этом случае (2.2) превращается в поточечное равенство

$$f(x + s) = \sum_{n=1}^m \sqrt{\lambda_n} \psi_n^\tau(x) \varphi_n^\tau(s),$$

$x \in [0, \tau], s \in [0, t - \tau]$. Если $f \in \mathbf{C}^{(m)}([0, t])$, то ψ_i^τ и φ_i^τ также будут m раз дифференцируемы. Отсюда легко выводится ([5]), что f удовлетворяет дифференциальному уравнению

$$\mathfrak{D}f = 0, \tag{2}$$

где $\mathfrak{D}f = \sum_{i=1}^m b_i f^{(i)}$, а коэффициенты b_i постоянны. Хорошо известно ([7], §13), что общее решение (3.2) представляет собой некоторую линейную комбинацию произведений полиномов, экспонент, а также синусов и косинусов. Аналогичный результат имеет место и для считающих мер μ_1 и μ_2 ([5]).

Таким образом, мы имеем достаточно полное описание класса функций, порождающих конечномерные собственные подпространства для аддитивной "Гусеницы".

3.2 Выделение слагаемых

3.2.1 Общие условия

Остановимся теперь на условиях выделения слагаемых в сумме. В отличие от предыдущих примеров, где роль мер μ_1 и μ_2 была относительно небольшая, условия разделимости существенно зависят от этих мер. Рассмотрим сначала

случай, когда μ_1 и μ_2 являются мерами Лебега. Поскольку $f_1, f_2 \in \mathbf{C}([0, t])$, то равенства (1.10) и (1.11) превращаются в поточечные и упрощаются. В самом деле, для непрерывных f_1 и f_2 и аддитивной "Гусеницы" (1.11) переходит в равенство

$$\int_0^\tau f_1(x+u)f_2(x+v) dx = 0$$

$(u, v \in [0, t - \tau])$, которое заменой переменных сводится к

$$\int_v^{\tau+v} f_1(\delta+z)f_2(z) dz = 0 \quad (3)$$

для любых $v \in [0, t - \tau]$ и $\delta \in [-v, t - \tau - v]$. В свою очередь (3.3) эквивалентно равенству

$$\int_0^\tau f_1(\delta+z)f_2(z) dz = 0 \quad (4)$$

при любом $\delta \in [0, t - \tau]$ и "условию периодичности"

$$f_1(\delta + \tau + v)f_2(\tau + v) = f_1(\delta + v)f_2(v), \quad (5)$$

которое должно выполняться при всех $v \in [0, t - \tau]$ и $\delta \in [-v, t - \tau - v]$. Аналогично (1.10) переходит в

$$\int_0^{t-\tau} f_1(\alpha+y)f_2(y) dy = 0 \quad (6)$$

при любом $\alpha \in [0, \tau]$ и

$$f_1(\alpha + t - \tau + y)f_2(t - \tau + y) = f_1(\alpha + y)f_2(y) \quad (7)$$

при всех $y \in [0, \tau]$ и $\alpha \in [-y, \tau - y]$.

Случай абсолютно непрерывных мер μ^λ представляет, как правило, лишь методический интерес. На практике наблюдаются и обрабатываются временные ряды с дискретным временем и меры μ^λ в этом случае должны быть дискретны и сосредоточены на правильной решетке с некоторым шагом δ . Не умоляя общности, мы положим $\delta = 1$, так что параметр λ (а также числа t и τ) будет принимать только целые значения.

Здесь нам будет удобно несколько изменить обозначения. Положим $t = n - 1$, (так что решетка имеет вид $\{0, \dots, n - 1\}$ и $0 \leq \tau \leq n - 1$), и будем называть *длиной гусеницы* число $d = \tau + 1$. Соответственно введем $h = \lambda + 1$ и будем писать μ^h вместо μ^λ и т. д.

Выберем меры μ^h , сосредоточенные на множествах $\{0, \dots, h - 1\}$ и приписывающие каждому элементу этих множеств вес, равный единице, и кратко остановимся на выделении слагаемых из суммы.

Условия разделимости функций f_1 и f_2 (теперь эти функции достаточно рассматривать только в точках $0, \dots, n - 1$) превратятся в следующие равенства

$$\sum_{i=0}^{d-1} f_1(i+j)f_2(i+k) = 0 \quad (8)$$

для любых $0 \leq j, k \leq n - d$ и

$$\sum_{j=0}^{n-d} f_1(i+j) f_2(k+j) = 0 \quad (9)$$

для любых $0 \leq i, k \leq d-1$. Иначе говоря, если F_1 и F_2 — матрицы, составленные из элементов $f_1(i+j)$ и $f_2(i+j)$ соответственно ($0 \leq i \leq d-1, 0 \leq j \leq n-d$), то условиями разделимости f_1 и f_2 будут являться равенства нулю матриц $F_1^T F_2$ и $F_1 F_2^T$. Аналоги (3.4)–(3.7) также легко записываются.

Кратко остановимся на центральном и бицентральном вариантах. Рассмотрим сначала меру Лебега и введем два условия.

1. Условие A1.

Будем говорить, что функция f удовлетворяет Условию A1, если

$$\int_0^{t-\tau} f(s) ds = 0 \quad (10)$$

и f периодична с периодом $t - \tau$.

2. Условие A2.

Функция f удовлетворяет условию A2, если

$$\int_0^{\tau_1} f(y) dy = 0,$$

f имеет некоторый период τ_1 , а числа t и τ пропорциональны τ_1 .

Как уже указывалось, использование центрального ковариационного оператора для аддитивной гусеницы связано с заданием функции g в виде

$$g(x, s) = f(x + s) - \mathcal{E}_{t-\tau} f(x), \quad (11)$$

где

$$\mathcal{E}_{t-\tau} f(x) = \frac{1}{t-\tau} \int_0^{t-\tau} f(x+s) ds. \quad (12)$$

Условие инвариантности относительно центрирования сводится к тождеству

$$\int_0^{t-\tau} f(x+s) ds = 0, \quad x \in [0, \tau]. \quad (13)$$

Тем самым имеет место следующее утверждение.

Предложение 3.1 *Функция f является инвариантной относительно однократного центрирования тогда и только тогда, когда она удовлетворяет условию A1.*

Если f_1 удовлетворяет A1, условия разделимости принимают вид

$$\int_0^\tau f_1(x+u) f_2(x+v) dx = \int_0^\tau f_1(x+u) \mathcal{E}_{t-\tau} f_2(x) dx \quad (14)$$

и

$$\int_0^{t-\tau} f_1(z+s) f_2(x+s) ds = 0. \quad (15)$$

Отсюда, в частности, следует, что функция f_2 также обладает периодом $t - \tau$, и, следовательно, $\mathcal{E}_{t-\tau}f_2 = const.$

При двойном центрировании условия инвариантности превращаются в

$$\int_0^\tau f_2(z) dz = 0, \quad f_2(\tau + v) = f_2(v), \quad (16)$$

и

$$\int_0^{t-\tau} f_2(y) dy = 0, \quad f_2(t - \tau + y) = f_2(y). \quad (17)$$

Это означает, что доказан следующий факт.

Предложение 3.2 *Функция f является инвариантной относительно двойного центрирования тогда и только тогда, когда f удовлетворяет условию А2.*

Если f_1 удовлетворяет условию А2, то условия разделимости превращаются в (3.14) и

$$\int_0^{t-\tau} f_1(x + s)(f_2(y + s) - \mathcal{E}_\tau f_2(s)) ds = 0 \quad (18)$$

при любых $x, s \in [0, \tau]$.

В дискретном случае все происходит совершенно аналогично и все интегралы переходят в соответствующие суммы.

Перейдем теперь непосредственно к примерам.

3.2.2 Выделение константы

Наша задача — найти семейство функций, отделимых от константы.

1. *Нецентральный ковариационный оператор. Мера Лебега.*

Пусть $f_1(x) = c > 0$. Ненулевое собственное подпространство в этом случае будет одномерным, причем $\lambda_1 = c^2\tau(t - \tau)$, собственная функция имеет вид $\psi_1 = 1/\sqrt{\tau}$, а функция первой главной компоненты φ_1 равна $1/\sqrt{t - \tau}$.

Условия (3.4)–(3.7), как нетрудно видеть, в точности соответствуют инвариантности функции $g_2(x, s) = f_2(x + s)$ относительно двукратного центрирования, то есть (3.16) и (3.17). Следовательно, (ненулевая) константа отделяется от ненулевой непрерывной функции f тогда и только тогда, когда f удовлетворяет условию А2.

2. *Центральный ковариационный оператор. Мера Лебега.*

Как уже говорилось, при центрировании константа переходит в нуль. Следовательно, константа отделяется от любой ненулевой непрерывной функции, удовлетворяющей А1.

Отметим, что условия разделимости стали слабее по сравнению с нецентрализмом случаем. А именно, ослабло ограничение на длину ряда: если t больше периода функции, то всегда можно подобрать длину гусеницы τ , обеспечивающую разделимость.

3. *Бицентральный ковариационный оператор. Мера Лебега.*

Как и в центральном случае, двойное центрирование (2.12) превращает константу в тождественный ноль. Тем самым условия выделения константы в бицентральном случае оказываются такими же, как и в нецентрализмом.

4. *Дискретная мера.*

Здесь и ниже мы будем весьма кратко останавливаться на дискретном случае, так как он вполне аналогичен абсолютно непрерывному.

Для нецентрального ковариационного оператора из (3.8) и (3.9) легко получается, что *ненулевая константа отделяется от ненулевой функции f тогда и только тогда, когда f имеет некоторый период d_1 , n и d делятся на d_1* и

$$\sum_{j=0}^{d_1-1} f(j) = 0. \quad (19)$$

В бицентральном случае этот результат сохраняется. Что касается центрального, то здесь функция f должна иметь период $n - d$ и удовлетворять равенству

$$\sum_{j=0}^{n-d-1} f(j) = 0.$$

3.2.3 Выделение линейной функции

Займемся теперь условиями выделения линейной функции.

1. *Нецентральный ковариационный оператор. Мера Лебега.*

Как уже говорилось, линейная функция $f(y) = ay + c, a \neq 0$ порождает двумерное собственное подпространство, причем собственные функции ψ_i и функции главных компонент $\varphi_i, i = 1, 2$, также являются линейными. Условие (3.5) превращается в

$$(a(\delta + \tau + v) + c)f_2(\tau + v) = (a(\delta + v) + c)f_2(v). \quad (20)$$

Дифференцируя обе части этого равенства по δ , получим, что $f_2(\tau + v) = f_2(v)$, откуда сразу же следует, что $f_2 \equiv 0$. Следовательно, *линейная функция, не являющаяся константой, не отделима ни от одной ненулевой непрерывной функции f .* Естественно, тот же результат получится и для полиномов больших степеней.

2. *Центральный ковариационный оператор. Мера Лебега.*

Преобразование (3.11) приводит к

$$g_1(x, s) = a(s - (t - \tau)/2).$$

Собственное подпространство, как нетрудно видеть, при этом становится одномерным, причем функция главной компоненты φ_1 остается линейной, а собственная функция ψ_1 превращается в константу.

Как уже указывалось, необходимым условием отделимости от инвариантной относительно однократного центрированной функции является периодичность. Поэтому *при однократном центрировании линейная функция по-прежнему не отделяется ни от одной ненулевой непрерывной функции.*

3. *Бицентральный ковариационный оператор. Мера Лебега.*

Двойное центрирование переводит линейную функцию в ноль. Поэтому условия выделения линейной функции совпадает с условием А2. Заметим, что полиномы старших степеней по-прежнему не будут отделяться ни от одной непрерывной функции.

4. *Дискретная мера.*

Как и в непрерывном случае, в нецентральном и центральном вариантах "Гусеницы" линейная функция не отделяется ни от одной ненулевой функции f . При

двойном центрировании условие отделимости линейной функции совпадает с условием выделения константы.

3.2.4 Выделение вещественной экспоненты

Перейдем к рассмотрению вещественной экспоненты.

1. *Нецентральный ковариационный оператор. Мера Лебега.*

В случае $f_1(y) = \exp(\beta x)$, $\beta \neq 0$, равенства (3.4)–(3.7) имеют вид

$$\int_0^\tau e^{\beta z} f_2(z) dz = 0, \quad e^{\tau+v} f_2(\tau + v) = e^v f_2(v),$$

и

$$\int_0^{t-\tau} e^{\beta y} f_2(y) dy = 0, \quad e^{\beta(t-\tau+y)} f_2(t - \tau + y) = e^{\beta y} f_2(y).$$

Эти равенства в точности соответствуют условиям выделения константы в сумме

const + $e^{\beta y} f_2(y)$. Таким образом, *вещественная экспонента $e^{\beta y}$ отделяется от ненулевой непрерывной функции f тогда и только тогда, когда*

- a) *функция $e^{\beta y} f(y)$ является периодической на $[0, t]$ с периодом τ_1 ;*
- b) *числа τ и t пропорциональны τ_1 ;*
- c)

$$\int_0^{\tau_1} e^{\beta y} f(y) dy = 0.$$

Например, если исходная сумма имеет вид $e^{\beta y} + e^{-\beta y} \cos(2\pi y/\tau_1)$, а t и τ пропорциональны τ_1 , то (аддитивная нецентральная) гусеница выделит в этой сумме экспоненту.

2. *Центральный ковариационный оператор. Мера Лебега.*

Функция g в этом случае принимает вид

$$g_1(x, s) = e^{\beta x} (e^{\beta s} - c_{t-\tau}),$$

где $c_r = (e^{\beta r} - 1)/\beta r$, $r > 0$. Так же, как и в случае линейной функции, при однократном центрировании вещественная экспонента не отделяется ни от одной ненулевой непрерывной функции, так как не удовлетворяет необходимому условию периодичности. Собственное подпространство остается одномерным.

3. *Бицентральный ковариационный оператор. Мера Лебега.*

Функция g_1 имеет вид

$$g_1(x, s) = (e^{\beta x} - c_\tau) (e^{\beta s} - c_{t-\tau}).$$

Условия разделимости переходят в равенства

$$\int_0^\tau e^{\beta x} f_2(x + s) dx = 0, \quad \int_0^{t-\tau} e^{\beta s} f_2(x + s) ds = 0,$$

а также

$$\int_0^{t-\tau} f_2(x + s) ds = 0, \quad \int_0^\tau f_2(x + s) dx = 0.$$

Первые два из этих четырех равенств в точности совпадают с условиями выделения экспоненты в нецентральном случае. В частности, из них вытекает, что функция $e^{\beta y} f_2(y)$ является периодической с периодом $t - \tau$. В то же время третье условие имеет следствием периодичность самой функции f_2 с тем же периодом. Следовательно, выделение экспоненты снова оказывается невозможным. Собственное подпространство остается одномерным.

4. Дискретная мера.

Как и раньше, в нецентральном случае вещественная экспонента $f_1(k) = e^{\beta k}$ отделяется от функции $f_2(k)$ тогда и только тогда, когда функция $e^{\beta k} f_2(k)$ отделяется от константы. Легко провести и более детальную расшифровку этих условий, абсолютно аналогичную непрерывному случаю. Оба центральных варианта приводят к невозможности выделения вещественной экспоненты.

3.2.5 Выделение мнимой экспоненты.

Перейдем теперь к последнему примеру — мнимой экспоненте.

1. Нецентральный ковариационный оператор. Мера Лебега.

Пусть $f_1(y) = \sin(\omega y + \alpha)$. Как обычно в этом случае, все вычисления упрощаются, если мы введем соответствующую мнимую экспоненту. Таким образом, положим $f_1(y) = e^{i\omega y}$.

Собственное подпространство для мнимой экспоненты будет двумерным (напомним, что мы рассматриваем вещественные собственные функции), а собственные функции и функции главных компонент будут линейными комбинациями синусов и косинусов с той же самой частотой ω . Отметим, что в случае, когда t и τ кратны периоду T мнимой экспоненты, эти линейные комбинации выродятся в соответствующие синусы и косинусы.

Что касается выделения функции $e^{i\omega y}$, то все результаты предыдущего пункта сохранятся с заменой вещественной экспоненты на мнимую. Отсюда легко получается, что *мнимая экспонента $e^{i\omega y}$ с периодом T отделяется от вещественной ненулевой непрерывной функции функции f тогда и только тогда, когда t и τ кратны T , функция f имеет период T* и

$$\int_0^T e^{i\omega y} f(y) dy = 0. \quad (21)$$

2. Центральный ковариационный оператор. Мера Лебега.

Пусть $f_1(y) = e^{i\omega y}$. Сразу же заметим, что условие инвариантности f_1 относительно центрирования сводится в данном случае к

$$\omega = 2\pi k/(t - \tau), \quad (22)$$

k — целое. Иначе говоря, $t - \tau$ кратно периоду $T = 2\pi/\omega$. Предположим сначала, что это условие не выполняется. Тогда функция f_2 должна удовлетворять А1,

$$\int_0^\tau e^{i\omega(x+v)} f_2(x + v) dx = 0,$$

и

$$\int_0^{t-\tau} e^{i\omega(x+s)} f_2(y + s) ds = 0.$$

Второе из этих равенств, в частности, имеет следствием условие периодичности $e^{i\omega(t-\tau+x)} f_2(t - \tau + x) = e^{i\omega x} f_2(x), x \in [0, \tau]$. Так как f_2 имеет период $t - \tau$, то

отсюда сразу же следует, что ω имеет вид (3.22) при некотором целом k . Таким образом, функция f_1 обязана быть инвариантной относительно центрирования.

