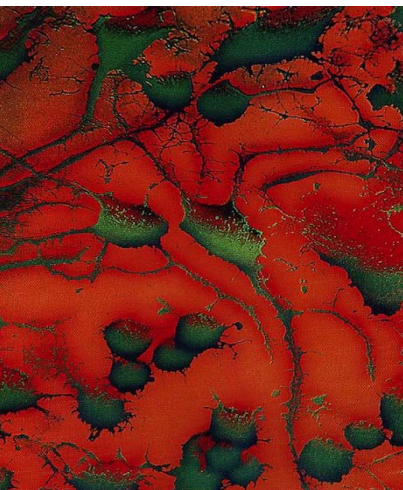
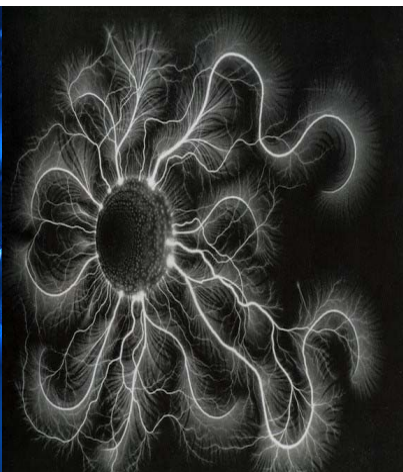


Пузыри в жидкости



Разряд в магнитном поле



Трещины в сухой глине

Нейроны в мозге

Н.В. Латыпова

**КОМПЬЮТЕРНАЯ ОБРАБОТКА
ДАННЫХ
ФРАКТАЛЫ**



Ижевск 2012

Министерство образования и науки Российской Федерации
ФГБОУ ВПО «Удмуртский государственный университет»
Кафедра математического анализа

Н. В. Латыпова

**КОМПЬЮТЕРНАЯ ОБРАБОТКА ДАННЫХ
ФРАКТАЛЫ**

Учебное пособие

Ижевск 2012

УДК 519.6я7
ББК 22.19с51я7
Л 278

Рекомендовано к изданию Учебно–методическим советом УдГУ

Н. В. Латыпова

Л 278 Компьютерная обработка данных. Фракталы: учеб. пособие. Ижевск: Изд-во «Удмуртский университет», 2012. 78 с., прил.

Предлагаемое учебное пособие посвящено изучению курса по компьютерной обработке данных с помощью нового современного аппарата — теории фракталов.

Пособие включает конспект лекций с основными понятиями курса, лабораторный практикум и упражнения для самостоятельной работы. В Приложении даются иллюстрации некоторых фракталов. Может быть использовано для курсовых и выпускных квалификационных работ.

Предназначено для студентов математического факультета.

УДК 519.6я7
ББК 22.19с51я7

©Н. В. Латыпова, 2012
©Изд-во «Удмуртский университет», 2012

ОГЛАВЛЕНИЕ

Предисловие	4
Лекция 1. Введение в теорию фракталов	5
Лекция 2. Конструктивные фракталы	7
Лекция 3. Динамические фракталы	11
Лекции 4. Мера и размерность Хаусдорфа–Безиковича. Размерность подобия	16
Лекции 5. Фрактальная размерность	19
Лекция 6. Случайные фракталы	24
Лекция 7. Временные ряды, их фрактальность. Метод R/S-анализа Херста	27
Лекция 8. Методы определения фрактальной размерности для временной рядов	31
Лекция 9. Массовые фракталы	42
Лекция 10. Принцип Парето	44
Лекция 11. Методы исследования документальных информационных потоков	47
Лекция № 12. Фрактальные свойства информационного пространства	54
Задачи и упражнения	59
Лабораторный практикум	61
Исследовательские работы	62
Темы докладов и рефератов	63
Вопросы к зачету	64
Список рекомендуемой литературы	65
Приложение	66

Предисловие

Теория фракталов — пример развития нового направления науки, основанного как на достижениях в весьма абстрактных областях математики, так и на новом взгляде на давно известный эмпирический материал, который до создания адекватных моделей не поддавался даже научному описанию и интерпретации. В данном пособии предложено применение фрактального подхода к компьютерной обработке данных.

В отличие от традиционного формата многих математических курсов теорема — доказательство — пример — задача, большую роль при изучении фракталов предполагается уделить компьютерному моделированию. Компьютер позволяет сейчас увидеть потрясающие и впечатляющие картинки, связанные с фракталами. Курс дает возможность попробовать придумать свои фрактальные множества и структуры. В процессе изучения теоретических разделов курса используются новые образовательные технологии обучения — презентации с использованием ноутбука, медиапроектора и экрана. Содержание курса постоянно обновляется в соответствии с последними достижениями в данной области. Теоретическая часть курса активно иллюстрируется прикладными задачами.

Для успешного усвоения курса достаточно знаний по курсу математического и комплексного анализа, теории вероятностей, численных методов и линейной алгебры.

Для самостоятельной работы студента предложены задачи и упражнения, а также тематика исследовательских работ, которые направлены на приобретение вычислительного опыта решения конкретных прикладных задач. Для углубления и расширения лекционного материала даются лабораторный практикум и темы докладов и рефератов.

Предназначено для студентов математического факультета. Но может быть полезно и интересно всем, кто интересуется как вычислительной математикой, так и тем, как устроен мир.

Лекция № 1. Введение в теорию фракталов

Понятия фрактал и фрактальная геометрия, появившиеся в конце 70-х, с середины 80-х прочно вошли в обиход математиков и программистов. В 1975 году Бенуа Мандельбротом было предложено определение фрактала. Согласно Мандельброту, слово фрактал происходит от латинских слов *fractus* — дробный и *frangere* — ломать, что отражает суть фрактала, как «изломанного», нерегулярного множества. Мандельброт дал строгое математическое определение фрактала, как множества, хаусдорфова размерность которого, строго больше топологической размерности. Наряду с этим определением существует еще одно определение фрактала. Фрактал — структура, состоящая из частей, которые в каком-то смысле подобны целому. Это определение содержит существенный отличительный признак — фрактал выглядит одинаково, в каком бы масштабе мы его не наблюдали.

Структуры, похожие на фракталы, можно обнаружить в окружающей нас природе: границы облаков, границы морских побережий, турбулентные потоки в жидкостях, трещины в некоторых породах, зимние узоры на стекле и т. п. Но эти объекты невозможно описать с помощью классической геометрии. Как построить модель горного хребта или кроны дерева в терминах геометрии? Как описать всю сложность системы кровообращения, состоящей из множества капилляров и сосудов и доставляющей кровь к каждой клеточке человеческого тела? Как представить почки и легкие, напоминающие по структуре деревья с ветвистой кроной? Какая математика отвечает за ритмы сердца и головного мозга, наблюдаемые на электрокардиограмме и энцефалограмме? Можно ли математически описать внезапное возникновение волны паники на финансовых рынках или построить математическую модель социального поведения?

Одним из основных свойств фракталов является самоподоб-

бие (симметрия) — инвариантность относительно параллельных переносов и скейлинга (изменения масштаба). В самом простом случае небольшая часть фрактала содержит информацию о всем фрактале. Так при увеличении маленькие фрагменты фрактала получаются очень похожими на большие. В идеальном случае такое самоподобие приводит к тому, что фрактальный объект оказывается инвариантным относительно растяжений, то есть ему присуща дилатационная симметрия, предполагающая неизменность основных геометрических особенностей фрактала при изменении масштаба. Большинство природных объектов демонстрирует самоподобие — главный организующий принцип фракталов. Поэтому фракталы будут напоминать друг друга, независимо от используемой шкалы.

Фракталы делятся на два больших класса: конструктивные и динамические. С другой стороны, фракталы по способу задания подразделяются на группы: геометрические, алгебраические, системы итерированных функций, стохастические (случайные), массовые фракталы.

Понятие «фрактал» уже доказало свою пользу в ряде прикладных областей. Например, фрактальное броуновское движение позволяет моделировать лесные и горные ландшафты, флуктуации уровней рек и биение сердца. Обладая большой объяснительной силой, теория фракталов стимулирует развитие естествознания. Применение теории фракталов позволяет вскрыть огромные, неиспользуемые ранее резервы, и применить их в области физических и технических приложений (например, исследование волновых явлений в неоднородных средах, распознавание образов и создание новых сред для передачи информации, задачи искусственного интеллекта и т. д.), в естествознании (фрактальная обработка медицинской информации, задачи классификации и кластеризации земных покровов и объектов, моделирование каскадных водопадов и турбулентных процессов, определяющих погоду и т. д.) и исследование социальных процессов (описание ситуаций на финансовых

рынках, построение математических моделей социального поведения). Впечатляюще красивые картинки фрактальных множеств могут стимулировать любого художника или дизайнера.

Лекция № 2. Конструктивные фракталы

Рассмотрим примеры построения классических конструктивных фракталов. Начнем с *множества Кантора*.

Георг Кантор (1845–1918) явился одним из основателей теории множеств. Он также придумал один из старейших фракталов — множество Кантора (описано им в 1883). На Западе подобные множества называют иногда пылью Кантора. Это множество хорошо известно из курса функционального анализа как пример множества нулевой меры Лебега, чья мощность равна мощности континуума $[0, 1]$. Фрактальные свойства пыли Кантора имеют огромное значение, особенно учитывая тот факт, что многие известные фракталы являются близкими родственниками этого фрактала. Способ построения этого множества следующий. Из исходного отрезка единичной длины выбрасывается интервал $\left(\frac{1}{3}, \frac{2}{3}\right)$. Далее из каждого оставшегося отрезка выбрасываем средние трети и т. д. (см. рис. 1 Приложения). В пределе получим фрактал Кантора. На первом шаге имеем два отрезка, длины $\frac{1}{3}$ каждый. На втором шаге имеем $2^2 = 4$ отрезка длиной $\frac{1}{3^2}$. После трех шагов будет $2^3 = 8$ отрезков, и каждый имеет длину $\frac{1}{3^3}$. После n шагов получим 2^n отрезков, каждый длины $\frac{1}{3^n}$. Общая длина оставшихся отрезков равна $\left(\frac{2}{3}\right)^n$. Она стремится к нулю, когда $n \rightarrow \infty$. Это означает, что множество Кантора имеет меру Лебега (т. е., грубо говоря, общую длину), равную нулю, и нулевую топологическую размерность.