Пусть теперь (3.22) выполнено. Следовательно, искомая функция f_2 не обязана быть инвариантной относительно операции центрирования и условия разделимости имеют вид (3.14) и (3.15). В данном случае это означает, что f_2 имеет период $t - \tau$,

$$\int_0^{t-\tau} e^{i\omega s} f_2(s) ds = 0, \quad (23)$$

а также

$$\int_0^\tau e^{i\omega x} \tilde{f}_2(x + v) dx = 0, \quad (24)$$

где $\tilde{f}_2 = f_2 - B_{t-\tau}$, $B_{t-\tau} = \mathcal{E}_{t-\tau} f_2(x) = const$ и $v \in [0, t - \tau]$. В свою очередь (3.24) эквивалентно периодичности функции f_2 с периодом τ и

$$\int_0^\tau e^{i\omega x} \tilde{f}_2(x) dx = 0. \quad (25)$$

Отсюда сразу же следует, что t и τ кратны некоторому τ_1 и f_2 имеет период τ_1 . Отметим, что ввиду (3.22) равенство (3.23) выполняется автоматически.

Рассмотрим (3.25). Нетрудно видеть, что оно сводится к

$$(e^{i\omega\tau} - 1) \int_0^{\tau_1} e^{i\omega z} \tilde{f}_2(z) dz = 0.$$

Равенство $e^{i\omega\tau} = 1$ эквивалентно требованию, чтобы τ было кратно периоду T . Если же $e^{i\omega\tau} \neq 1$, то (3.25) перейдет в (3.21) с заменой f на \tilde{f}_2 .

Таким образом мы получили, что в центральном случае для отдельности мнимой экспоненты $e^{i2\pi y/T}$ от непрерывной функции f необходимы и достаточные следующие условия:

1. функция f является периодической;
2. числа t и τ кратны периоду функции f ;
3. $t - \tau$ кратно периоду T мнимой экспоненты;
4. если τ не кратно T , то \tilde{f} удовлетворяет (3.21).

Тем самым есть два случая выделения мнимой экспоненты. В одном из них длина ряда и длина гусеницы кратны как периоду экспоненты, так и периоду функции f . В другом — длины ряда и гусеницы кратны периоду f , их разность кратна периоду мнимой экспоненты, а центрированная функция \tilde{f} удовлетворяет (3.21).

3. Бицентральный ковариационный оператор. Мера Лебега.

Как нетрудно видеть, условия инвариантности относительно двойного центрирования для мнимой экспоненты с периодом T сводятся к тому, что t и τ кратны T . Пусть эти условия не выполняются. Тогда, используя условия А2 инвариантности функции f_2 и действуя так же, как в случае однократного центрирования, мы приходим к противоречию.

Итак, пусть t и τ кратны T . Условия разделимости мнимой экспоненты и функции f_2 переходят в

$$\int_0^\tau e^{i\omega x} f_2(x + s) dx = \int_0^\tau e^{i\omega x} \mathcal{E}_{t-\tau} f_2(x) dx,$$

$$\int_0^{t-\tau} e^{i\omega s} f_2(x+s) ds = \int_0^{t-\tau} e^{i\omega s} \mathcal{E}_\tau f_2(s) ds.$$

Предположим, что $f_2 \in \mathbf{C}^{(1)}([0, t])$, так что предыдущие равенства можно дифференцировать соответственно по s и x под знаком интеграла. Сразу же получаем, что производная f'_2 функции f_2 имеет период T и удовлетворяет равенству (3.21). Следовательно, сама функция f_2 имеет вид

$$f_2(y) = ay + b + \tilde{f}_2(y),$$

где функция \tilde{f}_2 также имеет период T и удовлетворяет (3.21). Тем самым мы получили, что *мнимая экспонента с периодом T отделяется от функции $f \in \mathbf{C}^{(1)}([0, t])$ тогда и только тогда, когда t и τ кратны T , а функция f имеет вид $ax + b + \tilde{f}$, где \tilde{f} имеет период T и удовлетворяет (3.21).*

Заметим, что эти условия согласуются с условиями выделения линейной функции.

4. Дискретная мера

Как всегда, случай дискретной меры вполне аналогичен варианту с мерой Лебега. Выписывание конкретных условий выделения мнимой экспоненты не предоставляет труда.

4 Приближенное выделение слагаемых

В предыдущем разделе было показано, что условия разделимости слагаемых в сумме являются весьма ограничительными. На практике, однако, наблюдается обратное (см. иллюстрации к предыдущим главам). Это объясняется тем, что "практическая" разделимость слагаемых является приближенной. Как будет показано ниже, условия такой приближенной разделимости являются гораздо более слабыми.

4.1 Устойчивость метода "Гусеница"

Все предыдущие результаты относились к "идеальному" разделению слагаемых. На практике мы встречаемся с искажениями и шумами, и нужно быть уверенным, что при сравнительно небольших погрешностях результат работы метода "Гусеница" будет не слишком сильно отличаться от "идеального".

Введем сначала несколько обозначений. Пусть h_1 и h_2 — две функции, принадлежащие \mathbf{L}_1^2 , причем $\|h_1\|_1 \|h_2\|_1 \neq 0$. Положим

$$\rho_1(h_1, h_2) = \frac{\langle h_1, h_2 \rangle_1}{\|h_1\|_1 \|h_2\|_1}$$

и будем называть величину $\rho_1(h_1, h_2)$ (некентральным) коэффициентом корреляции между h_1 и h_2 . Аналогичным образом будем рассматривать коэффициенты корреляции в \mathbf{L}_2^2 .

Общие условия разделимости слагаемых в сумме являются условиями ортоизональности функций $g_1(\cdot, u)$ и $g_2(\cdot, v)$ в \mathbf{L}_1^2 для $\mu_2^{\otimes 2}$ -почти всех (u, v) , а также функций $g_1(x, \cdot)$ и $g_2(y, \cdot)$ в \mathbf{L}_2^2 для $\mu_2^{\otimes 2}$ -почти всех (x, y) .

В связи с этим дадим следующее определение. Будем говорить, что функции $g_1(x, s)$ и $g_2(x, s)$ являются ε -разделимыми в сумме $g = g_1 + g_2$, если функции

$|\rho_1(g_1, g_2)(u, v)|$ и $|\rho_2(g_1, g_2)(x, y)|$ не превосходят ε при почти всех (u, v) и (x, y) соответственно.

Пусть теперь функции g_1 и g_2 являются разделимыми, $\tilde{g}_1 = g_1 + \varepsilon_1$, $\tilde{g}_2 = g_2 + \varepsilon_2$ и "погрешности" $|\varepsilon_1|, |\varepsilon_2|$ равномерно не превосходят $\delta/2$. В этом случае можно ожидать, что (главные) собственные функции, собственные числа и функции главных компонент интересующих нас возмущенных операторов будут мало отличаться от соответствующих характеристик невозмущенных операторов.

Предположим дополнительно, что при $i = 1, 2$ $\|g_i(\cdot, u)\|_1 > \delta$ равномерно по $u \in D_2$ (поскольку нас интересуют непрерывные функции g_i , это не является слишком ограничительным предположением) и аналогичные неравенства имеют место для $g_i(x, \cdot)$ при $x \in D_1$. Тогда нетрудно видеть, что функции $\rho_i(\tilde{g}_1, \tilde{g}_2)$, $i = 1, 2$, будут иметь вид $O(\delta)$. Таким образом, понятие разделимости слагаемых в сумме является устойчивым относительно небольших "шумов".

Можно ввести и другое определение ε -разделимости, если рассматривать малость коэффициентов корреляции не поточечно, а в метриках соответствующих L^2 . Конечно, понятие "малости" погрешностей должно в этом случае измениться. Аналогичным образом исследуется устойчивость разделимости относительно небольших изменений длины "Гусеницы" и длины ряда.

4.2 Асимптотическая разделимость слагаемых

Гораздо более важной в практическом плане является асимптотическая разделимость слагаемых. Введем соответствующие понятия и определения для аддитивной "Гусеницы", обобщение на общий случай проводится непосредственно.

Пусть $f : [0, \infty) \mapsto \mathbb{R}$ — некоторая измеримая функция, f_t — ее сужение на $[0, t]$ и μ^t , $t > 0$ — семейство вероятностных мер на σ -алгебре борелевских подмножеств $[0, t]$. При каждом t зададим длину "Гусеницы" $\tau = \tau(t) \in (0, t)$, положим $\mu_1 = \mu^\tau$, $\mu_2 = \mu^{t-\tau}$ и (в нецентральном случае) определим функцию g_τ равенством

$$g_\tau(x, s) = f_t(x + s), \quad (1)$$

$x \in [0, \tau]$, $s \in [0, t - \tau]$, предполагая, что $g_\tau \in L^2_{1,2}$ при любом t . Естественно потребовать, чтобы меры μ^t были нормированными сужениями на $[0, t]$ некоторой меры μ , определенной на борелевских подмножествах $[0, \infty)$.

Пусть $f = f_1 + f_2$, $g_\tau^{(1)}$ и $g_\tau^{(2)}$ определены аналогично (4.1). Будем говорить, что функции f_1 и f_2 *асимптотически разделяются при $t \rightarrow \infty$* , если функции $\rho_1^{(t)}(g_\tau^{(1)}, g_\tau^{(2)})(u, v)$ и $\rho_2^{(t)}(g_\tau^{(1)}, g_\tau^{(2)})(x, y)$ стремятся к нулю при $t \rightarrow \infty$ соответственно почти всюду относительно (u, v) и (x, y) . Иначе говоря, асимптотическая разделимость $g_\tau^{(1)}$ и $g_\tau^{(2)}$ означает, что для любого ε существует такое t_0 , что при $t > t_0$ функции $g_\tau^{(1)}$ и $g_\tau^{(2)}$ являются ε -разделимыми.

С практической точки зрения условия асимптотической разделимости являются гораздо менее жесткими, чем весьма ограничительные условия точной разделимости. Поясним это на простейших примерах, связанных с нецентральной "Гусеницей" в том случае, когда μ является мерой Лебега.

Предположим, что $\tau(t) \rightarrow \infty$ и $t - \tau(t) \rightarrow \infty$ при $t \rightarrow \infty$. Тогда, ввиду полной симметрии "индивидуов" и "признаков", условие асимптотической разделимости

непрерывных функций f и g сводится к единственному требованию

$$\frac{\int_0^T f(x+s)g(x+s)ds}{\sqrt{\int_0^T f^2(x+s)ds} \sqrt{\int_0^T g^2(x+s)ds}} \xrightarrow{T \rightarrow \infty} 0, \quad (2)$$

которое должно выполняться при почти всех $x \geq 0$ по мере Лебега.

Конечно, условие (4.2) легко проверяется в конкретных случаях. Приведем соответствующие примеры.

1. Отделимость от константы.

- Любая функция вида

$$g(x) = \sum_{k=0}^N c_k \cos(\omega_k x), \quad \omega_k \neq 0, \quad (3)$$

асимптотически отделима от константы;

- Вещественная экспонента $g(x) = \exp(\alpha x)$ с $\alpha \neq 0$ асимптотически отделима от константы;
- Любой полином

$$g(x) = \sum_{k=0}^N c_k x^k, \quad c_N \neq 0,$$

не отделяется от константы. Любопытно, что предел в (4.2) для полинома имеет вид $\sqrt{2N+1}/N$, то есть чем больше степень полинома, тем степень его отделимости от константы лучше.

2. Отделимость от суммы косинусов (4.3).

- Любой полином отделим от функции (4.3);
- Любая вещественная экспонента с ненулевым показателем отделима от функции (4.3).

3. Отделимость от полинома.

- Любая вещественная экспонента с ненулевым показателем отделима от любого полинома;
- Любые два ненулевых полинома асимптотически не разделяются.

4. Отделимость от вещественной экспоненты.

- Вещественные экспоненты с показателями α и β асимптотически разделяются тогда и только тогда, когда $\alpha = -\beta \neq 0$.

Кратко остановимся на двух центральных вариантах аддитивной "Гусеницы", где существенную роль играет понятие инвариантности относительно центрирования.

Здесь это понятие, конечно, также должно принять асимптотическую форму.

В случае однократного центрирования и меры Лебега асимптотическая инвариантность функции f относительно центрирования означает, что

$$\frac{1}{t-\tau} \int_0^{t-\tau} f(x+s) ds \rightarrow 0 \quad (4)$$

при $t \rightarrow \infty$ для почти всех s по мере Лебега. В бицентральном случае к (4.4) прибавляется соотношения

$$\frac{1}{\tau} \int_0^\tau f(x+s) dx \rightarrow 0$$

при $t \rightarrow \infty$ для почти всех s по мере Лебега, а также

$$\frac{1}{\tau(t-\tau)} \int_0^\tau dx \int_0^{t-\tau} ds f(x+s) \rightarrow 0. \quad (5)$$

Отсюда, например, сразу же следует, что при однократном центрировании достаточными условиями отделимости константы от функции f является (4.4) и $t-\tau \rightarrow \infty$ (то есть длина ряда должна быть большой, а длина гусеницы может быть маленькой), а при двукратном центрировании достаточные условия отделимости линейной функции могут быть записаны в виде

$$\frac{1}{T} \int_0^T f(x+s) ds \xrightarrow{T \rightarrow \infty} 0,$$

(4.5) и $\tau \rightarrow \infty$, $t-\tau \rightarrow \infty$.

4.3 Выделение сигнала из шума

В статистике естественной моделью шумов являются случайные процессы, обладающие некоторыми условиями эргодичности. В нашем случае введение такой модели приводит к понятию асимптотической стохастической разделимости слагаемых в сумме. Дадим соответствующие определения, снова ограничиваясь вариантом с мерой Лебега.

Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — некоторое вероятностное пространство, $f_1(s)$, $f_2(s)$ — два случайных процесса, определенных при $s \geq 0$. Положим $f = f_1 + f_2$ и, выбрав некоторую функцию $\tau = \tau(t) \in (0, t)$, определим случайные поля g_τ , $g_\tau^{(1)}$ и $g_\tau^{(2)}$ аналогично (4.1). Будем предполагать, что эти функции суммируемы с квадратом по прямому произведению соответствующих мер Лебега для любого $t > 0$ при почти всех $\omega \in \Omega$. Тем самым, как и в детерминированном случае, мы можем рассматривать коэффициенты корреляции $\rho_1^{(t)}$ и $\rho_2^{(t)}$ между случайными полями $g_\tau^{(1)}$ и $g_\tau^{(2)}$, которые сами также являются случайными полями над $[0, \tau]^2$ и $[0, t-\tau]^2$. Естественно, здесь должны быть наложены дополнительные условия существования этих коэффициентов корреляции для почти всех траекторий $f_1(s)$ и $f_2(s)$. В конкретных условиях, однако, формализация и проверка таких условий не представляет труда.

Будем говорить, что случайные процессы f_1 и f_2 являются *стохастически разделимыми* при $t \rightarrow \infty$, если коэффициенты корреляции $\rho_1^{(t)}(u, v)$ и $\rho_2^{(t)}(x, y)$ сходятся по вероятности к нулю при $t \rightarrow \infty$ при почти всех (u, v) и (x, y) .

Отметим, что приведенное выше определение предполагает, что длина "Гусеницы" τ является детерминированной. Более реалистическим с точки зрения

практики является выбор τ в зависимости от конкретного вида наблюдаемых реализаций процессов. Не представляет труда дать соответствующий вариант определения стохастической разделимости.

Приведем примеры, показывающие содержательность введенных выше понятий.

Пусть $f(s)$, $s \geq 0$ — некоторая неслучайная функция, интегрируемая с квадратом на любом конечном отрезке $0 \leq a < b < \infty$, а $\xi(s)$ — измеримый случайный процесс, интерпретируемый как "чистый шум". Пусть $E\xi(s) = 0$ при любом s и $R_\xi(x, y)$ — ковариационная функция процесса $\xi(s)$. Обозначим для краткости $r(x, u; y, v) = R_\xi(x + u, y + v)$.

Предположим, что функция f и случайный процесс $\xi(s)$ обладает следующими свойствами.

1. Условие А3.

Для любого $s \in [0, \infty)$ существует такое число $\delta = \delta(s) > 0$, что

$$\mathbf{P} \left(\int_0^T \xi^2(x + s) dx / T < \delta \right) \xrightarrow{T \rightarrow \infty} 0. \quad (6)$$

2. Условие А4.

Обозначим

$$\tilde{f}_T(x, s) = f(x + s) / \sqrt{\frac{1}{T} \int_0^T f^2(y + s) dy}.$$

Тогда для любых $u, v \in R$ при $T \rightarrow \infty$

$$F_{u,v}^T \stackrel{\text{def}}{=} \frac{1}{T^2} \iint_{[0,T]^2} dx dy \tilde{f}_T(x, u) \tilde{f}_T(y, u) r(x, v; y, v) \rightarrow 0. \quad (7)$$

Имеет место следующее утверждение.

Лемма 4.1 *Если $\tau \rightarrow \infty$, $t - \tau \rightarrow \infty$, а функция $f(s)$ и случайный процесс $\xi(s)$ удовлетворяют А3 и А4, то они стохастически разделяются при $t \rightarrow \infty$.*

Доказательство. Обозначим

$$A(T) = \left(\frac{1}{T} \int_0^T \tilde{f}_T(x, u) \xi(x + v) dx \right)^2.$$

Нам достаточно доказать, что для любых $u, v \in [0, \infty)$ и любого $\varepsilon > 0$

$$\mathbf{P} \left(A(T) > \frac{\varepsilon}{T} \int_0^T \xi^2(x + v) dx \right) \xrightarrow{T \rightarrow \infty} 0. \quad (8)$$

Ввиду условия (4.6) доказательство (4.8) сводится к проверке сходимости по вероятности к нулю величины $A(T)$. Поскольку

$$\mathbf{P}(A(T) > \varepsilon) \leq F_{u,v}^T / \varepsilon,$$

то требуемое немедленно следует из А4. \square

Прокомментируем утверждение предыдущей Леммы. Прежде всего отметим, что условия А3 и А4 выполняются, если процесс $\xi(s)$ является стационарным в широком смысле,

$$\frac{1}{T} \int_0^T \xi^2(s) ds \xrightarrow{P} R_\xi(0) > 0, \quad (9)$$

$$\frac{1}{T^2} \int_0^T dx \int_0^T dy |R_\xi(x-y)| \rightarrow 0$$

при $T \rightarrow \infty$, а функция f является ограниченной и удовлетворяет соотношению

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \int_0^T f^2(x+s) dx > 0$$

при любом $s \geq 0$.

В случае неограниченной или быстро убывающей функции f "наглядные" достаточные условия выполнения А3 и А4, естественно, изменятся. Скажем, если функция f является линейной, а предположение о стационарности $\xi(s)$ вместе с (4.9) остается в силе, то (4.7) превратится в

$$\frac{1}{T^2} \int_0^T dx \int_0^T dy (x+1)(y+1) |R_\xi(x-y)| = o(T^2).$$

Для вещественной экспоненты с показателем β аналогичное соотношение примет вид

$$\frac{1}{T^2} \int_0^T dx \int_0^T dy e^{\beta(x+y)} |R_\xi(x-y)| = o(T^{-1} e^{2\beta T}).$$

Тем самым в стационарном случае условия асимптотической разделимости "сигнала" и "шума" описываются, грубо говоря, в терминах скорости убывания ковариационной функции "шума". Ясно также, что можно рассматривать различные нестационарные варианты процесса $\xi(s)$, обладающие регулярным поведением величины $\int_0^T \xi^2(s) ds/T$ при больших T .

Отметим, что в лемме 4.1 рассматривается детерминированный вариант выбора τ . Переход к случайной длине "Гусеницы" не представляет большой сложности, если, например, потребовать выполнения неравенства $t/a \leq \tau(t, \omega) \leq t/b$ для некоторых a и b при любых ω и t .

Сделаем теперь несколько замечаний по поводу роли центрирования в стохастическом разделении "сигнала" $f(s)$ и "шума" $\xi(s)$. Естественно потребовать, чтобы процесс $\xi(s)$ обладал свойством *асимптотической стохастической инвариантности относительно центрирования*, то есть

$$\frac{1}{T} \int_0^T \xi(x+s) dx \xrightarrow{P} 0 \quad (10)$$

при $T \rightarrow \infty$ при любом $s \in [0, \infty)$. Отсюда сразу же следует, что утверждение леммы 4.1 остается в силе с заменой функции f на ее однократно центрированный вариант. При двукратном центрировании к (4.10) должен быть добавлен стохастический аналог (4.5).

Естественно, конкретные условия разделимости будут меняться. В частности, для линейной функции f при двукратном центрировании выполнение

закона больших чисел (4.10) и стохастического аналога (4.5) автоматически обеспечивает асимптотическое выделение f из "шума". С практической точки зрения это означает, что выделенное "Гусеницей" среднее будет мало отличаться от линейной функции f , а остальные члены разложения будут определяться шумовыми эффектами. Остановимся на этой ситуации подробнее.

4.4 Выделение линейного тренда

Пусть $f(s) = as + b$ и $\xi(s)$ — измеримый случайный процесс с нулевым средним и ковариационной функцией, равной $R(u, v)$. Предположим, что каждая реализация этого процесса интегрируема по мере Лебега на любом конечном промежутке. Для каждого $t > 0$ определим длину гусеницы $\tau = \tau(t)$, положим $\tilde{f}(s) = f(s) + \xi(s)$ и рассмотрим бицентральный вариант метода "Гусеница" для анализа функции \tilde{f} . Нетрудно видеть, что выделяемое среднее равно

$$\begin{aligned} \mathcal{E}_\tau \tilde{f}(s) + \mathcal{E}_{t-\tau} \tilde{f}(x) - \mathcal{E}_{\tau, t-\tau} \tilde{f} = \\ = a(x+s) + b + \mathcal{E}_\tau \xi(s) + \mathcal{E}_{t-\tau} \xi(x) - \mathcal{E}_{\tau, t-\tau} \xi, \end{aligned}$$

причем $x \in [0, \tau]$, $s \in [0, t - \tau]$,

$$\mathcal{E}_\tau \xi(s) = \frac{1}{\tau} \int_0^\tau \xi(x+s) dx,$$

$\mathcal{E}_{t-\tau} \xi$ определяется аналогично с заменой τ на $t - \tau$, а

$$\mathcal{E}_{\tau, t-\tau} \xi = \frac{1}{\tau(t-\tau)} \int_0^\tau dx \int_0^{t-\tau} \xi(x+s) ds.$$

Поскольку линейная функция полностью перешла в среднее, нам нет нужды проводить разложение Шмидта. Таким образом, используя терминологию раздела 2, нам нужно восстановить функцию \tilde{f} "по среднему" на промежутке $[0, t]$. Результат восстановления легко записывается, причем всегда можно считать, что $\tau \leq t/2$. Он имеет вид

$$f_t(u) = au + b + \eta_t(u),$$

где $u \in [0, t]$, а случайный процесс $\eta_t(u)$ (новый "шум") равен

$$\eta_t(u) = -\mathcal{E}_{\tau, t-\tau} \xi + \eta'_t(u),$$

где

$$\eta'_t(u) = \begin{cases} \int_0^u \zeta(u, s) ds/u, & \text{при } u < \tau, \\ \int_{u-\tau}^{t-\tau} \zeta(u, s) ds/(t-u), & \text{при } u > t - \tau, \\ \int_{u-\tau}^u \zeta(u, s) ds/\tau, & \text{иначе,} \end{cases}$$

а

$$\zeta(u, s) = \mathcal{E}_{t-\tau} \xi(u-s) + \mathcal{E}_\tau \xi(s)$$

(см. формулу (2.5)). В частности, $\eta_t(0) = \mathcal{E}_\tau \xi(0) + \mathcal{E}_{t-\tau} \xi(0) - \mathcal{E}_{\tau, t-\tau} \xi$, а $\eta_t(t) = \mathcal{E}_\tau \xi(t-\tau) + \mathcal{E}_{t-\tau} \xi(\tau) - \mathcal{E}_{\tau, t-\tau} \xi$.

Ясно, что $\mathbf{E}\eta_t(u) = 0$. Ковариационная функция $R_\eta(u, v)$ процесса $\eta_t(u)$ легко выписывается, поэтому не представляет принципиальных затруднений нахождение оценок наименьших квадратов для a и b (а для гауссовского процесса $\xi(u)$ и доверительных эллипсов для (a, b)). В общем случае, однако, вид $R_\eta(u, v)$ весьма громоздок. Приведем выражение для дисперсии $\mathbf{D}\eta_t(u) = R_\eta(u, u)$ в случае $\tau = t/2$. Нетрудно видеть, что при $u \leq \tau$

$$\begin{aligned} \mathbf{D}\eta_t(u) &= \frac{4}{\tau^2 u^2} \int_{[0, \tau]^2} dx dx_1 \int_{[0, u]^2} dy dy_1 r(x, y; x_1, y_1) \\ &- \frac{2}{\tau^3 u} \int_{[0, \tau]^3} dx dx_1 dy \int_{[0, u]} dy_1 r(x, y; x_1, y_1) \\ &+ \frac{1}{\tau^4} \int_{[0, \tau]^4} dx dx_1 dy dy_1 r(x, y; x_1, y_1). \end{aligned} \quad (11)$$

Аналогичным образом при $u \geq \tau$

$$\begin{aligned} \mathbf{D}\eta_t(u) &= \frac{4}{\tau^2(t-u)^2} \int_{[0, \tau]^2} dx dx_1 \int_{[u-\tau, \tau]^2} dy dy_1 r(x, y; x_1, y_1) \\ &- \frac{2}{\tau^3(t-u)} \int_{[0, \tau]^3} dx dx_1 dy \int_{[u-\tau, \tau]} dy_1 r(x, y; x_1, y_1) \\ &+ \frac{1}{\tau^4} \int_{[0, \tau]^4} dx dx_1 dy dy_1 r(x, y; x_1, y_1). \end{aligned}$$

Конечно, если τ и $t - \tau$ стремятся к бесконечности, а процесс $\xi(u)$ обладает некоторыми эргодическими свойствами (например, если он стационарен в широком смысле и его ковариационная функция достаточно быстро убывает), то дисперсия $\mathbf{D}\eta_t(u)$ будет стремиться к нулю при $t \rightarrow \infty$.

Для выяснения скорости сходимости процесса η_t к нулю упростим ситуацию, перейдя к дискретному случаю и предположив, что процесс $\xi(u)$ является "белым шумом" с дискретным временем. Тем самым мы имеем дело с некоррелированными случайными величинами ξ_0, \dots, ξ_{n-1} , причем $\mathbf{D}\xi_i = \sigma^2 > 0$. Будем считать, что n является нечетным и длина "Гусеницы" d имеет вид $d = (n+1)/2$. Ясно, что здесь достаточно рассматривать $u \leq d$.

Выражение для дисперсии $\mathbf{D}\eta_t(u)$, аналогичное (4.11), в дискретном случае также легко выписывается, при этом $R_\xi(x, y) = \sigma^2 \delta(x - y)$. Приведем окончательный результат, пренебрегая членами порядка $o(1/d)$. При $0 \leq u \leq d$

$$\mathbf{D}\eta_t(u) \approx \sigma^2 \frac{11d^2 - 7ud + 2u^2}{3d^3}, \quad (12)$$

при $u > d$ выражение (4.12) сохраняется с заменой u на $2d - u$. Таким образом, дисперсия по порядку уменьшается в d раз, правда, несколько неравномерно: в середине ряда (при $u \approx d$) примерно в $d/2$ раз, а в начале и конце (при $u \approx 0$ и $u \approx 2d$) примерно в $3d/11$ раз.

В целом приведенный метод аналогичен "скользящему суммированию", однако "Гусеница" позволяет выбирать длинные участки осреднения, не теряя начала и конца ряда. Что касается выбора длины "Гусеницы" $\tau = t/2$, то, как нетрудно видеть, этот выбор является оптимальным по крайней мере по порядку.

* * *

Подведем некоторые итоги. Полученные результаты показывают принципиальную разницу в использовании метода "Гусеница" для анализа коротких и длинных временных рядов. В первом случае очень велика роль длины "Гусеницы" (и длины самого ряда), условия разделения слагаемых весьма ограничительны, существенную роль играют периодические составляющие ряда. Класс "естественных" функций, выделяемых методом, невелик.

Другая картина возникает при работе с длинными рядами, где "Гусеница" способна выделить сигнал из шума достаточно общего вида и одновременно разложить его на простые и интерпретируемые составляющие. При этом выбор длины "Гусеницы" существенно облегчается, что не исключает роли "удачного попадания" при анализе, скажем, зашумленных периодических функций.

Следует помнить, однако, что отнесение некоторого члена разложения Гильберта-Шмидта к "сигналу" или "шуму" происходит отнюдь не автоматически и зависит от опыта и искусства статистика. Кроме того, работа с большими длинами "Гусеницы" на практике требует нахождения собственных чисел и собственных векторов матриц больших размерностей, что порождает чисто вычислительные трудности.

5 Мультипликативная "Гусеница"

В этом разделе мы рассмотрим мультипликативный вариант метода "Гусеница". Хотя этот пример не имеет большого практического значения, он показывает, что выбор функции $\theta(x, s)$, отличной от $x + s$, также может привести к содержательным результатам.

Здесь нам будет удобно несколько изменить область определения основных функций. Подлежащая разложению функция f (как и раньше, она считается непрерывной) будет задана на отрезке $[-t, t]$, а множества D_1 и D_2 приобретут вид $D_1 = [\tau, \tau]$, $D_2 = [-t/\tau, t/\tau]$, причем $\tau \in (0, t)$. Задав на борелевских подмножествах D_1 и D_2 меры μ_1 и μ_2 , определим $\theta(x, s) = xs$, $x \in D_1$, $s \in D_2$. Тем самым в нецентральном случае

$$g(x, s) = f(xs).$$

Докажем сначала, что в случае, когда μ_1 и μ_2 являются мерами Лебега, а функция f — непрерывна, всегда можно считать, что каждая из функций ψ_n и φ_n является либо четной, либо нечетной. Прежде всего заметим, что при этих предположениях любая четная функция будет ортогональна любой нечетной (конечно, если они суммируемы с квадратом). Проведем доказательство для собственных функций ψ_n .

Действительно, пусть h — непрерывная собственная функция оператора $\mathbf{G}\mathbf{G}^*$. Это значит, что

$$\int_{-\tau}^{\tau} h(y) dy \int_{-t/\tau}^{t/\tau} f(xs) f(ys) ds = \lambda h(x).$$

Непосредственно видно, что функция $h_1(x) = h(-x)$ также является собственной и соответствует тому же самому собственному числу λ . Пусть функция h не

является ни четной, ни нечетной. Тогда λ имеет кратность, большую единицы, причем функции $h(x) + h(-x)$ и $h(x) - h(-x)$ также являются собственными, соответствуют собственному числу λ , первая из них является четной, а вторая — нечетной. Следовательно, любая собственная функция кратности один является или четной, или нечетной, а для любого собственного подпространства большей размерности можно выбрать соответствующий базис.

Остановимся кратко на условиях конечномерности ненулевого собственного подпространства оператора $\mathbf{R}_\tau = \mathbf{G}_\tau \mathbf{G}_\tau^*$ и разделимости слагаемых в сумме.

Непосредственно проверяется, что при $f = a|y|^\beta$, $\beta \geq 0$, собственное подпространство оператора \mathbf{R}_τ будет одномерным при любых μ_1 и μ_2 (в случае целого β знак модуля можно убрать). Аналогичным образом конечная линейная комбинация подобных функций с неотрицательными (не обязательно целыми) показателями будет порождать конечномерное собственное подпространство. Все дальнейшие рассмотрения посвящены только варианту с мерой Лебега.

Пусть $f \in \mathbf{C}^{(m)}([-t, t])$ и

$$f(xs) = \sum_{k=0}^{m-1} \eta_k(x) \gamma_k(s).$$

Повторяя выкладки ([5]), мы получим, что вместо (3.2) f удовлетворяет дифференциальному уравнению

$$\sum_{j=0}^m b_j y^j f^{(j)}(y) = 0. \quad (1)$$

Пусть не все коэффициенты (5.1) равны нулю. Не ставя задачу отыскания общего решения этого уравнения, покажем, что если ненулевая функция f разлагается в ряд Тейлора в некоторой окрестности нуля и удовлетворяет (5.1), то она является некоторым полиномом вида $f(y) = \sum_{j=1}^m c_j y^{n_j}$, где $n_j \geq 0$, целые.

Действительно, пусть $f(y) = \sum_{k=0}^{\infty} c_k y^k$. Подставляя это равенство в (5.1) и приравнивая коэффициенты при степенях y к нулю, получим, что

$$c_i \sum_{k=0}^i b_k k! \binom{k}{i} = 0$$

при $0 \leq i < m$ и

$$c_i \sum_{k=0}^m b_k k! \binom{k}{i} = 0$$

при $i \geq m$. Отсюда следует, что самое большое m чисел c_i отлично от нуля.

Перейдем теперь к разделению слагаемых в сумме. Условия разделимости (как уже говорилось, мы ограничиваемся нецентральным ковариационным оператором, непрерывными функциями и мерой Лебега) сводятся к равенствам

$$\int_{-\tau}^{\tau} f_1(xu) f_2(xv) dx = 0 \quad (2)$$

при любых $u, v \in [-t/\tau, t/\tau]$ и

$$\int_{-t/\tau}^{t/\tau} f_1(xs) f_2(ys) ds = 0$$

при любых $x, y \in [-\tau, \tau]$.

Прежде всего заметим, что если одна из функций f_1 или f_2 четная, а другая нечетная, то они всегда будут разделяться. Оказывается, эти условия являются не только достаточными, но и необходимыми.

Разложим обе функции на четную и нечетную составляющую, $f_1 = f_1^+ + f_1^-$, $f_2 = f_2^+ + f_2^-$. Тогда (5.2) превратится в

$$\int_0^\tau F(x; u, v) dx = 0,$$

где четная функция $F(\cdot; u, v)$ имеет вид

$$F(x; u, v) = f_1^+(xu)f_2^+(xv) + f_1^-(xu)f_2^-(xv).$$

Заменив u на $-u$, получим, что

$$\int_0^\tau f_1^+(xu)f_2^+(xv) dx = 0.$$

Сделав замену переменных $xu = z$ и обозначив $\alpha = v/u$, будем иметь

$$\int_0^{vu} f_1^+(z)f_2^+(\alpha z) dz = 0.$$

Ввиду произвольности u при фиксированном α $f_1^+(z)f_2^+(\alpha z) = 0$. В свою очередь, из-за произвольности α получаем, что либо $f_1^+ \equiv 0$, либо $f_2^+ \equiv 0$. Следовательно, одна из функций является четной, а другая — нечетной.

Указатель литературы

- [1] Рид М., Саймон Б., (1977), *Методы современной математической физики, т.1 (Функциональный анализ)*, М.: — Мир.
- [2] Лоэв М., (1962), *Теория вероятностей*, М.: — ИЛ.
- [3] Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д., (1989), *Прикладная Статистика. Классификация и снижение размерности*, М.: — Финансы и Статистика.
- [4] Форсайт Дж., Малькольм М., Моулер К., (1980), *Машинные методы математических вычислений*, М.: — Мир.
- [5] Buchstaber V.M., (1994), *Time Series Analysis and Grassmannians*, Amer. Math. Soc. Transl., (2), **162**, 1- 17.
- [6] Vautard, R., Ghil, M., (1989), Singular-spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D*, **35**, 395-424.
- [7] Камке Э., (1971), *Справочник по обыкновенным дифференциальным уравнениям*, М.: — Наука.