Свойства фрактала Кантора:

1⁰. Канторова пыль есть самоподобный фрактал размерности $d = \frac{\ln 2}{\ln 3} \approx 0.6309\dots$

2⁰. Канторова пыль не содержит интервалов положительной длины (это следует из построения).

3⁰. Сумма длин интервалов, удаленных при построении в точности равна 1.

4⁰. Мощность множество Кантора равна мощности континуума $[0, 1]$.

5⁰. Классическая канторова пыль представляет собой пример компактного, совершенного и вполне разрывного множества. (Множество называется компактным, если оно замкнуто и ограничено. Множество называется совершенным, если оно замкнуто и не содержит изолированных точек. Говорят, что множество вполне разрывно, если все компоненты этого множества есть одиночные точки.)

Следующий пример построения конструктивного фрактала — это *кривая Коха*. Эту кривую построил Хельге фон Кох в 1904 году, как пример кривой, которая нигде не имеет касательной, т. е. нигде не дифференцируема, хотя всюду непрерывна. Эта кривая самоподобна: каждая ее часть является копией целого. К тому же, каждая часть кривой имеет бесконечную длину. Процесс ее построения выглядит следующим образом: берем единичный отрезок, разделяем на три равные части и заменяем серединный интервал равносторонним треугольником без этого сегмента. В результате образуется ломаная, состоящая из четырех звеньев длины $\frac{1}{3}$. На следующем шаге повторяем операцию для каждого из четырех получившихся звеньев и т.д. (см. рис. 2 Приложения). Предельная кривая и есть кривая Коха. В своем исследовании, Мандельброт много экспериментировал с кривыми Коха, и получил фигуры, такие как Острова Коха, Кресты Коха, Снежинки Коха и даже трехмерные представления кривой Коха, используя тетраэдр и прибавляя

меньшие по размерам тетраэдры к каждой его грани.

Рассмотрим еще один пример простого самоподобного фрактала — *ковер Серпинского*, построенный в 1915 году польским математиком Вацлавом Серпинским. Для его построения берется квадрат с длиной стороны, равной единице. Затем каждая из сторон квадрата делится на три равные части, а весь квадрат, соответственно, на девять одинаковых квадратиков со стороной равной $\frac{1}{3}$. Из полученной фигуры вырезается центральный квадрат. Затем такой же процедуре подвергается каждый из 8 оставшихся квадратиков и т.д.; этот процесс продолжается неограниченно. Пересечение полученных на каждом шаге объединений оставшихся квадратов называется ковром Серпинского (рис. 3).

Аналогичным образом можно построить *решето Серпинского*, используя вместо квадрата треугольник. Процесс построения начинается с равностороннего треугольника. На первом шаге из центра удаляется перевернутый треугольник, который меньше первоначального в два раза. На втором шаге удаляется три перевернутых треугольника из трех оставшихся треугольников. Продолжая этот процесс, на n -ом шаге удаляем перевернутых треугольников из центров 3^{n-1} оставшихся треугольников. Продолжая этот процесс, получаем объект состоящий из одних только дырок. Это и есть решето Серпинского (см. рис. 4).

Существуют объемные представления решета Серпинского — *пирамида Серпинского* и ковра Серпинского — *губка Менгера* (рис. 5). При построении трехмерной версии ковра используют обыкновенный куб, из которого вырезают маленькие кубы. В своей пространственной форме, губка Менгера преобразуется в систему сквозных форм, в которой каждый сквозной элемент постоянно заменяется себе подобным. Эта структура очень похожа на разрез костной ткани.

При построении конструктивных фракталов используют L-

системы и аффинные преобразования.

Понятие L-систем, тесно связанное с самоподобными фракталами, появилось только в 1968 году благодаря Аристриду Линденмайеру. Изначально L-системы были введены при изучении формальных языков, а также использовались в биологических моделях селекции. С их помощью можно строить многие известные самоподобные фракталы, включая снежинку Коха и ковер Серпинского. Некоторые другие классические построения, например кривые Пеано, Гильберта, Серпинского, также укладываются в эту схему. И конечно, L-системы открывают путь к построению новых фракталов, что и послужило причиной их широкого применения в компьютерной графике для построения фрактальных деревьев и растений. Для графической реализации L-систем в качестве подсистемы вывода используется так называемая тертл-графика (turtle — черепаха). При этом точка (черепашка) движется по экрану дискретными шагами, как правило, прочерчивая свой след, но при необходимости может перемещаться без рисования. В нашем распоряжении имеются три параметра (x, y, α) , где (x, y) — координаты черепашки, α — направление, в котором она смотрит. Черепашка обучена интерпретировать и выполнять последовательность команд, задаваемых кодовым словом, буквы которого читаются слева направо. Кодовое слово представляет собой результат работы L-системы и может включать следующие буквы: F — переместиться вперед на один шаг, прорисовывая след, b — переместиться вперед на один шаг, не прорисовывая след, $[$ — открыть ветвь, $]$ — закрыть ветвь, $+$ — увеличить угол α на величину θ , $-$ — уменьшить угол α на величину θ . Размер шага и величина приращения по углу θ задаются заранее и остаются неизменными для всех перемещений черепашки. Если начальное направление движения α (угол, отсчитываемый от положительного направления оси X) не указано, то полагают α равным нулю.

Рассмотрим пример L-системы, соответствующей снежинке

Коха. Она задается следующим образом:

$$\theta = \frac{\pi}{3}; \quad \text{axiom} = F + +F + +F; \quad \text{newf} = F - F + +F - F.$$

Графическое представление аксиомы $F + +F + +F$ есть равносторонний треугольник. Черепашка делает один шаг вперед, затем угол α увеличивается на $\frac{2\pi}{3}$ и черепашка делает еще один шаг вперед, угол α снова увеличивается на $\frac{2\pi}{3}$ и черепашка делает еще шаг. На первом шаге каждая буква F в словесно-инициаторе $F + +F + +F$ заменяется на $F - F + +F - F$:

$$(F - F + +F - F) + +(F - F + +F - F) + +(F - F + +F - F).$$

Убирая скобки, получаем:

$$F - F + +F - F + +F - F + +F - F + +F - F + +F - F.$$

Продолжая процесс, получим в пределе снежинку Коха.

Общая схема построения конструктивных фракталов использует задание *основы* — фигура, к которой будут применяться преобразования, и *фрагмента*, повторяющегося при каждом уменьшении масштаба. Вершины для основы и фрагмента должны быть заданы явно. В лекции использованы материалы [5,9].

Лекция № 3. Динамические фракталы

В случае динамических фракталов речь идет о некоторых классах двумерных нелинейных отображений, которые относятся к дискретным динамическим системам. Анализ таких систем сопряжен со значительными трудностями. Рассмотрим наиболее изученный пример нелинейного отображения, которое можно представить в виде одномерного комплексного отображения, сопоставляющего одному комплексному числу $z_n = x_n + iy_n$ другое комплексное число $z_{n+1} = x_{n+1} + iy_{n+1}$ по итерационному правилу $z_{n+1} = f(z_n)$, где $f(z)$ — некоторая

нелинейная функция переменной z , n — номер итерации. Комплексные отображения такого рода порождают популярные в последнее время фракталы Жюлиа, Мандельброта и Ньютона.

Неподвижной точкой \tilde{z} отображения $z_{n+1} = f(z_n)$ называется корень уравнения $f(z_n) = z_n$ (его также называют неподвижной точкой функции). Если, стартовав в непосредственной близости от неподвижной точки, мы будем в процессе итераций к ней неограниченно приближаться, то такая неподвижная точка называется *притягивающей*. Соответственно, стартовав в непосредственной близости от *отталкивающей неподвижной точки*, мы будем от нее удаляться. *Нейтральная неподвижная точка* характеризуется тем, что, стартовав в достаточно малой ее окрестности, мы будем все время находиться в этой окрестности, не приближаясь и не удаляясь от точки. Чтобы определить характер неподвижной точки \tilde{z} отображения, надо вычислить производную $f'(\tilde{z})$. Имеет место следующая теорема.

Теорема. *Если $|f'(\tilde{z})| < 1$, то точка \tilde{z} является притягивающей, если $|f'(\tilde{z})| > 1$, то отталкивающей, и, наконец, если $|f'(\tilde{z})| = 1$, то неподвижная точка является нейтральной.*

При анализе отображений наряду с неподвижными точками важную роль играют так называемые периодические точки и циклы, состоящие из этих периодических точек.

Аттрактором на комплексной плоскости называется точка (или точки), к которой сходится процесс итераций $z_{n+1} = f(z_n)$ при $n \rightarrow \infty$. В качестве такого аттрактора может выступать притягивающая неподвижная точка или притягивающий цикл определенного периода. Иногда таких аттракторов может быть несколько, они также могут состоять из бесчисленного количества точек и представлять собой непрерывную линию или какое-нибудь другое (например, канторовское) множество. Если же в процессе итераций изображающая их точка уходит на бесконечность, то аттрактором такого процесса считается бесконечно удаленная точка.

Множества Жюлиа дают наиболее впечатляющие иллюстрации того, как простой процесс может привести к чрезвычайно запутанным множествам. Итерации (отображения) вида $z_{n+1} = f(z_n)$, $n = 0, 1, 2, \dots$ с простой функцией на комплексной плоскости вызывают появление различных экзотических фракталов. Наиболее изученным примером отображений такого вида является квадратичное комплексное отображение

$$z_{n+1} = f(z_n) = z_n^2 + c,$$

где $c = a + ib$ — ненулевая комплексная константа. После разделения вещественной и мнимой частей отображение запишется в виде:

$$x_{n+1} = x_n^2 - y_n^2 + a, \quad y_{n+1} = 2x_n y_n + b.$$

Такое отображение дает множество фракталов, соответствующих множеству Жюлиа — $J(a, b)$. Множество Жюлиа функции f — это граница множества тех точек z , которые при итерировании $f(z)$ стремятся к бесконечности. Жюлиа подробно исследовал свойства таких множеств.

Введем в рассмотрение функцию

$$f^{\{n\}}(z) = f(\dots(f(f(z)))\dots)$$

и зададим последовательность $z_n = f^{\{n\}}(z)$ для $n = 1, 2, 3, \dots$

Теорема 1. Пусть $|c| < 2$, $z \in C$ и $z_n = f^{\{n\}}(z)$ для $n = 1, 2, 3, \dots$. Если существует такое n_0 , что $|z_{n_0}| \geq 2$, то имеет место $\lim_{n \rightarrow \infty} z_n = \infty$, т. е. последовательность итераций $\{f^{\{n\}}(z)\}_{n=1}^{\infty}$ стремится к бесконечности и z не принадлежит множеству Жюлиа $J(f)$.