V.2 Метода “Гусеница” для прогнозирования временных рядов: элементы операторной теории.

Д.Л. Данилов

1 Схема “гусеничной” экстраполяции в общем случае.

Предлагается описание методов экстраполяции функций “гусеничного” типа на языке функционального анализа, разработанном в применении к методу “Гусеница” в разделе V.1. Для упрощения выкладок изложение большей части результатов ведется для случая непрерывной “гусеницы” без центрирования. В качестве непосредственного приложения рассматривается аддитивная гусеница.

Рассмотрим разложение Гильберта-Шмидта некоторой функции $g \in L^2([0, T - \tau] \times [0, \tau])$:

$$g(x, s) = \sum_n \sqrt{\lambda_n} \phi_n(s) \psi_n(x), \text{ при } s \in [0, \tau], x \in [0, T - \tau]. \quad (1)$$

Здесь ϕ_n — собственные функции ковариационного оператора G^*G , имеющего ядро $k(u, v) = \int g(x, u)g(x, v)dx$, и

$$\psi_n(x) = \frac{1}{\sqrt{\lambda_n}} \int_0^\tau g(x, r) \phi_n(r) dr. \quad (2)$$

С учетом (2) разложение (1) можно представить в виде

$$g(x, s) = \sum_n \phi_n(s) \int_0^\tau g(x, r) \phi_n(r) dr. \quad (3)$$

Предположим, что мы хотим восстановить значения ядра g на $[\tau - \delta, \tau] \times [0, T - \tau]$ по имеющейся системе функций ϕ_n и значениям g на $[0, \tau - \delta] \times [0, T - \tau]$. Преобразуем выражение в правой части (3). Разобьем интегралы следующим образом:

$$g(x, s) = \sum_n \phi_n(s) \int_0^{\tau - \delta} g(x, r) \phi_n(r) dr + \sum_n \phi_n(s) \int_{\tau - \delta}^\tau g(x, r) \phi_n(r) dr. \quad (4)$$

Перенося вторую сумму в левую часть, получим

$$g(x, s) - \sum_n \phi_n(s) \int_{\tau - \delta}^\tau g(x, r) \phi_n(r) dr = \sum_n \phi_n(s) \int_0^{\tau - \delta} g(x, r) \phi_n(r) dr. \quad (5)$$

Ограничим область изменения аргументов (x, s) прямоугольником $[0, T - \tau] \times [\tau - \delta, \tau]$. Тогда выражение в левой части (5) представляет собой результат

действия на функцию g интегрального оператора $E - \bar{P}$, где \bar{P} имеет ядро $\bar{p}(s, r) = \sum_n \phi_n(s)\phi_n(r)$, и действует по формуле

$$[\bar{P}g](x, s) = \sum_n \phi_n(s) \int_{\tau-\delta}^{\tau} g(x, r)\phi_n(r)dr. \quad (6)$$

(При фиксированном x оператор \bar{P} есть интегральный оператор на $L^2[\tau - \delta, \tau]$.) Выясним условия существования $(E - \bar{P})^{-1}$. Запишем ряд Неймана этого оператора:

$$(E - \bar{P})^{-1} = E + \sum_{k=1}^{\infty} \bar{P}^k. \quad (7)$$

Критерием его сходимости является условие: $r_{\sigma}(\bar{P}) < 1$. (Здесь $r_{\sigma}(\bar{P})$ - спектральный радиус \bar{P} .) По определению

$$r_{\sigma}(\bar{P}) = \sup_{\|h\|=1} |(\bar{P}h, h)| = \sum_n \left(\int_{\tau-\delta}^{\tau} h(r)\phi_n(r)dr \right)^2. \quad (8)$$

Однако последнее выражение может быть переписано как

$$\sum_n \left(\int_0^{\tau} \hat{h}(r)\phi_n(r)dr \right)^2,$$

где $\hat{h} \in L^2[0, \tau]$: $\hat{h}|_{[\tau-\delta, \tau]} = h$ и $\hat{h}|_{[0, \tau-\delta]} = 0$. Поскольку интегралы $\int_0^{\tau} \hat{h}(r)\phi_n(r)dr$ представляют собой коэффициенты Фурье функции \hat{h} по ортонормальной системе функций $\{\phi_n\}$, то $r_{\sigma}(\bar{P}) \leq \int_0^{\tau} \hat{h}^2(r)dr = 1$. Очевидно, сходимость имеет место при $r_{\sigma}(\bar{P}) < 1$.

Выясним смысл условия $r_{\sigma}(\bar{P}) = 1$. Выполнение этого равенства означает существование такой функции h_* , что $\sum_n (\int_{\tau-\delta}^{\tau} h_*(r)\phi_n(r)dr)^2 = 1$, и, следовательно, функции $\hat{h}_*(s) = \mathbb{I}_{[\tau-\delta, \tau]}(s)h_*(s)$, ортогональной ядру ковариационного оператора G^*G . Последнее означает существование такой функции из $L^2[0, \tau]$, которая полностью раскладывается по системе векторов $\{\phi_n\}$ и при этом обращается в 0 на отрезке $[0, \tau - \delta]$, что немедленно влечет линейную зависимость системы функций $\{\phi_n\}$ на отрезке $[0, \tau - \delta]$, т.е. выполнение тождества $\sum_n \alpha_n \phi_n(s) = 0$ на $[0, \tau - \delta]$ для некоторых α_n , не равных нулю одновременно. Таким образом можно считать доказанной следующую лемму:

Лемма 1.1 Следующие утверждения эквивалентны:

1. $r_{\sigma}(\bar{P}) = 1$.
2. Существует такая функция $h_* \in L^2([0, \tau])$, ортогональная ядру ковариационного оператора G^*G , что $h_*(s) = \mathbb{I}_{[\tau-\delta, \tau]}(s)h(s)$, $h \in L^2([\tau - \delta, \tau])$.
3. Система собственных функций ковариационного оператора G^*G линейно зависит на отрезке $[0, \tau - \delta]$.

Рассмотрим несколько примеров.

Пример 1.1 Пусть $g(x, s) = f(x+s)$, $f(t) = \cos(\omega t)$. В этом случае собственное подпространство G^*G двумерно и образовано собственными функциями \cos , \sin . Указанная пара функций линейно независима на любом конечном интервале. Поскольку утверждение 3 Леммы 1.1 не выполняется, то $r_\sigma(\bar{P}) < 1$.

Пример 1.2 Пусть $g(x, s) = f(x+s)$, и $f(t) = \mathbb{1}_{[\tau-\rho, \tau]}(s)$, где $\rho < \delta$. В этом случае системы собственных функций и главных компонент порождаемых g , очевидно, обладают свойством:

$$g(x, s) = \mathbb{1}_{[\tau-\rho, \tau]}(x+s) = \sum_n \sqrt{\lambda_n} \phi_n(s) \psi_n(x) \text{ при } s \in [0, \tau], x \in [0, T-\tau].$$

Пусть теперь $x \in [0, T-\tau]$ и $s \in [0, \tau-\delta]$. Тогда $f(x+s) = 0$ и, следовательно,

$$\sum_n \phi_n(s) \alpha_n = 0 \text{ при } s \in [0, \tau-\delta].$$

Пример 1.3 Пусть $g(x, s) = f(x+s)$, и функция $f(t)$ обладает следующим свойством: размерность собственного подпространства ковариационного оператора конечна, не зависит от τ и, для определенности, равна m . В случае, когда f обладает еще и некоторыми свойствами гладкости, см. [1], такой класс функций описывается формулой $f(t) = \sum_{k=1}^m a_k(t) e^{\lambda_k t} \sin(\omega_k t + \varphi_k)$. Покажем, что в этом случае всегда $r_\sigma(\bar{P}) < 1$.

Действительно, рассмотрим порождаемые f поля $g(x, y) = f(x+y)$ при длинах гусеницы τ и $\tau-\delta$. Пусть система функций ϕ_n порождена g при длине гусеницы τ , и поэтому разложение (1) имеет место при $(x, s) \in [0, T-\tau] \times [0, \tau]$. Однако, это же разложение справедливо и на любом подмножестве $[0, T-\tau] \times [0, \tau]$, в том числе и на $[0, T-\tau] \times [0, \tau-\delta]$. Функции $\phi_n|_{[0, \tau-\delta]}$ уже не образуют ортогональную систему. Однако на $[0, T-\tau] \times [0, \tau-\delta]$ та же функция g имеет и биортогональное разложение. Пусть это разложение имеет вид:

$$g(x, s) = \sum_{n=1}^m \sqrt{\hat{\lambda}_n} \hat{\phi}_n(s) \hat{\psi}_n(x).$$

Пусть система функций $\phi_n|_{[0, \tau-\delta]}$ линейно зависима. Выделим из нее линейно независимую подсистему. Обозначим эти функции через $\{\phi_n^*\}_{n=1}^M$ и $M < m$. Выразив $\phi_n|_{[0, \tau-\delta]}$ через ϕ_n^* , получим:

$$g(x, s) = \sum_{n=1}^M \sqrt{\lambda_n^*} \phi_n^*(s) \psi_n^*(x).$$

Приравняем эти разложения:

$$\sum_{n=1}^m \sqrt{\hat{\lambda}_n} \hat{\phi}_n(s) \hat{\psi}_n(x) = \sum_{n=1}^M \sqrt{\lambda_n^*} \phi_n^*(s) \psi_n^*(x). \quad (9)$$

Умножая обе части (9) на $\hat{\phi}_k$ и интегрируя, получаем, с учетом ортогональности $\hat{\phi}_n$, что

$$\hat{\psi}_k(x) = \sum_{n=1}^M \sqrt{\frac{\lambda_n^*}{\hat{\lambda}_k}} \psi_n^*(x) \int \hat{\phi}_k(s) \phi_n^*(s) ds \text{ при } k = 1, \dots, m.$$

Последнее немедленно влечет линейную зависимость $\hat{\psi}_n$, что невозможно ввиду их ортогональности. Поскольку условие 3 Леммы 1.1 не выполняется, то $r_\sigma(\bar{P}) < 1$.

Далее будем предполагать, что условие $r_\sigma(\bar{P}) < 1$ выполнено. В этом случае можно записать:

$$g(x, s) = [(E - \bar{P})^{-1} h](x, s), \quad (10)$$

где

$$h(x, s) = \sum_n \phi_n(s) \int_0^{\tau-\delta} g(x, r) \phi_n(r) dr,$$

а оператор $(E - \bar{P})^{-1}$ имеет ядро $\sum_{n,m} \phi_n(s) b_{nm} \phi_m(r)$, где $B = (b_{nm}) = (E - A)^{-1}$,

$A = (a_{nm})$, $a_{nm} = \int_{\tau-\delta}^{\tau} \phi_n(r) \phi_m(r) dr$. Фактически можно считать, что $h = \hat{P}g$,

где $\hat{P}: \hat{P} + \bar{P} = P$, а P — проекционный оператор на $L^2[0, \tau]$ с ядром $p(s, r) = \sum_n \phi_n(s) \phi_n(r)$.

С учетом вышесказанного, формулу (10) можно переписать в виде

$$g(x, s) = \sum_{n,m} \phi_n(s) b_{nm} \int_0^{\tau-\delta} \phi_m(t) \sum_k \phi_k(t) \int_0^{\tau-\delta} g(x, r) \phi_k(r) dr dt. \quad (11)$$

Выражение (11) является основной формулой, используемой для прогнозирования (экстраполяции) ядра g . Исследуем некоторые свойства (11).

Лемма 1.2 Выражение (11) при $(x, s) \in [0, T - \tau] \times [\tau - \delta, \tau]$ однозначно определяет ядро g .

Доказательство

Из существования обратного оператора $(E - \bar{P})^{-1}$ следует, что уравнение (5) имеет единственное решение при любой заданной правой части. Пусть правая часть (5) задана значениями g на $[0, T - \tau] \times [0, \tau - \delta]$. В этом случае решение (5) есть некоторая функция \hat{g} на $[0, T - \tau] \times [\tau - \delta, \tau]$, определяемая по (10). Однако, одно решение (5) уже имеется: это само ядро g на соответствующем интервале. Поэтому $\hat{g} = g$. ■

Использование формулы (10) для определений значений g при $(x, s) \in [0, T - \tau] \times [\tau - \delta, \tau]$ предоставляет некую альтернативную формуле (1) возможность для восстановления g . На практике, сравнение этих двух вариантов восстановления g говорит о том, насколько большая информация о исходном поле была потеряна при использовании "усеченной" системы собственных векторов вместо "полней". Сильное расхождение восстановления и "псевдовосстановления", полученного с использованием (11), говорит о малой надежности получаемых прогнозов.

Рассмотрим следующую общую схему экстраполяции ядра $g(x, s)$. Пусть значения функции g известны на прямоугольнике $[0, T - \tau + \delta] \times [0, \tau]$, разложение (1) имеет место на $[0, T - \tau] \times [0, \tau]$, а на прямоугольнике $[0, T - \tau + \delta] \times [0, \tau]$ справедливо следующее разложение:

$$g(x, s) = \sum_n \sqrt{\lambda_n} \phi_n(s) \bar{\psi}_n(x), \quad (12)$$

где

$$\bar{\psi}_n(x) = \int_0^\tau g(x, r) \phi_n(r) dr, \quad (13)$$

и $\bar{\psi}_n$ — вообще говоря, не ортогональная и не нормированная система функций. Смысл условия прост: продолжение ядра на δ не влечет расширения линейной оболочки системы ϕ_n .

Лемма 1.3 *Если разложение (12) справедливо, то выражение (10) восстанавливает значения ядра g на $[0, T - \tau + \delta] \times [\tau - \delta, \tau]$ по его значениям на $[0, T - \tau] \times [0, \tau]$.*

Пример 1.4 Как и в примере (1.3), пусть $g(x, s) = f(x + s)$. Рассмотрим функции f , порождающие конечномерные (m -мерные) собственные подпространства. Докажем, что разложение (12) имеет место. Рассмотрим биортогональное разложение g на множестве $[0, T - \tau + \delta] \times [0, \tau]$:

$$g(x, s) = \sum_{n=1}^m \sqrt{\hat{\lambda}_n} \hat{\phi}_n(s) \hat{\psi}_n(x), \quad (14)$$

и разложение (1) на $[0, T - \tau] \times [0, \tau]$. Приравняем правые части разложений (12) и (14) на $[0, T - \tau + \delta] \times [0, \tau]$. Домножая на ϕ_k и интегрируя, получаем, с учетом ортогональности, что

$$\bar{\psi}_k(x) = \sum_{n=1}^m \sqrt{\hat{\lambda}_n} \hat{\psi}_n(x) \int_0^{T-\tau+\delta} \hat{\phi}_n(s) \phi_k(s) ds.$$

В этом случае справедливо

$$\begin{aligned} \sum_n \sqrt{\lambda_n} \phi_n(s) \bar{\psi}_n(x) &= \sum_{k=1}^m \phi_k(s) \sum_{n=1}^m \sqrt{\lambda_n} \hat{\psi}_n(x) \int_0^{T-\tau+\delta} \hat{\phi}_n(s) \phi_k(s) ds = \\ &= \sum_{n=1}^m \sqrt{\lambda_n} \hat{\psi}_n(x) \sum_{k=1}^m \phi_k(s) \int_0^{T-\tau+\delta} \hat{\phi}_n(s) \phi_k(s) ds. \end{aligned}$$

Последнее выражение совпадает с $g(x, s)$ тогда и только тогда, когда

$$\hat{\phi}_n(s) = \sum_{k=1}^m \phi_k(s) \int_0^{T-\tau+\delta} \hat{\phi}_n(s) \phi_k(s) ds, \quad (15)$$

т.е. базисы ϕ_n и $\hat{\phi}_n$ эквивалентны. Докажем, что это так. Легко установить, что всегда

$$\phi_n(s) = \sum_{k=1}^m \hat{\phi}_k(s) \int_0^{T-\tau+\delta} \hat{\phi}_k(s) \phi_n(s) ds \quad (16)$$

(следствие равенства разложений (1) и (14) на $[0, T - \tau] \times [0, \tau]$). Пусть теперь (15) не выполняется. Это означает, что матрица коэффициентов

$$\Gamma = (\gamma_{nk}), \text{ где } \gamma_{nk} = \int_0^{T-\tau+\delta} \hat{\phi}_n(s) \phi_k(s) ds,$$

вырождена. Однако, это равносильно тому, что некоторая нетривиальная линейная комбинация строк (столбцов) этой матрицы обращается в ноль, т.е. существует линейная комбинация $\omega = \sum_{k=1}^m c_k \phi_k$, где $\sum_{k=1}^m c_k^2 > 0$, ортогональная $\hat{\phi}_n$ при любом n . Но в этом случае ω ортогональна $g(x, s)$ на $[0, T - \tau + \delta] \times [0, \tau]$, а значит и на $[0, T - \tau] \times [0, \tau]$, что невозможно по построению базиса ϕ_n .

2 Прогнозирование слагаемого в сумме.

Напомним несколько важных понятий из основной схемы метода "Гусеница". Пусть $T > 0$ и $f : [0, T] \rightarrow \mathbb{R}$. Пусть имеется измеримое отображение $\theta : [0, T - \tau] \times [0, \tau] \rightarrow [0, T]$. Тогда функция g может быть определена как:

$$g(x, s) = f(\theta(x, s)). \quad (17)$$

Алгоритмы "гусеничного" типа основаны на разложении функции (17) по биортогональной системе функций порожденной операторами G^*G и GG^* , с последующим отбором информативных слагаемых. Запишем результат восстановления функции g при помощи формулы (10). Можно считать, что

$$g(x, s) = \sum_{n,m} \phi_n(s) b_{nm} \int_0^{\tau-\delta} \phi_m(t) \sum_k \phi_k(t) \int_0^{\tau-\delta} g(x, r) \phi_k(r) dr dt. \quad (18)$$

Рассмотрим замкнутое линейное подпространство $L_\theta^2([0, T - \tau] \times [0, \tau])$, зависящих от (x, s) через θ , и след этого подпространства $L_{(\theta,\delta)}^2$ на множестве $[0, T - \tau] \times [\tau - \delta, \tau]$. Проекция n -го члена ряда (18) на $L_{(\theta,\delta)}^2$ имеет вид $f_n(\theta(x, s))$, соответственно, $f(\theta(x, s)) = \sum_n f_n(\theta(x, s))$. Функцию f_n логично назвать псевдовосстановлением f по n -й собственной функции. Смысл операции псевдовосстановления заключается в следующем: перейти от восстановленного ядра $g(x, s)$ к восстановлению порождающей его функции f .

Рассмотрим важный частный случай аддитивной гусеницы $g(x, s) = f(x + s)$. В этом случае формула (18) приобретает вид:

$$f(x + s) = \sum_{n,m} \phi_n(s) b_{nm} \int_0^{\tau-\delta} \phi_m(t) \sum_k \phi_k(t) dt \int_0^{\tau-\delta} f(x + r) \phi_k(r) dr. \quad (19)$$

Можно заметить, что оператор, определяемый формулой (19), для каждого x действует на функцию f на $[x, x + \tau - \delta]$ и отображает ее на $[x + \delta, x + \tau]$, действуя как оператор сдвига на δ .

Обратимся теперь к вопросу об однозначной экстраполяции слагаемого в сумме. Будем считать, что у нас имеется функция $f = f^{(1)} + f^{(2)}$ (и $g = g^{(1)} + g^{(2)}$ соответственно), и мы хотим выяснить, при каких условиях мы можем

экстраполировать $f^{(1)}$ по наблюдаемым значениям f с помощью приведенных выше формул.

Предположим, что стандартные условия разделимости для слагаемых $f^{(1)}$ и $f^{(2)}$ выполнены. Будем обозначать собственные функции ковариационного оператора, соответствующие первому или второму слагаемому, верхним индексом 1 или 2. Выражение (19) при этом приобретает вид:

$$\begin{aligned} g^{(1)}(x, s) = & \sum_{n,m} \phi_n^{(1)}(s) b_{nm} \int_0^{\tau-\delta} \phi_m^{(1)}(t) \sum_k \phi_k^{(1)}(t) dt \int_0^{\tau-\delta} g^{(1)}(x, r) \phi_k^{(1)}(r) dr + \\ & + \sum_{n,m} \phi_n^{(1)}(s) b_{nm} \int_0^{\tau-\delta} \phi_m^{(1)}(t) \sum_k \phi_k^{(1)}(t) dt \int_0^{\tau-\delta} g^{(2)}(x, r) \phi_k^{(1)}(r) dr. \end{aligned} \quad (20)$$

Поэтому достаточным условием продолжимости $f^{(1)}$ в составе суммы служит условие:

$$\int_0^{\tau-\delta} p^{(1)}(s, r) g^{(2)}(t, r) dr = 0 \text{ при п.в. } (s, t) \in [\tau - \delta, \tau] \times [0, T - \tau + \delta]. \quad (21)$$

Смысл условия (21) прост: при п.в. (s, t) функция $g^{(2)}(t, \cdot)$ принадлежит ядру оператора $P^{(1)}$. Условие (21) гарантирует не только однозначность экстраполяции $g^{(1)}$ (когда $t \in [0, T - \tau]$), но и однозначное получение "псевдовосстановления" при $t \in [T - \tau, T - \tau + \delta]$.

Это условие, однако, представляется труднопроверяемым. Поэтому необходимо установить осмыслинный критерий его выполнения. Справедлива следующая лемма.

Лемма 2.1 Если система функций $\{\phi_n\}$ линейно независима на отрезке $[0, \tau - \delta]$, то условие (21) эквивалентно следующему условию:

$$\int_0^{\tau-\delta} g^{(1)}(s, r) g^{(2)}(t, r) dr = 0. \quad (22)$$

В случае "аддитивной гусеницы" последнее условие приобретает вид:

$$\int_0^{\tau-\delta} f^{(1)}(s + r) f^{(2)}(t + r) dr = 0. \quad (23)$$

Приведем несколько простых примеров, показывающих содержательность приведенного критерия.

Пример 2.1 Пусть функция $f^{(1)}$ — константа, а функция $f^{(2)}$ такова, что стандартные условия разделимости (см. раздел V.1) справедливы, т.е. $f^{(2)}$ имеет период τ_1 , T и τ кратны τ_1 и

$$\int_0^{\tau_1} f^{(2)}(y) dy = 0.$$

В этом случае условие (23) имеет вид:

$$\int_0^{\tau-\delta} f^{(2)}(y) dy = 0,$$

что выполняется при δ кратном τ_1 .

Пример 2.2 Пусть $f^{(1)} = \exp(\beta y)$ и $f^{(2)}$ такова, что стандартные условия разделимости справедливы, т.е. $\exp(\beta y)f^{(2)}(y)$ имеет период τ_1 , T и τ кратны τ_1 и

$$\int_0^{\tau_1} \exp(\beta y) f^{(2)}(y) dy = 0.$$

В этом случае условие (23) имеет вид:

$$\int_0^{\tau-\delta} \exp(\beta y) f^{(2)}(y) dy = 0,$$

что выполняется при δ кратном τ_1 .

Пример 2.3 Пусть $f^{(1)}(y) = \exp(i\omega y)$, и функция $f^{(2)}$ такова, что выполняются условия: $f^{(2)}$ имеет период $\tau_1 = 2\pi/\omega$, t и τ кратны τ_1 и

$$\int_0^{\tau_1} f^{(2)}(y) \exp(i\omega y) dy = 0.$$

В этом случае условие (23) приобретет вид:

$$\int_0^{\tau-\delta} f^{(2)}(y) \exp(i\omega y) dy = 0.$$

Это условие снова выполняется при δ кратном τ_1 .

Если условие (23) не выполняется точно, оно тем не менее может быть выполнено в каком-то смысле приближенно. Это соответствует случаю "зашумленного" ряда.

Заметим, что условие (23) может быть проинтерпретировано как равенство нулю нецентрального коэффициента корреляции ρ между функциями $g^{(1)}(x, \cdot)$ и $g^{(2)}(y, \cdot)$, т.е.

$$\rho[g^{(1)}, g^{(2)}](x, y) = \frac{< g^{(1)}(x, \cdot), g^{(2)}(y, \cdot) >}{\|g^{(1)}(x, \cdot)\| \|g^{(2)}(y, \cdot)\|}.$$

(Подробнее об условиях такого вида см. раздел V.1.) Эта интерпретация может быть положена в основу определения приближенной прогнозируемости функции $g^{(1)}$ в сумме $g^{(1)} + g^{(2)}$. Будем говорить, что $g^{(1)}$ является ε -прогнозируемой в сумме $g^{(1)} + g^{(2)}$, если при п.в. (x, y) справедливо $|\rho[g^{(1)}, g^{(2)}](x, y)| \leq \varepsilon$. Введенное таким образом понятие прогнозируемости будет устойчиво относительно малых отклонений в функциях $g^{(1)}, g^{(2)}$.

Дальнейшим развитием понятия приближенной прогнозируемости является асимптотическая прогнозируемость функции в сумме. Действительно, пусть $f^{(1)}, f^{(2)} : \mathbb{R}^+ \rightarrow \mathbb{R}$ измеримые функции, и их сужения $f_T^{(1)}, f_T^{(2)}$ на $[0, T]$ суммируемы относительно лебеговских мер на $[0, T]$. Пусть длина гусеницы $\tau = \tau(T)$, и $g_T^{(i)}(x, s) = f_T^{(i)}(\theta(x, s))$. Пусть $f = f^{(1)} + f^{(2)}$ и функции $f^{(1)}, f^{(2)}$ асимптотически разделяются при $T \rightarrow \infty$. Пусть также $\delta = \delta(T)$, $0 < \delta < \tau$, и оператор $\bar{P}_T^{(1)}$ обратим для любого T . Будем говорить, что $f^{(1)}$ асимптотически прогнозируема в сумме, если $|\rho[g^{(1)}, g^{(2)}](x, y)| \rightarrow 0$ при $T \rightarrow \infty$.

Пример 2.4 Пусть $f^{(1)}$ - константа, $f^{(2)}$ - непрерывна, и стандартные условия асимптотической разделимости справедливы, т.е.

$$\frac{1}{T} \int_0^T f^{(2)}(s+y) dy \rightarrow 0$$

при $T \rightarrow \infty$, и

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \int_0^T (f^{(2)}(s+y))^2 dy > 0,$$

при любом $s \geq 0$ и $\tau \rightarrow \infty$, $T - \tau \rightarrow \infty$. Тогда условием асимптотической прогнозируемости является:

$$\frac{1}{\tau - \delta} \int_0^{\tau - \delta} f^{(2)}(s+y) dy \rightarrow 0,$$

что выполняется при выборе δ таком, что $\tau - \delta \rightarrow \infty$. Заметим, что здесь, вообще говоря, возможно $\delta \rightarrow \infty$.

С практической точки зрения представляет интерес случай, когда $f^{(2)}$ является случайным процессом. В этом случае будем говорить, что функция $f^{(1)}$ асимптотически стохастически прогнозируема, если $|\rho[g^{(1)}, g^{(2)}](x, y)| \rightarrow 0$ по вероятности при $T \rightarrow \infty$.

А именно: пусть $f^{(2)}(s) = \xi(s)$ есть измеримый случайный процесс с нулевым средним $E\xi(s) = 0$ и известной ковариационной функцией процесса $R_\xi(x, y)$. Пусть выполняются стандартные условия асимптотической стохастической разделимости:

CSD1: для любого $s \geq 0$, существует $\gamma = \gamma(s) > 0$:

$$P \left\{ \frac{1}{T} \int_0^T \xi^2(x+s) dx < \gamma \right\} \rightarrow 0, \text{ при } T \rightarrow \infty.$$

CSD2: для любых вещественных u, v выполняется

$$\frac{1}{T^2} \int_0^T dx \int_0^T dy \hat{f}_T(x, u) \hat{f}_T(y, u) R_\xi(x+v, y+v) \rightarrow 0 \text{ при } T \rightarrow \infty,$$

$$\text{где } \hat{f}_T(x, s) = \frac{f(x + s)}{\sqrt{\frac{1}{T} \int_0^T f^2(y + s) dy}},$$

и, кроме этого, $\tau \rightarrow \infty$, и $T - \tau \rightarrow \infty$. Тогда для асимптотической прогнозируемости f в сумме $f + \xi$ достаточно выполнения условия $\tau - \delta \rightarrow \infty$. Опять-таки заметим, что последнее условие оставляет возможность для роста δ с увеличением T . Это позволяет надеяться, что с ростом наблюдаемого отрезка исследуемого случайного процесса возможно предсказывать дальнейшее поведение его тренда на все больший промежуток времени вперед.

Литература

- [1] Buchstaber, V.M., (1994), Time Series Analysis and Grassmannians. *Amer. Math. Soc. Transl.* (2), **162**.

V.3 Метод "Гусеница" для обнаружения разладки в тренде.

Д.Л. Данилов А.А. Жиглявский

1 Постановка задачи

Предположим, что $\xi(t)$, $t \in [0, T]$ — случайный процесс со средним (трендом) $\lambda(t) = E\xi(t)$ и известной ковариационной функцией

$$R(t, s) = E(\xi(t) - \lambda(t))(\xi(s) - \lambda(s)),$$

причем время t может быть как дискретным, так и непрерывным.

Рассмотрим задачу о разладке в тренде $\lambda(t)$. Конкретно, рассмотрим задачу проверки гипотезы об отсутствии разладки

$$H_0 : \lambda(t) = \lambda_0(t) \quad \text{при всех } t \in [0, T]$$

при альтернативе о ее наличии:

$$H_1 : \exists \kappa \in (0, T) \quad \lambda(t) = \begin{cases} \lambda_0(t), & 0 \leq t < \kappa, \\ \lambda_1(t), & \kappa \leq t \leq T. \end{cases}$$

Здесь $\lambda_0(t)$ и $\lambda_1(t)$ — функции (возможно, неизвестные) из некоторых классов функций Λ_0 и Λ_1 , такие, что $\lambda_0 \neq \lambda_1$, а κ — неизвестный момент разладки (в случае справедливости гипотезы H_1).

Ниже используются два основных класса функций Λ_0 . В разделе 3 $\Lambda_0 = \mathcal{P}_l$ — множество полиномов порядка не более чем l , а в разделе 2 Λ_0 состоит из функций вида

$$\lambda(t) = \sum_{k=1}^m a_k(t) e^{-\lambda_k t} \sin(\omega_k t + \varphi_k), \quad (1)$$

где для всех $k = 1, \dots, m$ функции $a_k(t)$ являются (неизвестными) полиномами от t , а $\lambda_k, \omega_k, \varphi_k$ — неизвестные параметры.

Для обнаружения разладки, т.е. для проверки гипотезы H_0 , ниже применяются следующие два подхода:

- основанный на свойствах рядов конечного ранга (раздел 2);
- использующий свойства "гусеничного центрирования" (раздел 3).

В случае отсутствия случайного шума первый из указанных подходов основан на простом факте, состоящем в том, что после разладки ранг функции конечного ранга начинает увеличиваться при каждом новом измерении в случае дискретного времени и становится, вообще говоря, бесконечным в случае непрерывного времени. Этот подход может быть использован и в случае наличия небольшого аддитивного шума, возможность его применения в этом случае обосновывается с помощью формул теории возмущений.

Второй из указанных подходов, основанный на использовании свойств "гусеничного центрирования", может быть использован, в частности, для обнаружения разладки в коэффициентах полиномиального тренда. Идея метода состоит в том, что t -кратное "гусеничное центрирование" преобразует случайный процесс с полиномиальным трендом степени не выше t в некоторый другой процесс с нулевым средним. Изучение характеристик преобразованного процесса, и в первую очередь его ковариационной функции, позволяет сформулировать целый ряд алгоритмов обнаружения разладки и сделать некоторые выводы о свойствах этих алгоритмов.

2 Подход, основанный на свойствах рядов конечного ранга

Предположим сначала, что исходный временной ряд $f(t)$, ($t \in [0, T]$) имеет вид (1) с некоторыми $a_k(t), \lambda_k, \omega_k, \varphi_k$, а время t может быть как дискретным, так и непрерывным.

Как следует из раздела VI.1, секция 3.1, в случае непрерывного времени при любой длине гусеницы τ только конечное число собственных чисел гусеничного ковариационного оператора (см. раздел VI.2) отлично от нуля. В случае дискретного времени ($t = 1, \dots, T$) число измерений T и длина гусеницы τ должны быть достаточно велики для того, чтобы определить ранг ряда (1), который равен числу ненулевых собственных чисел гусеничного ковариационного оператора.

С другой стороны, если κ, τ достаточно велики для того, чтобы определить ранг временного ряда (1), для определенности равный r при $t < \kappa$, а в момент времени κ наступает разладка (т.е. $f(\kappa)$ определяется по формуле отличной от (1) с заданными $a_k(t), \lambda_k, \omega_k, \varphi_k$), то ранг ряда $f(t)$, $t \leq \kappa$, автоматически увеличивается на единицу. Более того, каждое из следующих измерений $f(t)$, $t = \kappa + 1, \kappa + 2, \dots$ также увеличивает ранг ряда.

В случае непрерывного времени наличие разладки в момент времени κ (т.е. непредставимость $f(t)$, $t > \kappa$, в виде (1), с теми же $a_k(t), \lambda_k, \omega_k, \varphi_k$) делает процесс конечного ранга $f(t)$, $0 < t < \kappa$, процессом $f(t)$, $0 < t \leq T$, $\kappa < T$, имеющим, вообще говоря, бесконечный ранг.

Таким образом, как в дискретном, так и в непрерывном случаях изменение ранга процесса $f(t)$ в момент времени κ свидетельствует о наличии разладки в этот момент времени.

Предложенный метод может быть использован и в том случае, когда представление (1) выполняется приближенно (например, в случае наличия небольшого аддитивного шума). В этом случае собственные числа невозмущенного ковариационного оператора (соответствующего представлению (1)) изменятся незначительно, и величина их изменения может быть определена с помощью формул теории возмущений, см. [1]. Приведем одну из наиболее простых формул подобного типа.

Пусть T — ковариационная матрица невозмущенного ряда (конечного ранга) и пусть T_1 — возмущающий оператор с матрицей $T_1 = Y^T X + X^T Y + Y^T Y$, где X — матрица наблюдений невозмущенного ряда, Y — матрица наблюдений

шума. В этом случае мажорирующая функция для возмущений полупростого собственного числа λ имеет вид

$$\psi(\Delta) = p\Delta + pq\Delta^2 + \frac{2pq(ps+q)\Delta^3 + 2p^3q^2s\Delta^4}{1 - (ps+q)\Delta - 2pq s\Delta^2 + \Omega(\Delta)},$$

где

$$\Omega(\Delta) = \sqrt{(1 - (ps+q)\Delta)^2 - 4pq s\Delta^2},$$

$$p = \|T_1 P\|, \quad q = \|T_1 S\|, \quad s = \|S - \alpha P\|,$$

P — невозмущенный собственный проектор соответствующий собственному числу λ , S — приведенная резольвента оператора T относительно собственного значения λ в точке λ , α — вещественное число. (Приведенной резольвентой называют голоморфную часть ряда Лорана резольвенты $R(\xi, T) = (T - \xi)^{-1}$ оператора T .) В этом случае первые три коэффициента ряда теории возмущений для λ оцениваются следующим образом:

$$|\hat{\lambda}^{(1)}| \leq p, \quad |\hat{\lambda}^{(2)}| \leq pq, \quad |\hat{\lambda}^{(3)}| \leq pq(ps+q).$$

Ряд теории возмущений имеет, в общем случае, вид

$$\lambda(\Delta) - \lambda = \sum_{k=1}^{\infty} |\hat{\lambda}^{(k)}| \Delta^k.$$

При этом, если норма оператора T_1 мала, либо $\|T_1 P\|$ мало, то коэффициенты ряда теории возмущений малы, также как и суммарное возмущение любого собственного значения.