Как следует из приводимой теоремы (см. [5]), в случае $|c| < 2$ можно прекратить вычисление последовательности итераций, как только величины достигают значения 2 по модулю. Последовательность итераций таких точек гарантированно стремится к бесконечности.

Множества Жюлиа функции $z^2 + c$ обладают большим разнообразием (см. рис. 7–8 Приложения). Для каждого значения c получаем новое изображение, тем не менее существует всего два типа изучаемого множества. Каждое множество Жюлиа либо связно, либо вполне несвязно. Индикатором для двух типов множеств Жюлиа служит множество Мандельброта. Каждая точка в множестве Мандельброта представляет значение c , для которого J связно. Каждая точка из дополнения к множеству Мандельброта представляет значение c , для которого множество J вполне не связно. Таким образом, множество Мандельброта (см. рис. 9) — это точки, расположенные в черной области. Это множество симметрично относительно оси a . Граница множества Мандельброта представляет собой фрактал Мандельброта.

Множество Мандельброта M для полинома $f(z) = z^2 + c$ есть множество всех $c \in C$, для которых последовательность итераций точки $z = 0$ ограничена, то есть

$$M = \left\{ c \in C : \left\{ f^{\{n\}}(0) \right\}_{n=0}^{\infty} \text{ограничена} \right\}.$$

Выбор точки $z = 0$ в качестве начальной связан с тем обстоятельством, что точка 0 — единственная точка, в которой производная функции $f(z) = z^2 + c$ обращается в нуль.

Определение множества Мандельброта, приведенное выше, может быть использовано для написания программы, определяющей принадлежность точки множеству M . Задача проверки последовательности итераций на ограниченность упрощается при использовании следующей теоремы (см. [5]).

Теорема 2. *Если $|c| > 2$ и $|z| \geq |c|$, то последовательность итераций устремляется к ∞ . В частности из этого следует, что точка c не принадлежит M .*

Объединяя полученный результат с теоремой 1, получаем, что проверять нужно точки $|c| \leq 2$. Причем в случае $|c| < 2$, если последовательность итераций достигает состояния, когда

ее величина превосходит 2, то это означает, что она стремится к бесконечности, и, следовательно, проверяемая точка не принадлежит M . Выбранные при таком условии c , как и в случае $|c| > 2$ всегда соответствуют фракталу Жюлиа, называемому пылью Фату. Точка $c = -2$ — единственная точка окружности, которая принадлежит множеству Мандельброта.

Рассмотрим еще один вид динамических фракталов — это фракталы Ньютона. Для их построения понадобится метод Ньютона–Рафсона, который для функции действительной переменной часто называют методом касательных. Метод определяется с помощью итерационного процесса $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$. В курсе «Численные методы» была доказана сходимость метода Ньютона к единственной неподвижной точке независимо от выбора начальной точки $x_0 \in [a, b]$. Области с фрактальными границами появляются при приближенном нахождении корней нелинейного уравнения $f(z) = 0$ алгоритмом Ньютона–Рафсона на комплексной плоскости. Стартовав в непосредственной близости от значения корня уравнения $f(z) = 0$, и, используя итерационный алгоритм:

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)}, n = 0, 1, 2, \dots,$$

мы получаем последовательность комплексных чисел, быстро сходящуюся к этому корню. При этом возникает вполне закономерный вопрос, а что будет, если начальная точка z_0 в плоскости комплексных чисел выбрана не вблизи от корня, а произвольным образом? И мы приходим к так называемой проблеме Кэли, которую в 1977 году удалось разрешить экспериментально (с помощью компьютера) американскому математику Джону Хаббарду (John Hubbard) и его студентам, которым он преподавал математику в Парижском университете Орсэй.

Рассмотрим простейшее уравнение третьей степени

$$z^3 - 1 = 0.$$

В этом случае имеем три кубических корня из 1, а именно, $w_1 = 1$, $w_2 = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$ и $w_3 = -\frac{1}{2} - i\frac{\sqrt{3}}{2}$. Нанесенные на комплексную плоскость, они образуют равносторонний треугольник, центр которого совпадает с началом координат. Эти корни действительно являются притягивающими неподвижными точками отображения $z_{n+1} = z_n - \frac{z_n^3 - 1}{3z_n^2}$. Так какова же геометрия областей притяжения корней кубического уравнения? По аналогии с уравнением второй степени можно предположить, что ответ выглядит примерно так, как изображено на рис. 10 Приложения. Именно такую картину первоначально представлял себе Хаббард (и многие другие до него). Однако такое предположение оказалось неверным. Более скрупулезное компьютерное исследование выявило, что геометрия границ областей притяжения имеет гораздо более сложную форму. Так, если раскрашивать разные области притяжения различным цветом, то мы получим картинку, изображенную на рис. 11 Приложения. Из этого рисунка видно, что граница областей притяжения состоит из сильно переплетенных самоподобных структур. Оказывается, что на границе между любыми двумя цветами всегда расположена гирлянда островков третьего цвета. Границы этих островков, в свою очередь, состоят из гирлянд островков меньшего размера соответствующего дополнительного цвета и т. д. Непрерывно уменьшаясь в размерах, детали границы постоянно воспроизводят сами себя. В результате оказывается, что каждая точка такой фрактальной границы соседствует сразу с тремя областями притяжения. В лекции использованы материалы [5,9].

Лекция № 4. Мера и размерность Хаусдорфа–Безиковича. Размерность подобия

В 1919 году Ф. Хаусдорф определил дробную s -меру для $\forall s \geq 0$ и на этой основе каждому множеству в евклидовом

пространстве сопоставил число, названное им метрической размерностью. Он же привел примеры с дробной размерностью (канторово множество, кривая Коха и др.). Идеи Хаусдорфа, не опубликовавшего больше ни одной работы в этом направлении, были развиты А.С.Безиковичем, который длительное время был автором или соавтором практически всех работ по данной тематике. В последующие годы размерность Хаусдорфа–Безиковича получила применение в некоторых разделах математики, но только благодаря научным трудам Б.Мандельброта стала настолько популярна в последние годы.

Для строгого определения размерности Хаусдорфа–Безиковича нам потребуется понятие хаусдорфовой меры. Если $U \subset \mathbb{R}^n$, то диаметр U определяется как

$$\text{diam } U = \sup\{\|x - y\| : \forall x, y \in U\},$$

т.е. наибольшее расстояние между любой парой точек этого множества. Если $F \subset \bigcup_i U_i$ с $0 < \text{diam } U_i \leq \delta \forall i$, то $\{U_i\}$ называют δ -покрытием множества F . Пусть $F \subset \mathbb{R}^n$ и $s \geq 0$. Для $\forall \delta > 0$ определяем

$$H_\delta^s(F) = \inf \left\{ \sum_i (\text{diam } U_i)^s : \{U_i\} - \delta\text{-покрытие } F \right\}.$$

Величину $H^s(F) = \lim_{\delta \rightarrow 0} H_\delta^s(F)$ называют s -мерной хаусдорфовой мерой множества F .

Масштабное свойство. Если $F \subset \mathbb{R}^n$, $\lambda F = \{\lambda x : x \in F\}$, где $\lambda > 0$, то $H^s(\lambda F) = \lambda^s H^s(F)$.

Хаусдорфова размерность определяется как

$$D_H = \dim_H F = \inf\{s : H^s(F) = 0\} = \sup\{s : H^s(F) = \infty\},$$

$$\text{т.е. } H^s(F) = \begin{cases} \infty, & s < \dim_H F, \\ 0, & s > \dim_H F. \end{cases}$$

Возьмем теперь произвольный набор преобразований подобия плоскости $f_1(x), f_2(x), \dots, f_m(x)$ с коэффициентами подобия $k_i < 1$, для любого $i = 1, 2, \dots, m$.

Теорема 3. *Существует единственное непустое компактное множество F , для которого $F = \bigcup_{i=1}^m f_i(F)$. (См. [13]).*

Множество F называется *инвариантным множеством*, или *аттрактором*, системы преобразований $\{f_i(x)\}$. Теорема 3 позволяет строить новые фрактальные объекты. Укажем способ построения аттрактора с помощью последовательных приближений. С этой целью для любого множества A определим множество $\Phi(A) = \bigcup_{i=1}^m f_i(A)$. Для аттрактора имеем равенство $F = \Phi(F)$. В качестве нулевого приближения выберем компактное множество F_0 такое, что $\Phi(F_0) \subset F_0$, и положим $F_1 = \Phi(F_0), F_2 = \Phi(F_1), \dots$. Тогда последовательность $F_0, F_1, \dots, F_n, \dots$ будет убывающей: $F_0 \supset F_1 \supset F_2 \supset \dots$ и пересечение $F = \bigcap F_n$ есть искомым аттрактор.

Итак, с набором преобразований подобия $\{f_i(x)\}$ можно единственным образом связать компактное множество F — аттрактор этой системы. В некоторых случаях аттрактор является фракталом. Всегда ли это так? Можно ли определить размерность аттрактора заранее, зная только набор коэффициентов подобия k_1, k_2, \dots, k_m ?

Определим неотрицательное число s как решение уравнения $k_1^s + k_2^s + \dots + k_m^s = 1$. Единственное решение s данного уравнения называется *размерностью подобия* аттрактора F : $s = s(F)$.

Для классических фракталов вычислить число s не представляет труда, и при этом мы всегда получаем равенство $s(F) = \dim_H F$. Однако в общем случае это равенство не выполняется. Приведем одно достаточное условие, обеспечивающее совпадение хаусдорфовой размерности и размерности подобия. Скажем, что выполнено *условие Морана*, если существу-

ет непустое открытое множество V такое, что $\bigcup_{i=1}^m f_i(V) \subset V$, $f_i(V) \cap f_j(V) = \emptyset$, $i \neq j$.

Теорема 4. *Если выполнено условие Морана, то размерность подобия аттрактора совпадает с его хаусдорфовой размерностью, т. е. $s(F) = \dim_H(F)$. (См. [13]).*

Для классических фракталов условие Морана легко проверяется. Так, для решета Серпинского множество V — внутренность исходного правильного треугольника, для ковра — это внутренность исходного квадрата. В лекции использованы материалы [4,9,13].