3 Применение алгоритма "аддитивная гусеница" для обнаружения разладки временных рядов

Пусть $\xi(t)$, $t \in [0, T]$ — исходный случайный процесс со средним $\lambda(t) = E\xi(t)$ и ковариационной функцией

$$R(t, s) = E(\xi(t) - \lambda(t))(\xi(s) - \lambda(s)).$$

Время t может быть как непрерывным, так и дискретным.

Рассмотрим схему метода "гусеница" с однократным центрированием. Зададим τ , $0 < \tau < T/2$ — "длину гусеницы". Проведем преобразование исходного процесса $\xi(t)$ в двумерное случайное поле по формуле

$$g(x, s) = \xi(x + s), \quad 0 \leq x \leq T - \tau, \quad 0 \leq s \leq \tau.$$

Определим "гусеничное среднее":

$$\mathcal{E}_\tau \xi(s) = \frac{1}{T - \tau} \int_0^{T-\tau} g(x, s) dx = \frac{1}{T - \tau} \int_0^{T-\tau} \xi(x + s) dx.$$

Здесь и далее dx — Лебегова мера в случае непрерывного времени и считающая мера — в случае дискретного времени (в этом случае интегралы являются конечными суммами).

Однократно центрированное случайное поле имеет вид

$$\bar{g}(x, s) = \xi(x + s) - \frac{1}{T - \tau} \int_0^{T-\tau} \xi(t + s) dt.$$

Проведем "ганделизацию" поля $\bar{g}(x, s)$, определяя тем самым процесс

$$\beta(t) = \frac{\int_0^{T-\tau} \int_0^\tau \delta(x + s = t) \bar{g}(x, s) dx ds}{\int_0^{T-\tau} \int_0^\tau \delta(x + s = t) dx ds}, \quad (2)$$

(здесь δ – дельта-функция Дирака, знаменатель в правой части (2) представляет собой длину отрезка прямой $x + s = t$ в прямоугольнике $(x, s) \in [0, T - \tau] \times [0, \tau]$). Проводя элементарные преобразования, легко показать что

$$\beta(t) = \xi(t) - \varphi(t),$$

где

$$\varphi(t) = \frac{1}{(T - \tau) \mu_t} \int_0^{T-\tau} dy \int_{M_1(t)}^{M_2(t)} \xi(y + s) ds,$$

$$\text{а } \mu_t = \mu_\tau(t) = M_2(t) - M_1(t),$$

$$M_1(t) = \max\{0, t - T + \tau\}, \quad M_2(t) = \min\{t, \tau\},$$

$$\mu_a(u) = \begin{cases} u, & u \leq a, \\ a, & a < u < T - a, \\ T - u & u \geq T - a. \end{cases}$$

Выражение для $\varphi(t)$ может быть переписано в виде:

$$\varphi(t) = \frac{1}{(T - \tau) \mu_t} \int_{M_1(t)}^{T-\tau+M_2(t)} \xi(z) \mu_t(z) dz.$$

Очевидно, что если $\lambda(t) = E\xi(t) = \text{const}$, то $E\beta(t) = 0$.

Ковариационная функция процесса $\beta(t)$ не зависит от $\lambda(t)$, поэтому при подсчете этой функции положим $\lambda(t) = 0$. Имеем:

$$\begin{aligned} R_\beta(t, s) &= E\beta(t)\beta(s) = R(t, s) - E\xi(t)\varphi(s) - E\xi(s)\varphi(t) + E\varphi(t)\varphi(s) = \\ &= E\xi(t)\xi(s) - \frac{1}{(T - \tau) \mu_t} \int_0^{T-\tau} dy \int_{M_1(t)}^{M_2(t)} E\xi(s)\xi(x + y) dx - \\ &\quad - \frac{1}{(T - \tau) \mu_s} \int_0^{T-\tau} dy \int_{M_1(s)}^{M_2(s)} E\xi(t)\xi(x + y) dx + \\ &\quad + \frac{1}{(T - \tau)^2 \mu_t \mu_s} \int_0^{T-\tau} dy_1 \int_{M_1(t)}^{M_2(t)} E\xi(x_1 + y_1)\xi(x_2 + y_2) dx_1 \int_0^{T-\tau} dy_2 \int_{M_1(s)}^{M_2(s)} dx_2 = \\ &= R(t, s) - \frac{1}{(T - \tau) \mu_t} \int_0^{T-\tau} dy \int_{M_1(t)}^{M_2(t)} R(s, x + y) dx - \\ &\quad - \frac{1}{(T - \tau) \mu_s} \int_0^{T-\tau} dy \int_{M_1(s)}^{M_2(s)} R(t, x + y) dx + \\ &\quad + \frac{1}{(T - \tau)^2 \mu_t \mu_s} \int_0^{T-\tau} dy_1 \int_0^{T-\tau} dy_2 \int_{M_1(t)}^{M_2(t)} dx_1 \int_{M_1(s)}^{M_2(s)} R(x_1 + y_1, x_2 + y_2) dx_2. \end{aligned}$$

В частности, дисперсия процесса $\beta(t)$ в точке t равна

$$\begin{aligned}\sigma^2(t) &= R(t,t) - \frac{2}{(T-\tau)\mu_t} \int_0^{T-\tau} dy \int_{M_1(t)}^{M_2(t)} R(t,x+y) dx + \\ &+ \frac{1}{(T-\tau)^2\mu_t^2} \int_0^{T-\tau} dy_1 \int_0^{T-\tau} dy_2 \int_{M_1(t)}^{M_2(t)} dx_1 \int_{M_1(t)}^{M_2(t)} R(x_1+y_1, x_2+y_2) dx_2.\end{aligned}$$

Рассмотрим гипотезу

$$H_0 : \lambda(t) = 0.$$

Естественными тестовыми статистиками для проверки гипотезы H_0 являются статистики

$$\sup_{t \in [0,T]} (\beta(t) - h(t)), \quad (3)$$

$$\sup_{t \in [0,T]} (\varphi(t) - h(t)), \quad (4)$$

где $h(t)$ — некоторая пороговая функция, а решение о разладке принимается в случае когда значение supremum'a положительно. В основных примерах, рассматриваемых ниже, $\xi(t)$ является "белым шумом", откуда следует, что статистика (3) с вероятностью 1 неограничена, т.е. $\Pr \left\{ \sup_t \beta(t) > h \right\} = 1$ для любого $h > 0$.

Для проверки гипотезы H_0 мы будем, следовательно, рассматривать статистику (4) и отвергать гипотезу H_0 в том случае, когда

$$\exists t \in [0,T] : \varphi(t) - h(t) \geq 0 \quad (5)$$

для некоторого порога $h(t)$. Вероятность ошибки I рода (отвергнуть H_0 , когда она верна) равна

$$\lambda_h = \Pr \left\{ \sup_{t \in [0,T]} (\varphi(t) - h(t)) \geq 0 \mid \lambda(t) = 0 \right\}.$$

Аппроксимируем эту вероятность для случая высокого порога h (т.е. $h \rightarrow \infty$). Для этого используем "касательную аппроксимацию" Дж.Дурбина (см. [2] и [3], стр. 56). Сначала представим λ_h в виде

$$\lambda_h = \int_0^T q_h(t) dt,$$

где $q_h(t)$ — плотность времени первого достижения барьера h процессом $\varphi(t)$ со средним 0 и ковариационной функцией

$$r(t,s) = \frac{1}{(T-\tau)^2\mu_t\mu_s} \int_0^{T-\tau} dy_1 \int_0^{T-\tau} dy_2 \int_{M_1(t)}^{M_2(t)} dx_1 \int_{M_1(s)}^{M_2(s)} dx_2 R(x_1+y_1, x_2+y_2).$$

Согласно аппроксимации Дурбина, плотность $q_h(t)$ представима в виде $q_h(t) = b(t)f(t)$, где

$$\begin{aligned} b(t) &= \lim_{s \uparrow t} \frac{1}{t-s} E \left[1_{[\sup_{t \in [0,s]} (\varphi(t)-h(t)) < 0]} \cdot (h(s) - \varphi(s)) \mid \varphi(t) = h(t) \right] \simeq \\ &\simeq b_0(t) = \lim_{s \uparrow t} \frac{1}{t-s} E [(h(s) - \varphi(s)) \mid \varphi(t) = h(t)] = \\ &= \frac{h(t)}{r(t,t)} \frac{\partial r(s,t)}{\partial s} \Big|_{s=t-} - h'(t), \\ \text{а } f(t) &= \frac{1}{\sqrt{2\pi r(t,t)}} \exp \left\{ -\frac{h^2(t)}{2r(t,t)} \right\} \text{ соответственно.} \end{aligned}$$

Поскольку ξ – "белый шум", то с учетом вида $r(t,s)$ легко получить формулы

$$r(t,s) = \frac{1}{(T-\tau)^2} \left\{ T - \tau + \frac{T-t-(T-s)}{2} - \frac{(T-t)^2}{3(T-s)} \right\}$$

при $t, s \in [T-\tau, T]$, и

$$\begin{aligned} \frac{\partial r(t,s)}{\partial s} &= \frac{\partial}{\partial s} \left\{ T - \tau + \frac{s-T}{2} - \frac{(T-t)^2}{3(T-s)} \right\} = \frac{1}{(T-\tau)^2} \left\{ \frac{1}{2} - \frac{(T-t)^2}{3(T-s)^2} \right\}, \\ \frac{\partial r(s,t)}{\partial s} \Big|_{s=t-} &= \frac{1}{(T-\tau)^2} \left\{ \frac{1}{2} - \frac{1}{3} \right\} = \frac{1}{6(T-\tau)^2}, \\ \frac{\partial r(s,t)}{\partial s} \Big|_{s=t-} &= 0 \quad \text{при } \tau < t < T-\tau. \end{aligned}$$

Из этого вытекает

$$\begin{aligned} b_0(t) &= \frac{h(t)}{\frac{1}{(T-\tau)^2} \left\{ \tau - T - \frac{T-t}{3} \right\}} \cdot \frac{1}{6(T-\tau)^2} - h'(t) = \frac{h(t)}{6 \left(T - \tau - \frac{T-t}{3} \right)} - h'(t), \\ f(t) &= \frac{1}{\sqrt{2\pi} \frac{1}{6(T-\tau)^2} \exp \left\{ -\frac{h^2}{2} \cdot \frac{6}{T-\tau-\frac{T-t}{3}} \right\}} = \sqrt{\frac{3}{\pi}} (T-\tau) \exp \left\{ -\frac{-3h^2}{T-\tau-\frac{T-t}{3}} \right\}. \end{aligned}$$

В частности, если $h(t) = h = \text{const}$, то

$$\begin{aligned} \alpha_h &\simeq \int_0^T f(t)b_0(t) dt = \\ &= \frac{2h(T-\tau)}{6\sqrt{2\pi}} \int_{T-\tau}^T \frac{1}{\left(T - \tau - \frac{T-t}{3} \right)^{3/2}} \exp \left\{ -\frac{h^2}{2} \frac{(T-\tau)^2}{\left(T - \tau - \frac{T-t}{3} \right)} \right\} dt = \\ &= \text{Erf} \left(\frac{\sqrt{3} h(T-\tau)}{\sqrt{6T-8\tau}} \right) - \text{Erf} \left(\frac{1}{\sqrt{2}} h\sqrt{T-\tau} \right), \end{aligned}$$

где

$$\text{Erf}(a) = \frac{2}{\sqrt{\pi}} \int_0^a e^{-t^2} dt.$$

Поскольку корреляционные функции процессов $\beta(t)$ и $\varphi(t)$ имеют сложный вид, то аппроксимации вероятностей ошибок при проверке гипотезы H_0 достаточно сложны. Существенно более простые формулы для вероятностей ошибок получаются в случае использования статистики

$$\gamma(T) = \beta(T) - \beta(0).$$

Изучим второй момент случайной величины $\beta(T) - \beta(0)$. Из формулы (2) вытекает что

$$\begin{aligned}\beta(0) &= \xi(0) - \frac{1}{T-\tau} \int_0^{T-\tau} \xi(x) dx, \\ \beta(T) &= \xi(T) - \frac{1}{T-\tau} \int_\tau^T \xi(x) dx.\end{aligned}$$

Следовательно,

$$\begin{aligned}E(\beta(T) - \beta(0))^2 &= E\left[\xi(T) - \xi(0) + \frac{1}{T-\tau} \int_0^{T-\tau} (\xi(x) - \xi(x+\tau)) dx\right]^2 = \\ &= R(T, T) + R(0, 0) - 2R(T, 0) - \\ &\quad - \frac{2}{T-\tau} \int_0^{T-\tau} [R(T, x+\tau) - R(T, x) - R(0, x+\tau) + R(0, x)] dx + \\ &\quad + \frac{1}{(T-\tau)^2} \int_0^{T-\tau} \int_0^{T-\tau} [R(x, y) + R(x+\tau, y+\tau) - 2R(x+\tau, y)] dx dy. \quad (6)\end{aligned}$$

Формула (6) будет использована при исследовании критерия, построенного по статистике $\gamma(T) = \beta(T) - \beta(0)$, для проверки гипотез о разладке процесса $\xi(t)$.

В предположении, что $\xi(t)$ — гауссовский процесс, статистика $\gamma(T) = \beta(T) - \beta(0)$ распределена по нормальному закону со средним

$$E(\gamma(T)) = \lambda(T) - \lambda(0) - \frac{1}{T-\tau} \int_0^{T-\tau} (\lambda(x+\tau) - \lambda(x)) dx$$

и вторым моментом (6).

Рассмотрим простую вспомогательную задачу.

Пусть ξ — значение случайной нормально распределенной величины,

$$\begin{aligned}H_0 : \quad \xi &\sim N(0, \Delta^2), \\ H_1 : \quad \xi &\sim N(a, \Delta^2),\end{aligned}$$

где $a \geq 0$, а для проверки гипотезы H_0 используется тест

$$\xi > \Delta t_\alpha, \quad (7)$$

где t_α — $(1 - \alpha)$ -квантиль стандартного нормального распределения:

$$\Phi(t_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_\alpha} e^{-t^2/2} dt = 1 - \alpha.$$

Тогда, очевидно, вероятность ошибки первого рода при проверке гипотезы H_0 при альтернативе H_1 равна α , а вероятность ошибки второго рода равна

$$\Pr\{\text{принять } H_0 | H_1\} = 1 - \Phi(t_\alpha + a/\Delta).$$

Если $a = a(\theta)$ зависит от параметра θ , то функция мощности критерия (7) равна

$$\beta(\theta) = \Phi\left(t_\alpha + \frac{a(\theta)}{\Delta}\right).$$

Если знак a неизвестен и рассматривается двусторонний критерий

$$|\xi| > t_{\alpha/2},$$

то вероятность ошибки I рода равна α и в этом случае, а функция мощности равна

$$\beta(\theta) = \Phi\left(t_{\alpha/2} + \frac{a(\theta)}{\Delta}\right).$$

Приведем несколько примеров обнаружения разладки с помощью статистики $\gamma(T)$.

Пусть $R(t, s) = \sigma^2 \delta(t - s)$, т.е. процесс $\xi(t)$ является белым шумом, и гипотеза H_0 об отсутствии разладки имеет вид

$$H_0 : \lambda(t) = 0 \quad \forall t.$$

Далее под γ_0 и γ_1 будем обозначать тестовую статистику γ в случаях справедливости гипотез H_0 и H_1 , соответственно. Заметим что

$$\gamma_1(T) = \gamma_0(T) + \lambda(\tau) - \lambda(0) + \frac{1}{T - \tau} \int_0^{T-\tau} [\lambda(t) - \lambda(t + \tau)] dt.$$

Формулы такого типа в дальнейшем будут называться (γ_0, γ_1) -формулами.

В предположении справедливости H_0 выполняется $E\gamma(T) = 0$,

$$\text{Var } \gamma(T) = \left(2 - \frac{4}{T - \tau}\right) \sigma^2 = \Delta^2.$$

Рассмотрим теперь различные альтернативные гипотезы H_1 , соответствующие различным видам разладок.

Пример 1. Разладка для ступеньки.

Пусть

$$H_1 : \lambda(t) = \begin{cases} 0, & t \leq \kappa, \\ \theta, & t > \kappa, \end{cases}$$

где κ — неизвестный момент разладки, $\tau \leq \kappa \leq T - \tau$, $\tau < T/2$, и значение параметра θ также может быть неизвестно. (Исследованию алгоритмов обнаружения разладки указанного вида посвящена, в значительной степени, монография [3].)

В этом случае

$$\lambda(0) = 0, \quad \lambda(T) = \theta \text{ и}$$

$$\lambda(t) - \lambda(t + \tau) = \begin{cases} 0 & 0 \leq t \leq \kappa - \tau, \\ -\theta & \kappa - \tau \leq t \leq \kappa, \\ 0 & \kappa \leq t \leq T. \end{cases}$$

Поэтому

$$\int_0^{T-\tau} \lambda(t) - \lambda(t + \tau) dt = -\theta \cdot \text{length}([\kappa - \tau, \kappa]) = -\theta\tau.$$

и

$$\gamma_1(T) = \gamma_0(T) + \theta - \frac{\theta\tau}{T-\tau} = \gamma_0(T) + \theta \frac{T-2\tau}{T-\tau},$$

$$E\gamma_1(T) = E\gamma_0(T) + \theta \frac{T-2\tau}{T-\tau} = E\gamma_0(t) + a(\theta, \tau, T).$$

Кроме того,

$$\text{Var } \gamma_1(T) = \text{Var } \gamma_0(T)$$

Заметим, что

$$a(\theta, \tau, T) = 0 \quad \text{при } \tau = \frac{T}{2},$$

$$\lim_{\tau \rightarrow 0} a(\theta, \tau, T) = \theta.$$

Следовательно, в предположении справедливости гипотезы H_1 статистика γ является нормально распределенной случайной величиной со средним $E\gamma = a(\theta, \tau, T)$ и дисперсией: $\text{Var } \gamma = \Delta^2$. Таким образом, получаем следующее выражение для функции мощности этого критерия (7):

$$\beta(\kappa, T, \tau) = \Phi \left(t_\alpha + \theta \frac{T-2\tau}{T-\tau} / \left(\sigma \sqrt{2 - \frac{4}{T-\tau}} \right) \right).$$

Пример 2. Разладка – прямоугольник

Пусть

$$H_1 : \lambda(t) = \begin{cases} 0 & t \leq \kappa_1, \\ \theta & \kappa_1 \leq t \leq \kappa_2, \\ 0 & \kappa_2 \leq t. \end{cases}$$

где $0 \leq \kappa_1 \leq \kappa_2$.

Пусть сначала $0 \leq \tau \leq \kappa_1 \leq \kappa_2 \leq T - \tau$ и $\tau < \frac{T}{2}$. Тогда $\lambda(\tau) = \lambda(0) = 0$,

$$\frac{1}{T-\tau} \int_0^{T-\tau} \lambda(t) dt = \frac{1}{T-\tau} \theta(\kappa_2 - \kappa_1) = \frac{\theta(\kappa_2 - \kappa_1)}{T-\tau},$$

$$\frac{1}{T-\tau} \int_0^{T-\tau} \lambda(t+\tau) dt = \frac{1}{T-\tau} \int_\tau^T \lambda(t) dt = \frac{\theta(\kappa_2 - \kappa_1)}{T-\tau}.$$

Таким образом, $\gamma_1(T) = \gamma_0(T)$.

Пусть теперь $\kappa_1 < \tau$, но $\kappa_2 \leq T - \tau$. Тогда

$$\frac{1}{T-\tau} \int_0^{T-\tau} \lambda(t) dt = \frac{\theta(\kappa_2 - \kappa_1)}{T-\tau},$$

$$\frac{1}{T-\tau} \int_0^{T-\tau} \lambda(t) dt = \frac{\theta(\kappa_2 - \tau)}{T-\tau} = \frac{\theta(\kappa_2 - \max(\tau, \kappa_1))}{T-\tau}.$$

В случае $\kappa_2 > T - \tau$:

$$\frac{1}{T-\tau} \int_0^{T-\tau} \lambda(t) dt = \frac{\theta(T - \tau - \kappa_1)}{T-\tau},$$

$$\frac{1}{T-\tau} \int_0^{T-\tau} \lambda(t) dt = \frac{\theta(\min(\kappa_2, T - \tau) - \min(\kappa_1, T - \tau))}{T-\tau},$$

$$\frac{1}{T-\tau} \int_0^{T-\tau} \lambda(t) dt = \frac{\theta(\max(\kappa_2, \tau) - \max(\tau, \kappa_1))}{T-\tau}.$$

Таким образом, мы получили следующую общую формулу для однократного центрирования и разладки вида (3),

$$\gamma_1(T) = \gamma_0(T) + \frac{\theta}{T-\tau} \{ \min(\kappa_2, T-\tau) - \min(\kappa_1, T-\tau) - \max(\kappa_2, \tau) + \max(\kappa_1, \tau) \}.$$

что соответствует

$$a(\kappa, \tau, T) = \frac{\theta}{T-\tau} \{ \min(\kappa_2, T-\tau) - \min(\kappa_1, T-\tau) - \max(\kappa_2, \tau) + \max(\kappa_1, \tau) \}.$$

Пример 3. Общий случай финитной разладки.

Пусть альтернативная гипотеза H_1 такая что

$$\lambda(t) : \text{ supp } \lambda \in [\kappa_1, \kappa_2] \subset [0, T].$$

Тогда

$$\begin{aligned} \frac{1}{T-\tau} \int_0^{T-\tau} \lambda(t) dt &= \frac{1}{T-\tau} \int_{\kappa_1}^{\min(\kappa_2, T-\tau)} \lambda(t) dt, \\ \frac{1}{T-\tau} \int_{\tau}^T \lambda(t) dt &= \frac{1}{T-\tau} \int_{\max(\kappa_1, \tau)}^{\max(\kappa_2, T)} \lambda(t) dt, \\ \gamma_1(T) &= \gamma_0(T) + \frac{1}{T-\tau} \left(\int_{\kappa_1}^{\min(\kappa_2, T-\tau)} \lambda(t) dt - \int_{\max(\kappa_1, \tau)}^{\max(\kappa_2, T)} \lambda(t) dt \right). \end{aligned}$$

Вывод (γ_0, γ_1) -формул для двукратного центрирования.

Рассмотрим схему аддитивной "Гусеницы" с двукратным центрированием:

$$\begin{aligned} g(x, s) &= \xi(x+s), \\ \bar{g}(x, s) &= g(x, s) - \frac{1}{T-\tau} \int_0^{T-\tau} g(t, s) dt, \\ \bar{\bar{g}}(x, s) &= \bar{g}(x, s) - \frac{1}{\tau} \int_0^{\tau} \bar{g}(t, s) ds. \end{aligned}$$

Перепишем последнее выражение в развернутом виде:

$$\begin{aligned} \bar{\bar{g}}(x, s) &= \bar{g}(x, s) - \frac{1}{T-\tau} \int_0^{T-\tau} \bar{g}(t, s) dt - \\ &- \frac{1}{\tau} \int_0^{T-\tau} (g(x, s) - \frac{1}{T-\tau} \int_0^{T-\tau} g(t, s) dt) ds = g(x, s) - \\ &- \frac{1}{\tau} \int_0^{\tau} g(x, s) ds - \frac{1}{T-\tau} \int_0^{T-\tau} g(t, s) dt + \frac{1}{\tau(T-\tau)} \int_0^{\tau} ds \int_0^{T-\tau} g(t, s) dt. \end{aligned}$$

Учитывая что $g(t, s) = \xi(t+s)$, получаем

$$\begin{aligned} \bar{\bar{g}}(x, s) &= \xi(x+s) - \frac{1}{\tau} \int_0^{\tau} \xi(x+s) ds - \\ &- \frac{1}{T-\tau} \int_0^{T-\tau} \xi(t+s) dt + \frac{1}{\tau(T-\tau)} \int_0^{\tau} ds \int_0^{T-\tau} \xi(t+s) dt. \end{aligned}$$

Проводя "ганкелизацию" g , получаем процесс $\beta(t)$, такой что

$$\begin{aligned}\beta(0) &= \xi(0) - \frac{1}{\tau} \int_0^\tau \xi(s) ds - \frac{1}{T-\tau} \int_0^{T-\tau} \xi(t) dt + \frac{1}{\tau(T-\tau)} \int_0^\tau ds \int_0^{T-\tau} \xi(t+s) dt, \\ \beta(T) &= \xi(T) - \frac{1}{\tau} \int_0^\tau \xi(T-\tau+s) ds - \\ &\quad - \frac{1}{T-\tau} \int_0^{T-\tau} \xi(\tau+t) dt + \frac{1}{\tau(T-\tau)} \int_0^\tau ds \int_0^{T-\tau} \xi(t+s) dt.\end{aligned}$$

В дальнейшем для обнаружения разладки будут использоваться тестовая статистика вида $\gamma(T) = \beta(T) - \beta(0)$. При отсутствии разладки (т.е. при выполнении гипотезы $H_0 : \lambda(t) = 0$)

$$\begin{aligned}\gamma_0(T) &= \beta_0(T) - \beta_0(0) = \\ &= \xi(T) - \xi(0) - \frac{1}{\tau} \int_0^\tau \xi(T-\tau+s) ds + \frac{1}{\tau} \int_0^\tau \xi(s) ds - \\ &\quad - \frac{1}{T-\tau} \int_0^{T-\tau} \xi(\tau+t) dt + \frac{1}{T-\tau} \int_0^{T-\tau} \xi(t) dt.\end{aligned}$$

При наличии разладки в среднем имеем следующее выражение для тестовой статистики:

$$\begin{aligned}\gamma_1(T) &= \gamma_0(T) + \lambda(T) - \lambda(0) = \\ &= -\frac{1}{\tau} \int_0^\tau \lambda(T-\tau+s) ds + \frac{1}{\tau} \int_0^\tau \lambda(s) ds - \\ &\quad - \frac{1}{T-\tau} \int_0^{T-\tau} \lambda(\tau+t) dt + \frac{1}{T-\tau} \int_0^{T-\tau} \lambda(t) dt,\end{aligned}$$

что представляет собой основную формулу типа (γ_0, γ_1) для двукратного центрирования. Рассмотрим несколько частных случаев.

Пример 4. (Разладка для ступеньки.)

Пусть

$$H_1 : \lambda(t) = \begin{cases} 0, & t \leq \kappa, \\ \theta, & t > \kappa, \end{cases}$$

где κ — неизвестный момент разладки и $\tau \leq \kappa \leq T - \tau$, $\tau < T/2$. Тогда

$$\lambda(0) = \lambda(T) = 0, \quad \int_0^\tau \lambda(s) (ds) = 0,$$

$$\int_0^{T-\tau} \lambda(t) dt = (T - \tau - \kappa) \theta,$$

$$\int_0^\tau \lambda(T-\tau+s) ds = \int_{T-\tau}^T \lambda(s) ds = \tau \theta,$$

$$\int_0^{T-\tau} \lambda(t+\tau) dt = \int_\tau^T \lambda(t) dt = (T - \kappa) \theta,$$

$$\begin{aligned}\gamma_1(T) &= \gamma_0(T) - \theta - \frac{T-\kappa}{T-\tau} \theta + \frac{T-\tau-\kappa}{T-\tau} \theta = \\ &= \gamma_0(T) - \theta \left(1 + \frac{T-\kappa}{T-\tau} - \frac{T-\tau-\kappa}{T-\tau} \right) = \gamma_0(T) - \frac{\theta T}{T-\tau}.\end{aligned}$$

Перейдем к выводу формулы дисперсии $\gamma_0(T)$ в случае двукратного центрирования.

$$\begin{aligned}
\text{Var}\gamma_0(T) &= R(T, T) + R(0, 0) + \\
&+ \frac{1}{\tau^2} \int_0^\tau \int_0^\tau R(t, s) ds dt + \frac{1}{\tau^2} \int_0^\tau \int_0^\tau R(T - t + s, T - \tau + t) ds dt + \\
&+ \frac{1}{(T - \tau)^2} \int_0^{T-\tau} \int_0^{T-\tau} R(s + \tau, t + \tau) ds dt + \frac{1}{(T - \tau)^2} \int_0^{T-\tau} \int_0^{T-\tau} R(t, s) ds dt + \\
&+ \frac{2}{\tau} \left[\int_0^\tau R(T, s) ds - \int_0^\tau R(T, T - \tau + s) ds \right] + \\
&\left[+ \frac{2}{\tau} \int_0^\tau R(0, T - \tau + s) ds - \int_0^\tau R(0, s) ds \right] + \\
&+ \frac{2}{T - \tau} \left[\int_0^{T-\tau} (R(T, s + \tau) - R(\tau, s) - R(0, s + \tau) + R(0, s)) ds \right] - \\
&- 2R(0, T) - \frac{2}{\tau^2} \int_0^\tau \int_0^\tau R(T - \tau + s, t) ds dt - \\
&- 2 \frac{2}{(T - \tau)^2} \int_0^{T-\tau} \int_0^{T-\tau} R(\tau + s, t) ds dt + \\
&+ \frac{2}{\tau(T - \tau)} \left[\int_0^\tau \int_0^{T-\tau} R(s, t + \tau) ds dt - \int_0^\tau \int_0^{T-\tau} R(T - \tau + s, t + \tau) ds dt + \right. \\
&\left. + \int_0^\tau \int_0^{T-\tau} R(T - \tau + s, t) ds dt - \int_0^\tau \int_0^{T-\tau} R(t, s) ds dt \right].
\end{aligned}$$

Пример 5. (Разладка в функции регрессии.)

Выведем общие формулы типа (γ_0, γ_1) для разладки в функции регрессии в случае двукратного центрирования. Пусть

$$H_1 : \lambda(t) = \begin{cases} 0 & t \leq \kappa \\ k(t - \kappa), & t > \kappa \end{cases}$$

где k – некоторый (возможно, неизвестный) параметр. Разладка в данном случае соответствует изменению угла наклона линейной функции.

Тогда

$$\lambda(0) = 0, \quad \lambda(T) = k(T - \kappa)$$

и задача сводится к вычислению следующих интегралов:

$$\begin{aligned}
I_1 &= \int_0^\tau \lambda(T - \tau + s) ds = \int_{T-\tau}^T \lambda(s) ds = \int_{\max(T-\tau, \kappa)}^T \lambda(s) ds, \\
I_2 &= \int_0^{T-\tau} \lambda(s + \tau) ds = \int_\tau^T \lambda(s) ds = \int_{\max(\tau, \kappa)}^T \lambda(s) ds, \\
I_3 &= \int_0^\tau \lambda(s) ds = \int_{\min(\kappa, \tau)}^T \lambda(s) ds, \\
I_4 &= \int_0^{T-\tau} \lambda(t) dt = \int_{\min(\kappa, T-\tau)}^T \lambda(t) dt.
\end{aligned}$$

В результате вычисления получаем

$$a(t, \tau, T) = k(T - \kappa) - I_1/\tau - I_2/(T - \tau) + I_3/\tau + I_4/(T - \tau) =$$

$$\begin{aligned}
&= \frac{k}{2\tau(T-\tau)} \left(6\tau^2\kappa + 2\kappa T^2 + 3\tau T^2 - 6\kappa\tau T - T^3 - 3T\tau^2 + 2\kappa\tau \max(\kappa, T-\tau) + \right. \\
&\quad + 2\kappa T \min(\kappa, \tau) + 2\kappa\tau \min(\kappa, T-\tau) + T(\max(\kappa, T-\tau))^2 + \tau(\min(\kappa, \tau))^2 + \\
&\quad + \tau(\max(\kappa, \tau))^2 - 2\kappa T \max(\kappa, T-\tau) - 2\kappa\tau \min(\kappa, \tau) - 2\kappa\tau \max(\kappa, \tau) - \\
&\quad \left. - \tau(\max(\kappa, T-\tau))^2 - T(\min(\kappa, \tau))^2 - \tau(\min(\kappa, T-\tau))^2 \right)
\end{aligned}$$

Пример 6. Разладка в регрессии и броуновское движение.

Рассмотрим случай когда ξ — броуновское движение и $\tau = T/2$. Тогда $R(t, s) = \min(t, s)$ и формула для дисперсии $\gamma_0(T)$ упрощается следующим образом:

$$\begin{aligned}
\text{Var } \gamma_0(T) &= T + \frac{1}{\tau^2} I_1 + \frac{1}{\tau^2} I_2 + \frac{1}{\tau^2} I_2 + \frac{1}{\tau^2} I_1 + \\
&+ \frac{2}{\tau} \left[\int_0^\tau s \, ds - \int_\tau^{2\tau} s \, ds \right] + \frac{2}{\tau} \left[\int_\tau^{2\tau} s \, ds - \int_0^\tau s \, ds \right] - \\
&- \frac{2}{\tau^2} I_4 - \frac{2}{\tau^2} I_4 + \frac{2}{\tau^2} [I_3 - I_2 + I_4 - I_1],
\end{aligned}$$

где

$$\begin{aligned}
I_1 &= \int_0^\tau dt \int_0^\tau \min(t, s) \, ds = \\
&= \int_0^\tau \left\{ \int_0^t s \, ds + \int_t^\tau t \, ds \right\} dt = \int_0^\tau \left\{ t^2/2 + t(\tau-t) \right\} dt = \frac{\tau^3}{3}, \\
I_2 &= \int_0^{2\tau} dt \int_\tau^{2\tau} \min(t, s) \, ds = \int_\tau^{2\tau} \left\{ t^2/2 - \frac{\tau^2}{2} + t\tau \right\} dt = \frac{4}{3}\tau^3, \\
I_3 &= \int_0^t dt \int_\tau^{2\tau} \min(t, s) \, ds = \tau \int_0^\tau t \, dt = \frac{\tau^3}{2}, \\
I_4 &= \int_\tau^{2\tau} dt \int_0^\tau \min(t, s) \, ds = \tau \int_0^\tau s \, ds = \frac{\tau^3}{2}.
\end{aligned}$$

Следовательно,

$$\text{Var } \gamma_0(T) = T + \frac{\tau}{3} + \frac{4}{3}\tau + \frac{\tau}{3} - \frac{2}{2}\tau - \frac{2}{2}\tau = T.$$

Если разладка в точке $\kappa = \frac{3}{2}\tau$ (в этой точке отклонение максимально, т.е. разладка проявляется наиболее сильно) то, очевидно, выполняется следующее равенство:

$$a(t, \tau, T) = \frac{k\tau}{4} = \frac{kT}{8}.$$

Поэтому $a/\sqrt{\text{Var } \gamma_0(T)} = k\sqrt{T}/8$, прямо пропорционально \sqrt{T} , и функция мощности равна

$$\Phi \left(t_\alpha + a/\sqrt{\text{Var } \gamma_0(T)} \right) = \Phi \left(t_\alpha + \frac{k\sqrt{T}}{8} \right).$$

В заключение приведем вывод формулы для ковариационной функции процесса $\beta(t)$ в случае двукратного центрирования.