Лекция № 5. Фрактальная размерность

По определению Мандельброта: *фрактал* — это множество, хаусдорфова (фрактальная) размерность которого больше топологической. *Топологическая размерность* — топологическое свойство фигур, не изменяющееся при любых деформациях, производимых без разрывов и склеиваний (при взаимно однозначных и непрерывных отображениях). Топологическая размерность всегда равна целому числу: для пустого множества — (-1) , для точки — 0 , для линии — 1 , для поверхности или фигуры — 2 , для тела — 3 и т. д.

В работе Льюиса Ричардсона (1881–1953), одного эксцентричного метеоролога, Мандельброт в 1961 году обнаружил формулу для определения длины западного Британского побережья и испано–португальской границы. Ричардсон заметил, что результаты сильно зависят от масштаба, который используется на карте. Карта с масштабом $1 \text{ см} — 100 \text{ км}$ менее подробная, чем туристическая карта с масштабом $1 \text{ см} — 1 \text{ км}$. Нетрудно убедиться, что на туристической карте побережье становится длиннее. Этот же феномен можно обнаружить, используя одну карту, но применяя каждый раз все более и более мелкие единицы измерения.

Полученные Ричардсоном значения лежат вблизи прямой линии в дважды логарифмической системе координат, удовлетворяющей уравнению: $\lg S \approx -0.22 \lg a + \lg S_1$, где S — длина побережья, S_1 — длина, при измерении которой использовались единицы измерения в 1 км, a — другие единицы измерения («линейка»). Число 0.22 определяет угол наклона прямой. Тогда $S \approx S_1 \left(\frac{1}{a}\right)^{0.22}$.

Поэтому, если a уменьшается в 32 раза, то S увеличивается примерно вдвое. Теоретически, если $a \rightarrow 0$, то $S \rightarrow \infty$. Но тогда западное побережье Британии имело бы бесконечную длину. Что в реальности не так. Поэтому криволинейность такой границы определяется степенью ее изгибания, т. е. некоторым числом. Мандельброт добавил 1 к углу наклона прямой и получил число, которое он назвал *фрактальной размерностью* побережья (границы).

Мы ввели понятие о фрактальной размерности береговой линии. Дадим теперь общее определение этой величины. Пусть d — обычная Евклидова размерность пространства, в которой находится наш фрактальный объект F . Покроем теперь этот объект целиком d -мерными «шарами» радиуса ε . Представим, что нам потребовалось для этого не менее, чем $N(\varepsilon)$ шаров. Тогда, если при достаточно малых ε величина $N(\varepsilon)$ меняется с ε по степенному закону $N(\varepsilon) \sim \frac{1}{\varepsilon^D}$, где $D \doteq \dim F$ называется *хаусдорфовой* или *фрактальной размерностью* этого объекта F . Формулу можно так же записать в виде

$$D = \lim_{\varepsilon \rightarrow 0} \frac{\ln N(\varepsilon)}{\ln \frac{1}{\varepsilon}}.$$

Это и служит нестрогим общим определением фрактальной размерности D . В соответствии с ним величина D является локальной характеристикой данного объекта.

Кроме фрактальной размерности, так же часто использу-

ется *клеточная* размерность, которая отличается от хаусдорфовой только тем, что при ее вычислении берутся не шары, а клетки (или кубы). Очевидно, если фрактал является подмножеством прямой, то вместо квадратов надо использовать отрезки. Если же фрактал — подмножество трехмерного пространства, то квадраты заменяются кубами. Главный недостаток этого определения заключается в том, что подсчитанное значение $N(\varepsilon)$ не является наименьшим числом клеток со стороной ε , покрывающих фрактал.

В строгой математической теории под метрической или фрактальной размерностью понимается размерность Хаусдорфа–Безиковича, которую мы уже рассматривали. Однако это довольно сложное абстрактное понятие и рассчитать численно хаусдорфову размерность для произвольного динамического фрактала не представляется возможным. По этой причине введены ряд определений фрактальной размерности, допускающих численный расчет. Все они являются некоторыми оценками размерности Хаусдорфа–Безиковича. Определения фрактальной размерности могут быть основаны только на метрических свойствах множеств, а могут учитывать вероятностную меру. Таким образом, все фрактальные размерности можно разделить на две группы: чисто метрические и вероятностно–метрические. Рассмотрим наиболее часто применяемые виды фрактальных размерностей (см. [1]).

Емкость множества D_c . Емкость — это упрощение понятия хаусдорфовой размерности. Другое название — клеточная размерность. Покроем множество $M \subset \mathbb{R}^n$ одинаковыми n -мерными кубиками с ребром ε . Тогда их количество $N \approx \frac{a}{\varepsilon^{D_c}}$, где $a = const$. Взяв логарифм и перейдя к пределу $\varepsilon \rightarrow 0$, получаем выражение для D_c , которое можно считать определением емкости множества: $D_c = \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log \varepsilon^{-1}}$. Для многообразий и, во многих случаях, для фракталов $D_c = D_H$. В общем случае имеет место неравенство: $D_c \geq D_H$, т. е. емкость служит оцен-

кой хаусдорфовой размерности сверху. Емкость является чисто метрической размерностью.

Информационная размерность D_I . Эта характеристика аналогична емкости, но учитывает вероятностную меру на исследуемом множестве. Покрываем множество точек в \mathbb{R}^n одинаковыми n -мерными кубиками с ребром ε и находим вероятность попадания точки в каждый кубик. Можно ввести энтропию распределения $H(\varepsilon) = -\sum_{i=1}^N P_i \log_2 P_i$, где P_i — вероятность обнаружения точки множества в i -м кубике. Информационная размерность определяется как $D_I = \lim_{\varepsilon \rightarrow 0} \frac{H(\varepsilon)}{\log \varepsilon^{-1}}$. Очевидно, что для равномерного распределения вероятности $\left(P_i = \frac{1}{N}\right)$ имеет место равенство $D_I = D_c$. При любом другом распределении энтропия будет меньше и, соответственно, $D_I < D_c$.

Корреляционная размерность D_{cor} . Этой характеристикой удобно пользоваться при анализе временных рядов (каких-либо экспериментальных данных). Пусть имеется последовательность из m точек, принадлежащих некоторому множеству в \mathbb{R}^n : $x_1, x_2, \dots, x_i, \dots, x_m$. Обозначим за $m_\varepsilon(x_i)$ число точек последовательности, попадающих в сферу радиуса ε с центром в точке x_i , принадлежащей той же последовательности. При $\varepsilon \rightarrow 0$ и $m \rightarrow \infty$ число $m_\varepsilon(x_i) \approx am\varepsilon^{D_i}$, где $a = const$. Величина D_i в общем случае зависит от выбора точки x_i . Рассмотрим среднее по всем точкам число $m_\varepsilon = \frac{1}{m} \sum_{i=1}^m m_\varepsilon(x_i)$ и, в пределе при $\varepsilon \rightarrow 0$, $m \rightarrow \infty$, представим его в виде $m_\varepsilon = am\varepsilon^{D_{cor}}$, где показатель D_{cor} может принимать нецелые значения. Тогда $\varepsilon^{D_{cor}} = \frac{1}{am^2} \sum_{i=1}^m m_\varepsilon(x_i)$. Число точек, попадающих в сферу радиуса ε с центром в точке x_i , можно представить как $m_\varepsilon(x_i) = \frac{1}{m} \sum_{j=1, j \neq i}^m \xi(\varepsilon - |x_i - x_j|)$,

где $\xi(\cdot)$ — функция Хевисайда (кусочно–постоянная функция, равная нулю для отрицательных значений аргумента, и единице — для неотрицательных). Получаем $\varepsilon^{D_{cor}} = \lim_{\varepsilon \rightarrow 0} \frac{1}{a} C(\varepsilon)$, где $C(\varepsilon) = \lim_{m \rightarrow \infty} \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1, j \neq i}^m \xi(\varepsilon - |x_i - x_j|)$. Функцию $C(\varepsilon)$ называют корреляционным интегралом. Корреляционная размерность множества определяется как $D_{cor} = \lim_{\varepsilon \rightarrow 0} \frac{\log C(\varepsilon)}{\log \varepsilon^{-1}}$.

Обобщенная размерность D_q . Можно обобщить размерности D_c , D_I , D_{cor} и ввести размерность порядка q , пользуясь обобщенной энтропией порядка q (энтропией Реньи): $H_q(\varepsilon) = \frac{1}{1-q} \log \left(\sum_{i=1}^N P_i^q \right)$, где P_i — вероятность обнаружения точки множества в i -м элементе покрытия. Тогда размерность порядка q есть $D_q = \lim_{\varepsilon \rightarrow 0} \frac{\log H_q(\varepsilon)}{\log \varepsilon^{-1}}$. Справедливы равенства $D_0 = D_c$, $D_1 = D_I$, $D_2 = D_{cor}$.

Размерность Ляпунова D_L . Фрактальную размерность аттрактора динамической системы в фазовом пространстве \mathbb{R}^n можно оценить с помощью спектра характеристических ляпуновских показателей (ЛХП). Такая оценка называется размерностью D_L . Выстроим спектр ЛХП в порядке убывания значений показателей: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Ляпуновская размерность задается следующей формулой Каплана–Йорка:

$$D_L = k + \frac{1}{|\lambda_{k+1}|} \sum_{i=1}^k \lambda_i,$$

где целая часть размерности k определяется из условий $\sum_{i=1}^k \lambda_i \geq 0$, $\sum_{i=1}^{k+1} \lambda_i < 0$. Таким образом, k — это число первых показателей в спектре ЛХП, сумма которых еще неотрицательна.

Заметим, что точный расчет какой-либо из размерностей в большинстве случаев невозможен. К тому же непонятно, какие из перечисленных размерностей ближе к хаусдорфовой и

лучше отражают фрактальную структуру исследуемого множества? Грассбергер и Прокачча показали, что в общем случае имеет место соотношение

$$D_{cor} \leq D_I \leq D_c.$$

Для большинства хаотических аттракторов значения всех перечисленных размерностей очень близки. При выборе, каким определением размерности лучше воспользоваться, обычно исходят из возможностей численных расчетов. При численном моделировании динамических систем наиболее удобно использовать размерность Ляпунова или емкость множества. Для оценки фрактальной размерности аттрактора, по экспериментальным данным, лучше всего подходит корреляционная размерность. Некоторые отличия между значениями различных размерностей, по-видимому, тоже несут определенную информацию о структуре исследуемого множества и могут служить важным инструментом для его детального изучения. Однако смысл этих тонких различий и их взаимосвязь со свойствами хаоса на сегодняшний день не достаточно исследованы. В лекции использованы материалы [1,9,10].

Лекция № 6. Случайные фракталы

Гауссовский процесс $X(t)$ называется *классическим броуновским движением (КБД)*, или *винеровским процессом на интервале $[a, b]$* , если он обладает следующими свойствами:

1. $X(a) = 0$, и функция $X(t)$ почти всегда непрерывна (т. е. почти все реализации процесса $X(t)$ непрерывны);

2. случайная величина $\Delta X = X(t_2) - X(t_1)$, $t_2 > t_1$, имеет гауссовское распределение с математическим ожиданием $M(\Delta X) = 0$ и дисперсией $D(\Delta X) = \sigma^2(t_2 - t_1)$, где σ — положительная константа, т. е.

$$P(\Delta X < x) = \frac{1}{\sqrt{2\pi\sigma^2(t_2 - t_1)}} \int_{-\infty}^x e^{-\frac{u^2}{2\sigma^2(t_2 - t_1)}} du.$$

Свойства классического броуновского движения

1⁰. Пусть $X(t)$ — КБД на интервале $[a; b]$. Тогда его приращения стационарны, и $D|X(t_2) - X(t_1)| = \sigma^2 |t_2 - t_1|$, $\forall t_1, t_2 \in [a; b]$.

2⁰. Пусть $X(t)$ — КБД, определенное на отрезке $[a; b]$, где $a \leq t_1 \leq t'_1 \leq t_2 \leq t'_2 \leq \dots \leq t_k \leq t'_k \leq b$. Тогда приращения $X(t'_1) - X(t_1), X(t'_2) - X(t_2), \dots, X(t'_k) - X(t_k)$ являются независимыми случайными величинами.

3⁰. КБД является марковским процессом.

4⁰. Пусть $X(t)$ — КБД на $[a; b]$. Тогда математическое ожидание модуля приращения равно

$$M(|X(t_2) - X(t_1)|) = \sigma \sqrt{\frac{2}{\pi} |t_2 - t_1|}.$$

5⁰. КБД $X(t)$ на отрезке $[a; b]$ недифференцируемо.

6⁰. Фрактальная размерность реализации КБД равна

$$D = 1,5.$$

7⁰. Приращения реализации КБД обладают свойством статистического самоподобия, т.е.

$$X(t + \Delta t) - X(t) \triangleq \frac{1}{\sqrt{r}} \left(X(t + r\Delta t) - X(t) \right)$$

для любого $r > 0$, где значок « \triangleq » означает, что две случайные величины имеют одинаковое распределение и, в частности, одни и те же математическое ожидание и дисперсию.

Классическое броуновское движение представляет собой хорошую модель марковских случайных фракталов, для которых условная вероятность того, что $X(t_2)$ достигнет определенного значения при заданном $X(t_1)$ ($t_1 < t_2$), зависит только от t_1 и t_2 , а не от поведения $X(t)$ при $t < t_1$. Существует необходимость введения такого случайного процесса, который обладал

бы некоторой памятью. Такой процесс получил название фрактального броуновского движения (ФБД) и был исследован Бенуа Мандельбротом и Ван Несом в 1968 году.

Для аппроксимации фрактального броуновского движения нет простого метода, вроде суммирования гауссовских случайных величин, как в случае классического броуновского движения. ФБД удобно определить при помощи параметра H , $0 < H < 1$. При $H = \frac{1}{2}$ фрактальное броуновское движение совпадает с классическим. Параметр H соответствует степени изрезанности графика. При малых H , близких к 0, график получается сильно изрезанным, а при больших H , близких к 1 — весьма плавным, хотя и не гладким.

Гауссовский процесс $X(t)$ называется *фрактальным броуновским движением* на $[a; b]$ с параметром $H \in (0, 1)$, если оно обладает следующими свойствами:

1. $X(a) = 0$, и функция $X(t)$ почти всегда непрерывна;
2. случайная величина $\Delta X = X(t_2) - X(t_1)$, $t_2 > t_1$, имеет гауссовское распределение с математическим ожиданием $M(\Delta X) = 0$ и дисперсией $D(\Delta X) = \sigma^2(t_2 - t_1)^{2H}$, где σ — положительная константа, т. е.

$$P(\Delta X < x) = \frac{1}{\sqrt{2\pi\sigma^2(t_2 - t_1)^{2H}}} \int_{-\infty}^x e^{-\frac{u^2}{2\sigma^2(t_2 - t_1)^{2H}}} du.$$

Параметр H называется *показателем Херста*.

Свойства фрактального броуновского движения

1⁰. Пусть $X(t)$ — ФБД на интервале $[a; b]$. Тогда его приращения стационарны, и $M[X(t_2) - X(t_1)]^2 = \sigma^2 |t_2 - t_1|^{2H}$, $\forall t_1, t_2 \in [a; b]$.

2⁰. Пусть $X(t)$ — ФБД с параметром H , $0 < H < 1$, на отрезке $[a; b]$. Приращения $X(t)$ независимы тогда, и только тогда, когда $H = \frac{1}{2}$.

3⁰. ФБД с параметром H , $0 < H < 1$, не является марковским процессом, за исключением случая $H = \frac{1}{2}$.

4⁰. Пусть $X(t)$ — ФБД с параметром H , $0 < H < 1$, на интервале $[a; b]$. Тогда математическое ожидание модуля приращения равно $M(|X(t_2) - X(t_1)|) = \sqrt{\frac{2}{\pi}} \sigma |t_2 - t_1|^H$.

5⁰. ФБД $X(t)$ с параметром H , $0 < H < 1$, на отрезке $[a; b]$ недифференцируемо.

6⁰. Фрактальная размерность реализации ФБД равна

$$D = 2 - H.$$

7⁰. Приращение реализации ФБД обладают свойством статистического самоподобия, т. е.

$$X(t + \Delta t) - X(t) \triangleq \frac{1}{r^H} \left(X(t + r\Delta t) - X(t) \right)$$

для любого $r > 0$, где « \triangleq » означает, что две случайные величины имеют одинаковое распределение.

В лекции использованы материалы [5].

Лекция № 7. Временные ряды, их фрактальность. Метод R/S -анализа Херста

Многие наблюдения природных процессов приводят к временным зависимостям или рядам изменений. *Временной ряд* — это совокупность наблюдаемых параметров изучаемой системы во времени.

Метод R/S -анализа, предложенный первоначально Херстом для анализа временной зависимости стока воды из различных рек и озер, может быть применен для определения фрактальной размерности D и других временных рядов и однозначных самоаффинных кривых. Пусть $X(t)$ — исследуемая временная зависимость некоторой величины X , измеренной в

дискретные целочисленные моменты времени t . Представим себе, что X является некоторой накопленной величиной и может быть представлена как сумма некоторых элементарных вкладов $\xi(t)$. Например, если X — координата броуновской частицы, то ξ — это элементарные смещения такой частицы. Рассмотрим поведение величин X и ξ только в некотором ограниченном интервале времени t от 1 до τ . Для этого интервала времени можно вычислить среднюю величину приращений ξ :

$$\langle \xi \rangle_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} \xi(t).$$

Для того же интервала времени можно построить новую величину $X(t, \tau)$, которая будет иметь смысл накопленных отклонений от среднего за данный период:

$$X(t, \tau) = \sum_{u=1}^t \{\xi(u) - \langle \xi \rangle_\tau\}.$$

Нетрудно заметить, что $X(0, \tau) = X(\tau, \tau) = 0$, поскольку суммарное отклонение от среднего за время, в течении которого это среднее изменилось, в точности равно 0. *Размахом* $R(\tau)$ для зависимости $X(t, \tau)$ называется величина максимального изменения $X(t, \tau)$ за рассматриваемый интервал времени:

$$R(\tau) = \max_{1 \leq t \leq \tau} X(t, \tau) - \min_{1 \leq t \leq \tau} X(t, \tau).$$

Для анализа фрактальной размерности D исходной кривой $X(t)$ необходимо построить зависимость безразмерного отношения R/S от величины временного интервала, где S — среднеквадратичное отклонение ξ от среднего за тот же интервал времени:

$$S(\tau) = \sqrt{\frac{1}{\tau} \sum_{u=1}^{\tau} \{\xi(u) - \langle \xi \rangle_\tau\}^2}.$$

Херстом был установлен следующий эмпирический закон:

$$R(\tau)/S(\tau) \sim \tau^H,$$

где H — показатель Херста, связанный с фрактальной размерностью D кривой соотношением: $D = 2 - H$.

Таким образом, для определения фрактальной размерности D необходимо построить зависимость $R/S(\tau)$ в двойном логарифмическом масштабе и провести линейную аппроксимацию. Тангенс угла наклона β непосредственно дает значение показателя Херста $H = \beta$.

В изложенной выше реализации метода Херста зависимости $R(\tau)$ и $S(\tau)$ всегда рассчитываются только для начального интервала времени τ . Метод может быть усовершенствован, если проводить усреднение $R(\tau)$ и $S(\tau)$ по нескольким временным окнам длительностью τ , взятым в разных положениях на анализируемой кривой $X(t)$.

Алгоритм. Для того чтобы найти a и H , необходимо сначала вычислить значение R/S некоторого временного ряда. Промежуток времени τ , на котором анализируется временная последовательность, называется запаздыванием. Сначала надо уменьшить запаздывание τ вдвое, вследствие этого получим из исходного временного ряда два независимых значения R/S , характеризующих по отдельности половинные выборки. Затем снова уменьшаем τ до тех пор, пока оно не будет меньше какого-то достаточно маленького положительного числа. Если на каком-то этапе значение запаздывания выражается нечетным числом, т. е. $\tau = 2k + 1$, $k \in \mathbb{N}$, то при делении его пополам к первой половине относятся первые k измерений, а ко второй — оставшиеся $k + 1$ измерений. Но при этом считается, что получившиеся подинтервалы имеют одно и то же значение запаздывания.