Обозначим

$$\mu_t = \begin{cases} t, & t \leq \tau, \\ \tau, & \tau \leq t \leq T - \tau, \\ T - t, & t > T - \tau, \end{cases}$$

$$\begin{aligned} L_2(t) &= \min(t, T - \tau), & M_2(t) &= \min(t, \tau), \\ L_1(t) &= \max(0, t - \tau), & M_1(t) &= \max(0, t - T + \tau). \end{aligned}$$

Тогда

$$\begin{aligned} \beta(t) &= \xi(t) - \frac{1}{\tau \mu_t} \int_0^\tau dv \int_{L_1(t)}^{L_2(t)} \xi(x + v) dx - \\ &\quad - \frac{1}{(T - \tau) \mu_t} \int_0^{T-\tau} dy \int_{M_1(t)}^{M_2(t)} \xi(y + s) ds + \frac{1}{\tau(T - \tau)} \int_0^\tau ds \int_0^{T-\tau} \xi(x + s) dx \\ E\beta(t)\beta(p) &= \\ &= R(t, p) - \frac{1}{\tau \mu_p} \int_0^\tau dv \int_{L_1(t)}^{L_2(t)} R(t, x + v) dx - \\ &\quad - \frac{2}{(T - \tau) \mu_p} \int_0^{T-\tau} dy \int_{M_1(p)}^{M_2(p)} R(t, y + s) ds + \\ &\quad + \frac{2}{\tau(T - \tau)} \int_0^\tau ds \int_0^{T-\tau} R(t, x + s) dx - \frac{1}{\tau \mu_p} \int_0^\tau dv \int_{L_1(t)}^{L_2(t)} R(x + v, p) dx + \\ &\quad + \frac{1}{\tau^2 \mu_p \mu_t} \int_0^\tau dv_1 \int_{L_1(t)}^{L_2(t)} R(x_1 + v_1, x_2 + v_2) dx_1 \int_0^\tau dv_2 \int_0^\tau dv \int_{L_1(p)}^{L_2(p)} dx_2 - \\ &\quad + \frac{2}{(T - \tau) \tau \mu_t \mu_p} \int_0^\tau dv \int_{L_1(t)}^{L_2(t)} R(v + x, y + s) dx \int_0^{T-\tau} dy \int_{M_1(p)}^{M_2(p)} ds + \\ &\quad + \frac{1}{(T - \tau)^2 \mu_t \mu_p} \int_0^{T-\tau} dy_1 \int_{M_1(p)}^{M_2(p)} R(y_1 + s_1, y_2 + s_2) ds_1 \int_0^{T-\tau} dy_2 \int_{M_1(t)}^{M_2(t)} ds_2 - \\ &\quad - \frac{1}{\tau(T - \tau)^2 \mu_p} \int_0^{T-\tau} dy \int_{M_1(t)}^{M_2(t)} R(y + s_1, x + s_2) \int_0^\tau ds_2 \int_0^{T-\tau} dx + \\ &\quad - \frac{1}{\tau^2 (T - \tau) \mu_t} \int_0^\tau dv \int_{L_1(t)}^{L_2(t)} R(x_1 + v, x_2 + s) dx_1 \int_0^\tau ds \int_0^{T-\tau} dx_2 - \\ &\quad - \frac{1}{\tau^2 (T - \tau)^2 \mu_t} \int_0^{T-\tau} dy \int_{M_1(t)}^{M_2(t)} R(y + s_1, x + s_2) ds_1 \int_0^\tau ds_2 \int_0^{T-\tau} dx + \\ &\quad \frac{1}{\tau^2 (T - \tau)^2} \int_0^\tau ds_1 \int_0^{T-\tau} R(x_1 + s_1, x_2 + s_2) dx_1 \int_0^\tau ds_2 \int_0^{T-\tau} dx_2. \end{aligned}$$

Последняя формула представляет собой ковариационную функцию процесса $\beta(t)$ и может быть использована для построения критериев обнаружения разладки, аналогично случаю однократного центрирования.

Литература

1. Като Т, *Теория возмущений линейных операторов*, М., Мир, 1972.
2. Durbin J. The first-passage density of a continuous Gaussian process to a general boundary, Journal of Applied Probability, 1985, v.22, No. 1.
3. Жиглявский А.А., Красковский А.Е. *Обнаружение разладки случайных процессов в задачах радиотехники*, Изд-во Ленинградского университета, 1988, 222 с.

V.4 Геометрический подход к системе анализов многомерной статистики

Солнцев В.Н.

1 Введение

В этом разделе излагается единый геометрический подход к системе анализов многомерной статистики, позволяющий использовать их для исследования структур как в статистических, так и в детерминированных задачах. В основе подхода лежит задача построения подпространства заданной размерности, наилучшим образом приближающего семейство подпространств. Описан "метод согласованных вращений" ортонормальных базисов [2], являющийся некоторой модификацией и уточнением процедуры, предложенной в [1].

2 Основная статистическая модель

Рассмотрим некоторую совокупность (будем считать ее для простоты конечной) статистически связанных случайных величин $\{\xi_1, \dots, \xi_p\}$, имеющих совместную функцию распределения $F(x_1, \dots, x_p)$ и конечный второй момент $E\xi_i^2 \leq c < \infty$, $i = 1, \dots, p$.

Дополним эту совокупность вырожденной случайной величиной ξ_0 , принимающей значение 1 с вероятностью 1.

Линейная оболочка этой совокупности образует $(p+1)$ -мерное векторное пространство случайных величин $\eta = \sum_{i=0}^p \tau_i \xi_i$.

Введем скалярное произведение

$$\langle \eta, \zeta \rangle = \iint y z dG(y, z) = \int \dots \int y(x) z(x) dF(x_1, \dots, x_p) = E(\eta \zeta)$$

и согласованную с ним норму

$$\|\eta\| = \langle \eta, \eta \rangle^{1/2} = [E(\eta^2)]^{1/2}.$$

Очевидно, что в получившемся конечномерном евклидовом пространстве $\|\xi_0\| = 1$ и для всякого элемента η

$$\langle \eta, \xi_0 \rangle = \Pr_{\xi_0} \eta = E\eta.$$

Ортогональное дополнение к ξ_0 образует подпространство $L_2(\xi) = \{\eta - E\eta\}$ с индуцированным в нем скалярным произведением

$$\langle \eta_1, \eta_2 \rangle = E((\eta_1 - E\eta_1)(\eta_2 - E\eta_2)) = \text{cov}(\eta_1, \eta_2).$$

Квадрат нормы в этом подпространстве $\|\eta\|^2 = D(\eta)$, а косинус угла между элементами η_1 и η_2 совпадает с коэффициентом корреляции между соответствующими случайными величинами.

Размерность $L_2(\xi)$ равна рангу матрицы Σ , составленной из

$$\sigma_{ij} = \text{cov}(\xi_i, \xi_j), \quad i, j = 1, \dots, p.$$

В $L_2(\xi)$ можно построить некоторую систему координат, для чего следует построить ортонормальный базис. Стандартный метод построения такой системы, процесс Грамма-Шмидта, не дает однозначного решения, так как зависит от исходного набора элементов и порядка их выбора. Более удобен другой метод – ортонормальный базис главных компонент (ОБГК) [4]. Обозначая исходный набор случайных величин как вектор ξ , а ОБГК как вектор g , можно записать линейное преобразование от ξ к g в виде

$$g = \Lambda_q^{-1/2} P_q^T (\xi - \mu),$$

где $\mu = E(\xi)$, Λ_q – диагональная матрица составленная из ненулевых собственных чисел матрицы Σ , расположенных в порядке неввозрастания, P_q – матрица, образованная из собственных векторов матрицы Σ , соответствующих ненулевым собственным числам и взятым в том же порядке, что в матрице Λ_q .

Важно отметить, что описанная модель в равной мере применима как в теоретическом, так и в выборочном случаях. В последнем случае для выборки наблюдений объема n над p -мерной случайной величиной всегда можно построить эмпирическую функцию распределения $F_n^*(x)$, которую затем можно рассматривать в качестве теоретической функции распределения некоторой дискретной случайной величины ξ^* . Вместо величины ξ_0 выбирается вектор из единиц $l1_n$. Пространство $L_2(\xi^*)$ при этом оказывается подпространством \mathbf{R}^n , являющимся линейной оболочкой ортогональных дополнений к проекциям элементов выборки x_i на ξ_0 .

Такой подход позволяет в выборочном случае обходиться без каких-либо аппроксимаций выборочных распределений, тем более, что и в теоретическом случае вид распределений случайных величин оказывается не существенным. Распределения могут быть заданы, например, с помощью имитационной модели и необходимо уметь вычислять только моменты первых двух порядков. Наиболее существенным следствием такого подхода является возможность исследовать свойства различных процедур анализа модели с помощью методов бутстрэпа и аналогичных им. Другим следствием является применимость описанной модели к эмпирическим данным, не имеющим очевидной статистической природы, т. к. формальное построение эмпирической функции распределения требует наличия только исходной матрицы наблюдений.

3 Некоторые геометрические соотношения

Пусть H – евклидово пространство со скалярным произведением $\langle x, y \rangle$ и нормой $\|x\| = \langle x, x \rangle^{1/2}$.

Для всякой системы векторов $S = \{x_1, \dots, x_l\}$ из H линейная оболочка $\mathcal{M}(S)$ является подпространством H и для него можно построить ортонормированный базис (ОБ). Пусть кортеж $P_S = \{p_1, \dots, p_k\}$ образует некоторый ОБ в $\mathcal{M}(S)$ и $k = \dim \mathcal{M}(S)$. Очевидно, $\mathcal{M}(S)$ изоморфно k -мерному вещественному евклидову пространству E_k , т. к. для любого $x \in \mathcal{M}(S)$ можно указать единственный вектор $a_x \in \mathbf{R}^k$, такой что $x = P_S a_x$.

Обозначим единичную сферу в $\mathcal{M}(S)$ через

$$D_S = \{x \in \mathcal{M}(S) : \|x\| = 1\} = \{a_x \in \mathbf{R}^k : \|a_x\| = 1\},$$

где $\|a_x\|^2 = a_x^T a_x$. Для любых $x, y \in H$ угол α_{xy} между ними определяется через

$$\cos \alpha_{xy} = \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|} = \frac{\|\text{Pr}_x y\|}{\|y\|}$$

(если $x, y \in D_S$, то $\cos \alpha_{xy} = \langle x, y \rangle$).

Можно определить угол α_{yS} между некоторым $y \in H$ и $\mathcal{M}(S) \subset H$ через соответствующий \cos_{yS} как решение экстремальной задачи:

$$\cos \alpha_{yS} = \sup_{x \in \mathcal{M}(S)} \cos \alpha_{xy} = \sup_{x \in D_S} \cos \alpha_{xy} = \frac{\|y_S\|}{\|y\|},$$

где $y_S = \text{Pr}_{\mathcal{M}}(S) = \text{Pr}_{D_S} y$ – проекция y на $\mathcal{M}(S)$ или D_S

Более общим является понятие угла α_{12} между двумя подпространствами $\mathcal{M}_1, \mathcal{M}_2 \subset H$:

$$\cos \alpha_{12} = \sup_{x \in \mathcal{M}_1} \sup_{y \in \mathcal{M}_2} \cos \alpha_{xy}.$$

Если P_1 и P_2 определяют ОБ (размерности, $\dim \mathcal{M}_1 = k_1$ соответственно, k_1 и k_2) в \mathcal{M}_1 и \mathcal{M}_2 , а a_x и a_y являются представлениями x и y в этих ОБ, то $\cos \alpha_{xy} = a_x^T Q a_y$, причем $\|a_x\| = \|a_y\| = 1$, а $Q = P_1^T P_2$ – матрица, составленная из косинусов углов между векторами соответствующих ОБ. Тогда

$$\cos \alpha_{12} = \max_{(a_x, a_y)} a_x^T Q a_y$$

при $\|a_x\| = \|a_y\| = 1$.

Решение последней задачи хорошо известно: $\cos \alpha_{12} = |\rho_1|$, где ρ_1^2 является максимальным собственным числом одновременно матриц $Q^T Q$ и $Q Q^T$ и экстремум достигается на соответствующих собственных векторах. Не останавливаясь подробно на интерпретации остальных собственных чисел ρ_2, ρ_3, \dots и соответствующих им пар собственных векторов, поскольку она аналогична та-ковой в методе канонических корреляций и переменных, рассмотрим обобщение этого подхода.

4 Центроид пучка подпространств

Пусть $\mathcal{M}_0, \mathcal{M}_1, \dots, \mathcal{M}_m$ – подпространства H размерности, соответственно k_o, k_1, \dots, k_m , а W_1, \dots, W_m – весовые множители и $W_i \geq 0$. Без потери общности можно считать $\sum_{i=1}^m W_i = 1$. Рассмотрим функционал $J_W(x) = \sum_{i=1}^m W_i \cos^2 \alpha_{xi}$, где α_{xi} – угол между $x \in \mathcal{M}_0$ и подпространством \mathcal{M}_i ($i \neq 0$). Тогда $x_0 = \arg \sup_{x \in \mathcal{M}_0} J_W(x)$ является вектором из \mathcal{M}_0 , наилучшим образом (в смысле величины функционала $J_W(x)$) приближающим одновременно весь набор подпространств $\{\mathcal{M}_i, i = 1, \dots, m\}$.

Можно показать, что x_0 является собственным вектором матрицы QWQ^T , соответствующим максимальному собственному числу $\rho_1^2 = J_W(x_0)$, где

$$W = \begin{pmatrix} W_1 I_{k_1} & 0 & \dots & 0 \\ 0 & W_2 I_{k_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_m I_{k_m} \end{pmatrix},$$

$Q = P_0^T P$, $P = (P_1 : P_2 : \dots : P_m)$ – кортеж, объединяющий ОБ подпространств $\mathcal{M}_1, \dots, \mathcal{M}_m$.

В каждом из \mathcal{M}_i ($i = 1, \dots, m$) при этом определяется соответственно векторы y_i , ближайшие к x_0 (проекции x_0 по \mathcal{M}_i), на которых достигается экстремум функционала $J_W(x)$. Оказывается возможным содержательно проинтерпретировать также и собственные векторы, соответствующие меньшим собственным числам ρ_2^2, ρ_3^2 , и т. д.

Заметим, что этот подход геометрически очень близок центроидному методу факторного анализа.

Рассмотрим некоторые частные случаи данной задачи.

a). Если $\mathcal{M}_0 \supset \cup_{i=1}^m \mathcal{M}_i$, то, очевидно, можно считать $\mathcal{M}_o = H$, и мы получим задачу построения вектора из H , ближайшего к пучку подпространств;

б). При $m = 2$ получается задача, являющаяся некоторым обобщением канонического анализа, т. е. нахождения набора углов между подпространствами \mathcal{M}_1 и \mathcal{M}_2 .

в). При $k_i = 1$ ($i = 1, \dots, m$) мы приходим к обобщению (за счет наличия весов W_i) процедуры построения ОБГК. При этом появляется возможность выбором набора $\{W_i\}$ получать ОБГК для ковариационной, корреляционной и других матриц. Отметим, что при фиксированном наборе $\{W_i\}$ такой обобщенный ОБГК является единственным с точностью до векторов, соответствующих кратным собственным числам, и при этом имеется хорошо разработанная система интерпретации главных компонент как со статистической, так и с геометрической точек зрения.

5 Метод согласованных вращений

Одним из приемов интерпретации главных компонент является ортогональное преобразование некоторого их подмножества (вращение или "ротация"). В зависимости от способа преобразования получаются как различные интерпретации факторов простой структуры, так и другие схемы факторного анализа. Однако, традиционно все эти схемы применяются только к одной выборке или теоретической модели.

По мере расширения статистических исследований, с применением ЭВМ и пакетов прикладных программ обработки многомерных данных все чаще приходится сталкиваться с необходимостью совместной интерпретации результатов факторного анализа для нескольких теоретических или выборочных совокупностей. При этом факторные структуры могут быть получены с разным числом информативных компонент и по разным (перекрывающимся) подмножествам основного множества исходных признаков. Однако, если отдельные

подмножества, определяемые выбранными для каждой структуры компонентами, оказываются достаточно близкими друг к другу, то естественно попытаться построить подпространство, близкое ко всем подпространствам семейства. При этом во всех подпространствах семейства оказываются построенными согласованные друг с другом ортонормальные базисы.

Аналогичный прием возможно применить и для сопоставления структур различных классов в линейном дискриминантном анализе в случае плохообусловленных или даже вырожденных ковариационных матриц классов при малых объемах исходных классов [3].

В применении к методу "Гусеница" этот подход является фактически геометрическим обоснованием процедуры "многомерная гусеница" (см. раздел III.2).

Литература

- [1] Krzanowski W.J. Between-groups comparison of principal components // J. Amer. Stat. Assoc. V.74, 1979, pp.703–707, Correction: V.76, 1981, p.1022.
- [2] Баланина Н.Ю., Походзей Б.Б., Солнцев В.Н.. Сравнение главных компонент, построенных по различным выборкам данных // Заводская лаборатория. 1986. т.52. N 3. с.49-51.
- [3] Ермаков С.М., Походзей Б.Б., Солнцев В.Н.. Метод главных компонент в многомерном статистическом анализе (Обзор) // Заводская лаборатория. 1983. т.49. N 5. с.46-52.
- [4] Рао С.Р. Линейные статистические методы и их применения. М.: Наука, Физматгиз, 1968. – 548 с.