На каждом шаге число независимых областей будет удваиваться. Результаты R/S , соответствующие одному и тому же

значению запаздывания τ усредняются и вместе со значениями τ откладываются в двойном логарифмическом масштабе как функция запаздывания. В результате получится некоторое множество точек на фрактальной плоскости, которая образована координатными осями $X = \ln \tau$, $Y = \ln(R/S)$. Показатель Херста H определяет закон аппроксимирующей прямой фрактальной линии к оси абсцисс, т.е. H равен угловому коэффициенту прямой линейной регрессии. Зная значение H , можно найти константу $a = \frac{\tau}{(R/S)^{\frac{1}{H}}}$.

Классификация для показателя Херста H и фрактальной размерности D

1. $H = \frac{1}{2}$ ($D = 1,5$) указывает на случайный ряд. События случайны и некоррелированы. Настоящее не влияет на будущее. Функция плотности вероятности может быть нормальной кривой, однако, это не обязательное условие. R/S -анализ может классифицировать произвольный ряд безотносительно к тому, какой вид распределения ему соответствует.

2. $0 < H < \frac{1}{2}$ ($1,5 < D < 2$). Данный диапазон соответствует *антиперсистентным*, или *эргодическим рядам*. Такой тип системы часто называют «возврат к среднему». Если система демонстрировала «рост» в предыдущий период, то, скорее всего, в следующем периоде начнется спад. И наоборот, если шло снижение, то вероятен близкий подъем. Устойчивость такого антиперсистентного поведения зависит от того, насколько H близко к нулю. Такой ряд более изменчив, чем ряд случайный, т.к. состоит из частых реверсов спад–подъем. Антиперсистентная величина H дает более высокую фрактальную размерность и более прерывистую линию, чем классическое броуновское движение, и, следовательно, характеризует систему, более подверженную переменам.

3. $\frac{1}{2} < H < 1$ ($1 < D < 1,5$). Данный диапазон соответ-

ствуует *персистентным*, или *трендоустойчивым* рядам. Если ряд возрастал (убывал) в предыдущий период, то, вероятно, что он будет сохранять эту тенденцию какое-то время в будущем. Чем ближе H к $\frac{1}{2}$, тем более зашумлен ряд и тем менее выражен его тренд. Персистентный ряд — это фрактальное броуновское движение, или смещенные случайные блуждания. Сила этого смещения зависит от того, насколько H больше $\frac{1}{2}$. Персистентный временной ряд даёт более гладкую, менее зазубренную линию, нежели классическое броуновское движение.

Для очень большого количества наблюдений τ можно ожидать сходимости ряда к величине $H = \frac{1}{2}$, т.к. эффект памяти уменьшается до того уровня, когда становится незаметным. Другими словами, в случае длинного ряда наблюдений можно ожидать, что его свойства станут неотличимы от свойств классического броуновского движения поскольку эффект памяти рассеивается.

Для экономики размерность определяет, например, как та или иная компания будет реагировать на микро- и макроэкономические изменения рынка. В результате нелинейных законов, по которым могут быть приняты решения на финансовых рынках, график динамики рыночных цен и других показателей имеет вид фрактального броуновского движения.

В лекции использованы материалы [5,9,10].

Лекция № 8. Методы определения фрактальной размерности для временных рядов

Для вычисления фрактальной размерности разработано множество методов.

Метод Ричардсона является исторически одним из первых методов определения фрактальной размерности самоподобных

кривых. Идея метода заключается в измерении длины L анализируемой кривой при варьировании длины эталонного отрезка r , который многократно укладывается вдоль кривой. Метод был впервые предложен для исследования геометрии береговых линий. При этом анализ выполнялся вручную — длина береговой линии измерялась по карте с помощью циркуля. Было продемонстрировано, что чем больше радиус циркуля r , который фиксировался при проведении измерений, тем меньше оказывалась результирующая длина кривой L . Такой результат вполне понятен — чем больше эталонный отрезок r , тем сложнее отследить мелкие детали геометрии кривой, тем меньше получается суммарная измеренная длина L . Однако, наряду с этим тривиальным результатом, был получен и более важный и интересный результат. Оказалось, что в двойном логарифмическом масштабе зависимость $L(r)$ представляет собой прямую линию, что говорит о фрактальности и самоподобии береговых линий — при любой степени увеличения береговые линии выглядят одинаково сложно.

Компьютерная реализация метода Ричардсона позволяет определять фрактальную размерность кривых с повышенной точностью. В этом случае кривая на плоскости задается последовательностью координат (x_i, y_i) точек, представляющих кривую с максимальной доступной точностью (пространственным разрешением). Необходимо измерить длину кривой с некоторым выбранным шагом. При этом стартуют с начальной точки кривой и ищут пересечение кривой, заданной набором отрезков, с окружностью заданного радиуса r . Далее ищется новое пересечение кривой с окружностью, построенной из ранее найденной точки пересечения, и т. д. При этом может возникнуть неоднозначность, связанная с тем, что точек пересечения может оказаться несколько. В этом случае следует каждый раз выбирать либо самую удаленную, либо наоборот самую ближнюю вдоль кривой точку к центру окружности. Оказывается, результат определения фрактальной размерности не зависит

от способа выбора пересечения в этом случае.

Суммарная длина L кривой рассчитывается как суммарная длина всех N полученных отрезков длины r плюс последний неполный отрезок r' от последней найденной точки пересечения до конечной точки кривой:

$$L(r) = N(r) \cdot r + r'.$$

В случае, если кривая является замкнутой, величина L для данного r может быть уточнена за счет усреднения результата измерения длины по нескольким произвольным стартовым точкам. Шаг r для расчета L выбирается в диапазоне от некоторого минимального значения порядка длины наибольшего отрезка между двумя соседними точками (x_i, y_i) , задающими кривую, до максимального, соответствующего приблизительно 0,3 от длины L_0 , измеренной с максимальной доступной точностью.

Для определения фрактальной размерности D исследуемой кривой зависимость $L(r)$ строят в двойном логарифмическом масштабе, т. е. строят зависимость $\log L(\log r)$ и проводят линейную аппроксимацию. Тангенс угла наклона β линейной аппроксимации связан с величиной $D = 1 - \beta$.

В случае, если точки (x_i, y_i) расположены достаточно равномерно вдоль кривой (расстояния между соседними точками приблизительно равны), то метод Ричардсона может быть реализован в упрощенном виде, более приспособленном для компьютерного анализа. Предлагается заменить итеративный поиск пересечений отрезков, составляющих кривую, с окружностями, на алгоритм подсчета длины L кривой, используя сначала каждую точку кривой, затем каждую вторую точку, каждую третью точку и т. д. При этом учете каждой n -ой точки величины L и r вычисляются следующим образом:

$$L = \sum_{i=1}^m \sqrt{(x_{ni} - x_{n(i-1)})^2 + (y_{ni} - y_{n(i-1)})^2} + r', \quad r = \frac{L - r'}{m},$$

где $m = [N_0/n]$ — целая часть от деления общего количества точек кривой N_0 на n , $r' = \sqrt{(x_{N_0-1} - x_{nm})^2 + (y_{N_0-1} - y_{nm})^2}$ — длина последнего отрезка, соединяющего последнюю точку с номером, кратным n , и последнюю точку кривой. Фактически, r в этом случае вычисляется как средняя длина всех отрезков, последовательно соединяющих все точки с номерами, кратными n , а L — суммарная длина этих отрезков плюс длина последнего остаточного отрезка.

Если кривая является замкнутой, то величины L и r могут быть дополнительно уточнены за счет усреднения результата по нескольким стартовым точкам. Причем для случая учета каждой n -ой точки имеется ровно n вариантов выбора стартовой точки. Процедура ускорения может быть произведена и для случая незамкнутой кривой, однако в этом случае необходимо учитывать длину двух остаточных отрезков в начале и в конце кривой. С полученной зависимостью $L(r)$ поступают точно так же, как и в предыдущем случае — из линейной аппроксимации зависимости $\log L(\log r)$ получают величину тангенса угла наклона β , дающего $D = 1 - \beta$.

Другим классическим методом анализа фрактальной размерности является *метод Колмогорова*, который, так же как и метод Ричардсона, применим для самоподобных фрактальных кривых. Идея метода заключается в измерении длины кривой при различной степени огрубления ее представления. С этой целью плоскость, на которой представлена кривая, разбивается на квадратную сетку с шагом, соответствующим наибольшему пространственному разрешению, с которым задана кривая. Оценка длины кривой производится подсчетом числа элементов сетки, через которые проходит эта кривая: $L(a) = N(a) \cdot a$, где a — размер элемента сетки. Затем сетка огрубляется вдвое и расчет длины кривой повторяется. Естественно, что после огрубления длина кривой окажется меньше, т. к. эта процедура приводит к потере информации о мелких деталях геометрии кривой. Таким образом, величина L рассчитывается для разно-

го шага сетки a (для разной степени огрубления). Далее зависимость $L(a)$ анализируется так же, как и зависимость $L(r)$ в методе Ричардсона.

Одним из наиболее точных методов анализа фрактальной размерности признан *метод Минковского*. В этом методе исследуемая кривая покрывается окружностями и рассчитывается зависимость площади полученного покрытия S от радиуса окружности r . В действительности, конструирование такого покрытия и расчет его площади является довольно сложной задачей. Это покрытие представляет собой наложение очень большого количества пересекающихся окружностей, построенных для всех точек анализируемой кривой.

На практике для построения зависимости $S(r)$ строят карту Евклидовых расстояний. Для этого плоскость, в которой построена кривая, разбивается на квадратную сетку и каждому элементу сетки ставится в соответствие число, равное минимальному расстоянию от центра этого элемента до кривой. Полученную матрицу расстояний часто представляют в виде полутонового изображения. Число элементов матрицы $N(s \leq r)$ со значением s , меньшим или равным радиусу r , умноженное на площадь одного элемента a^2 , совпадает площадью $S(r)$ покрытия кривой окружностями радиуса r : $S(r) = N(s \leq r) \cdot a^2$.

Для определения фрактальной размерности D кривой методом Минковского пользуются следующим эмпирическим законом: $S(r) \sim r^{2-D}$. Таким образом, линейная аппроксимация зависимости $\log S$ ($\log r$) должна давать величину тангенса наклона β , соответствующую $D = 2 - \beta$.

Необходимо отметить, что в изложенном выше виде метод Минковского применим только для анализа самоподобных кривых. Однако, достаточно простая модификация позволяет применять этот метод для самоаффинных кривых. Например, в случае временных рядов необходимо проводить покрытие кривой не окружностями, а горизонтальными отрезками разной длины, параллельными оси времени. А для построения карты

Евклидовых расстояний в этом случае необходимо искать минимальное расстояние от элемента матрицы до кривой только в горизонтальном направлении.

Следующая группа методов так же, как и метод Минковского, позволяет проводить анализ фрактальной размерности как самоподобных, так и самоаффинных кривых. Однако, эти методы применимы только к однозначным кривым.

В *методе Корчака* проводится сечение исследуемой кривой горизонтальными прямыми линиями, параллельными оси абсцисс. В результате такой процедуры получается такой набор отрезков — сечений, для которых строится кумулятивная функция распределения $N(w)$ — число отрезков N с длиной, превышающей величину w . Для получения достаточной для анализа статистики необходимо провести порядка 10 сечений в диапазоне от 10% до 90% от общего размаха высоты кривой в исследуемом интервале. Как правило, в статистику включаются отрезки, лежащие как над кривой, так и под ней. Однако, в некоторых случаях интересным оказывается исследование только одного типа отрезков.

Зависимость $N(w)$ для самоаффинной кривой с фрактальной размерностью D должна удовлетворять простому закону:

$$N(w) \sim w^{1-D},$$

т. е. в двойном логарифмическом масштабе зависимость $N(w)$ должна следовать линейному закону с тангенсом угла наклона $\beta = 1 - D$. Однако, линейный закон выполняется только для относительно небольшого начального участка графика $N(w)$, поэтому линейная аппроксимация здесь весьма затруднена. На практике аппроксимацию проводят выражением, учитывающим эффект ограничения диапазона скейлинга, который проявляется в этом методе:

$$N(w) = k \cdot w^{1-D} \cdot e^{-\frac{w}{\xi}},$$

где k — нормировочный коэффициент, а величина ξ имеет смысл длины корреляции. Во многих случаях определение параметра ξ оказывается не менее интересным, чем определение собственно фрактальной размерности D .

Метод Расса является упрощенной реализацией метода Херста, рассмотренного в предыдущей лекции. Предполагается анализировать зависимость $R(\tau)$, где R — максимальный размах исследуемой кривой $X(t)$, найденный в пределах временного интервала τ . При этом размах R для каждого τ осуществляется не только на начальном участке, а по всей кривой.

Для определения фрактальной размерности D зависимость $R(\tau)$ строится в двойном логарифмическом масштабе и из линейной аппроксимации находится тангенс угла наклона β , который дает показатель Херста $H = \beta$ и фрактальную размерность $D = 2 - H$.

Метод RMS так же, как и методы Херста и Расса, является оконным методом. Для анализа фрактальной размерности D некоторой самоаффинной кривой, заданной дискретным набором точек $y(x)$, расположенных линейно вдоль оси абсцисс X , необходимо построить следующую зависимость:

$$RMS(w) = \frac{1}{n_w} \sum_{j=1}^{n_w} \sqrt{\frac{1}{w-2} \sum_{i=1}^w [y_{i,j} - \langle y \rangle_j]^2},$$

где n_w — количество анализируемых окон размером w (под размером окна понимается количество содержащихся в нём точек) вдоль кривой, $\langle y \rangle_j$ — среднее значение в j -ом окне, $y_{i,j}$ — i -ая точка в j -ом окне. При этом перед началом вычислений в каждом окне из всех точек вычитается прямая линия, соответствующая либо аппроксимации точек окна, либо прямая соединяющая концевые точки этого окна.

Также, как и в случаях методов Херста и Расса, линейная аппроксимация зависимости $RMS(w)$, построенной в двойном

логарифмическом масштабе, даёт тангенс угла наклона $\beta = H$ и фрактальную размерность $D = 2 - H$.

Открытым остаётся вопрос о выборе набора окон для каждого размера w , которые будут участвовать в анализе и подсчёте величины RMS . Здесь существует некоторый произвол. Можно для каждого w выбирать одно и то же фиксированное количество окон n_w , размещённых либо линейно, либо случайным образом на кривой. Можно также выбирать все возможные окна размера w , которые могут быть размещены на кривой. Разумно при этом наложить ограничение, чтобы выбираемые окна для данного w были несмежными.

Метод разложения в ряд Фурье является наиболее быстрым способом определения фрактальной размерности самоаффинных кривых, поскольку быстрое Фурье преобразование присутствует в подавляющем большинстве математических пакетов. Однако, вместе с тем, этот метод является одним из наименее точных.

Процедура разложения в ряд Фурье даёт на выходе значения амплитуды c и фазы φ для некоторого фиксированного ряда частот f , по которому происходит разложение анализируемой кривой $y(t)$. Вид разложения совпадает с выражением

$$y(t) = \sum_{f=f_{min}}^{f_{max}} c(f) \cdot \sin(2\pi ft + \varphi(f)),$$

где $f \in [f_{min}, f_{max}]$ — набор частот, входящих в разложение Фурье, $c(f)$ и $\varphi(f)$ — амплитуда и фаза соответствующих гармоник. Анализ частотного спектра мощности $c^2(f)$ позволяет получить величину фрактальной размерности D для исследуемой кривой. Линейная аппроксимация зависимости $c^2(f)$, представленной в двойном логарифмическом масштабе, даёт тангенс угла наклона β , связанный с

$$D = \frac{4 + \beta}{2}.$$

Классификация фрактальных временных рядов

Наиболее известной классификацией фрактальных временных рядов является классификация по спектру Фурье временного ряда или $\frac{1}{f}$ -классификация. Фурье-спектр хаотических временных рядов обычно описывается соотношением

$$A \sim \frac{1}{f^d},$$

где A — амплитуда, f — частота, а d — показатель в зависимости от значения которого процесс относится к одному из процессов.

В 1968 году Б. Мандельброт и Ван Несс предположили, а в 1989 году Фландрин строго доказал связь показателя d , определенного из спектра Фурье временного ряда, с показателем Херста H , определенного по значениям этого же временного ряда:

$$d = 2H + 1.$$

Сформулируем принципиальные различия в поведении хаотических временных рядов разного типа.

При $d = 0$ наблюдаемый процесс является хорошо изученным **белым шумом** или последовательностью независимых, распределенных по нормальному закону с постоянным средним и некоторой дисперсией случайных величин. Спектр Фурье такого временного ряда является постоянной величиной при всех значениях частоты f .

Интеграл белого шума представляет собой другой не менее известный процесс: траекторию одномерного броуновского движения, или винеровский процесс. В этом случае $d = 2$ ($H = 0,5$), а тип поведения называется **коричневым шумом**. Основным свойством этого процесса является отсутствие памяти: следующее приращение ряда не зависит от всех предыдущих.

При $1 \leq d < 2$ ($0 \leq H < 0,5$) процесс называется **розовым шумом**. В таких процессах существует «отрицательная» память: если в прошлом наблюдалось положительное приращение, то в будущем с высокой вероятностью будет наблюдаться отрицательное, и наоборот.

При $2 < d \leq 3$ ($0,5 < H \leq 1$) процесс называется **черным шумом** и обладает «положительной» памятью: если в прошлом наблюдалось положительное приращение, то в будущем с высокой вероятностью будет наблюдаться положительное, и наоборот.

Большинство временных рядов наблюдаемых в природе и науке, обычно можно отнести к одному из рассмотренных классов. Так коричневому шуму соответствует одномерное броуновское движение (так же долгое время предполагалось, что этим процессом описывается поведение цен на бирже). Розовым шумам соответствуют временные ряды, наблюдаемые в процессах турбулентности, к этому же типу поведения относятся ряды, описывающие цены финансовых активов. Черные шумы регистрируются в разливах рек, солнечной активности, статистике природных и техногенных катастроф, некоторые исследователи полагают, что поведение цен акций на бирже относятся к этому классу процессов.

Рассмотрим еще *метод исключения трендов* (Detrended fluctuation analysis, далее — DFA) и две его модификации: с линейной аппроксимацией и аппроксимацией четвертого порядка. Следуя сложившейся в экономических исследованиях практике, указанные варианты метода для краткости обозначают соответственно как DFA-1 и DFA-4. Метод исключения трендов является широко известным средством исследования корреляционных зависимостей во временных рядах.

Алгоритм метода исключения трендов. В качестве $x(t)$ обозначим исследуемый временной ряд.

1. Ряд $x(t)$, содержащий N значений, разбивается на $\frac{N}{\tau}$

непересекающихся интервалов. При этом каждый интервал содержит τ значений временного ряда.

2. В пределах каждого интервала осуществляется линейная аппроксимация ряда $x(t)$ (например, методом наименьших квадратов): $y_\tau(t) = a_\tau t + b_\tau$, где a_τ и b_τ — постоянные для каждого интервала.

3. Для каждого интервала вычисляется величина:

$$F_i^2(\tau) = \frac{1}{\tau} \sum (x(t) - y_\tau(t))^2.$$

Далее рассчитывается следующий показатель:

$$F^2(\tau) = \frac{\tau}{N} \sum_{i=1}^{\frac{N}{\tau}} F_i^2(\tau).$$

4. Используя линейную аппроксимацию $\log F(\tau)$ и $\log \tau$ в дважды логарифмической системе координат и измеряя угол наклон полученной зависимости β , вычисляем размерность временного ряда как $D = 2 - \beta$.

Существует несколько модификаций метода исключения трендов. Наиболее распространенная состоит в том, что вместо линейной аппроксимации (второй этап описанной выше процедуры) применяется аппроксимация более высоких порядков (например, в виде многочлена четвертой степени).

Одно из практических применений фрактального анализа состоит в исследовании взаимосвязей во временном ряду. Доказано, что если фрактальная размерность принимает значение 1,5, то приращения во временном ряду независимы между собой. Случай, когда фрактальная размерность меньше 1,5, соответствует персистентному ряду, характеризующемуся эффектами «долговременной памяти». Если фрактальная размерность больше 1,5 — отвечает антиперсистентному поведению временного ряда.

Курсовая динамика валют развитых стран, не подвергшихся валютным кризисам, характеризуется наиболее стабильной фрактальной структурой: их фрактальная размерность практически не отличается от значения 1,5. Такое состояние обеспечивает устойчивость макроэкономических систем развитых стран.

Перед валютными кризисами курсы валют развивающихся стран обладают завышенными или заниженными значениями фрактальной размерности по отношению к 1,5. В большинстве случаев фрактальная размерность отклоняется от значения 1,5 в меньшую сторону, что свидетельствует о формировании инвесторами «стадного» поведения.

После завершения валютного кризиса и наступления фазы послекризисной стабилизации фрактальные размерности приближаются к значению 1,5.

В целях исследования нестационарной динамики фрактальной размерности применяют *метод «скользящего окна»*. Суть его в следующем. В начале ряда данных выделяется последовательность значений, и для этих данных рассчитывается фрактальная размерность, например, методом DFA-1 или DFA-4. Далее интервал передвигается вправо вдоль временного ряда на одно значение и снова вычисляется фрактальная размерность. Повторяя эту процедуру в течение исследуемого периода, получаем локальные значения фрактальной размерности. В лекции использованы материалы [5,9,10].

Лекция № 9. Массовые фракталы

Под массовыми фракталами понимаются моделируемые и реальные объекты, состоящие из дискретного набора частиц-мономеров, агрегированных в большие кластеры с фрактальной геометрией. Часто при исследовании таких объектов прибегают к анализу пространственного распределения плотности вещества (или распределению массы объекта). Отсюда и название — массовые фракталы.

Модель ограниченной диффузией агрегации (ОДА) является наиболее популярным подходом для описания образования массовых фракталов. Эта модель может быть легко реализована в виде компьютерного алгоритма и хорошо подходит для численного исследования фрактальной геометрии таких объектов. Рассмотрим классический вариант ОДА — модель частица–кластер.

Для описания фрактальных структур, встречающихся на микромасштабах, широко используют понятие кластер. Это скопление близко расположенных, тесно связанных друг с другом частиц любой природы (атомов, молекул, ионов и иногда ультрадисперсных частиц) общим количеством 2-100 частиц. В последнее время термин кластер распространяется и на системы, состоящие из большого числа связанных макроскопических частиц. Фрактальный кластер — это хаотический фрактал, воспроизводимый системой с множеством хаотически, броуновски движущихся частиц, которые слипаются при соприкосновении с центром агрегации, с образованием разветвлённого кластера. В литературе фрактальный кластер также называют фрактальным агрегатом. Этот процесс может протекать при разных условиях, что отражается на свойствах образуемого кластера. Совокупность разных физических условий, при которых происходит рост кластера, составляет основу моделей, используемых для анализа процесса роста кластера и его свойств. Компьютерные модели, использующие определённые алгоритмы для описания движения и слипания твёрдых частиц и кластеров в процессе роста фрактального кластера, позволяют сконструировать фрактальный кластер и проанализировать как его свойства, так и характер роста.

Исторически первая компьютерная модель для описания сборки кластера была предложена Виттенем и Сандером в 1981 году и была названа DLA-моделью (агрегация, ограниченная диффузией, сокращенно ОДА). Представим кластер, растущий следующим образом: время от времени к нему присоединяется

одна молекула, так что когда частица входит в контакт с растущим объектом, она прилипает к нему и не ищет другого места, а, попросту говоря, остается на месте. Такой процесс называется агрегацией. Он представляет собой крайний пример неравновесного процесса роста, поскольку в нем совершенно отсутствует переупорядочение. Теперь предположим, что частицы диффундируют к кластеру в ходе случайного движения, т.е. последовательности шагов, длина и направление которых определяются случайным образом. Фрактальная структура, полученная путем машинного моделирования процесса, называемого агрегацией, ограниченной диффузией (см. рис. 12).

Хотя процесс агрегации, ограниченной диффузией, легко поддается описанию и моделированию, его природа еще не достаточно изучена. Почему, например, процесс такой агрегации приводит к образованию фракталов, а не, скажем, аморфных комков, не обладающих никакой симметрией? Почему так редко образуются петли? Каким образом фрактальная размерность зависит от размерности пространства? Ответы на эти вопросы остаются открытыми. Однако в качественном смысле можно понять некоторые важные свойства процесса. Допустим, что процесс начинается с гладкого кластера, к которому затем присоединяются частицы в ходе агрегации, ограниченной диффузией; когда кластер еще мал, несколько частиц могут налипнуть чисто случайно на каком-то одном участке его поверхности. Другими словами, благодаря «шуму», т.е. наличию случайного элемента в поведении частиц, на поверхности объекта образуются крошечные бугорки и ямки. В лекции использованы материалы [10].

Лекция № 10. Принцип Парето

В 1897 году итальянский социолог и экономист Вильфредо Парето (1848-1923) обнаружил, что распределение богатства в обществе обычно подчиняется определенному закону — с удво-

ением размера контролируемой собственности (богатства), количество людей, достигших соответствующего уровня сокращается в геометрической прогрессии, причем с примерно постоянным множителем. Таким образом, Парето пришёл к выводу, что неравенство распределения богатства в обществе — нечто вроде естественного закона природы, эффект которого можно сгладить, но невозможно устранить. Закон Парето, или Принцип Парето, или принцип 20/80 — эмпирическое правило в наиболее общем виде формулируется как «20% усилий дают 80% результата, а остальные 80% усилий — лишь 20% результата». Может использоваться как базовый принцип для оптимизации какой-либо деятельности: правильно выбрав минимум самых важных действий, можно быстро получить значительную часть от планируемого полного результата, причём дальнейшие улучшения не всегда оправданы. Приводимые в законе цифры нельзя считать безусловно точными — это скорее просто мнемоническое правило, нежели реальные ориентиры. В действительности для каждой задачи следует проводить соответствующую математическую обработку.

Формализацией принципа Парето является одномерная функция непрерывного двухпараметрического (α, x_0) распределения $F(x) = P(X < x)$ (вероятность того, что случайная величина X принимает значение, меньшее x) и соответствующая ей плотность вероятности $\varphi(x)$ через производную функции $F(x)$, которые задаются соотношениями степенного вида:

$$F(x) = 1 - \left(\frac{x_0}{x}\right)^\alpha, \quad \varphi(x) = \frac{d}{dx}F(x) = \frac{\alpha}{x_0} \left(\frac{x_0}{x}\right)^{\alpha+1},$$

где $\alpha > 0$, $x \geq x_0 > 0$, или после нормировки $\frac{x}{x_0}$ для нормированной величины x

$$F(x) = 1 - x^{-\alpha}, \quad \varphi(x) = \alpha x^{-\alpha-1}, \quad \alpha > 0, \quad x \geq 1,$$

где α — показатель степени; x_0 — минимально возможное значение x . Подобные зависимости называют распределениями с

«тяжелыми хвостами» (ТХ, heavy tails), — их нельзя «обрезать», как и пренебрегать крупными, но редкими событиями. Степенной параметр α как раз и характеризует «тяжелый хвост».

Главная неприятность таких распределений состоит в том, что моменты достаточно высокого порядка $M_q = \int x^q dF(x)$ у них расходятся: $M_q \rightarrow \infty$ для $q \geq \alpha$. Для распределения Парето существуют только моменты порядка, меньшего степени α . При $\alpha \leq 1$ бесконечно уже математическое ожидание или среднее M_1 , при $\alpha \leq 2$ случайная величина X имеет бесконечную дисперсию, связанную с M_2 , и т. д. На расходимость моментов влияет именно «тяжелый хвост», когда вид «головы» оказывается не существенным, а главную роль играет лишь асимптотика хвоста. С ростом α распределение «вырождается» и математическое ожидание стремится к x_0 .

Само распределение обладает свойством самоподобия: локализация значений, превышающих величину $z_0 \geq x_0$, также характеризуется распределением Парето:

$$\psi(z, \alpha, z_0) = \frac{\psi(z, \alpha, x_0)}{1 - F(z_0, \alpha, x_0)} = \frac{\alpha}{z_0} \left(\frac{z_0}{z} \right)^{1+\alpha}.$$

«Тяжелые хвосты» степенной зависимости, которые обычно отсекаются в распределениях Гаусса, при больших значениях α могут проявить себя самым неожиданным способом: отдельные, хотя и маловероятные случаи в такой системе все же возможны и могут породить в ней катастрофические изменения. В случае ТХ-распределений выборочные средние значения неустойчивы и малоинформативны, и к ним неприменим закон больших чисел. Для временных последовательностей сумма элементов ряда имеет такой же порядок, что и максимальный элемент ряда, и они характеризуются медленным уменьшением числа редких событий. Степенные ТХ распределения являются эффективным средством для моделирования

многих процессов и потоков данных в самых разных науках, описывают абсолютное большинство катастрофических ситуаций, кризисных явлений и стихийных бедствий с большими ущербами и т. п.

Лекция № 11. Методы исследования документальных информационных потоков

Оказывается, все созданные человеком тексты построены по единым правилам! Никому не удастся обойти их. Какой бы язык ни использовался, кто бы ни писал — классик или графоман — внутренняя структура текста останется неизменной. Она описывается законами Дж. Зипфа (G.K. Zipf). Зипф предположил, что природная лень человеческая (впрочем, это свойство любого живого существа) ведет к тому, что слова с большим количеством букв встречаются в тексте реже коротких слов. Основываясь на этом постулате, Зипф вывел два универсальных закона.

Первый закон Зипфа «ранг–частота». Выберем любое слово и посчитаем, сколько раз оно встречается в тексте. Эта величина называется частота вхождения слова. Измерим частоту каждого слова текста. Некоторые слова будут иметь одинаковую частоту, т. е. входить в текст равное количество раз. Сгруппируем их, взяв только одно значение из каждой группы. Расположим частоты по мере их убывания и пронумеруем. Порядковый номер частоты называется ранг частоты. Так, наиболее часто встречающиеся слова будут иметь ранг 1, следующие за ними — 2 и т. д. Распределение слов естественного языка подчиняется простому закону: $f \cdot r = c$, где f — частота встречаемости слова в тексте; r — ранг (порядковый номер) слова в списке; c — эмпирическая постоянная величина. Значение константы c в разных языках различно, но внутри одной языковой группы остается неизменно, какой бы текст мы ни взяли: для английских текстов константа Зипфа равна приблизительно 0.1, для

русского языка коэффициент Зипфа получился равным 0.06–0.07.

Рассматривая первый закон, мы отмахнулись от факта, что разные слова входят в текст с одинаковой частотой. Зипф установил, что частота и количество слов, входящих в текст с этой частотой, тоже связаны между собой. Если построить график, отложив по одной оси (оси X) частоту вхождения слова, а по другой (оси Y) — количество слов в данной частоте, то получившаяся кривая будет сохранять свои параметры для всех без исключения созданных человеком текстов! Как и в предыдущем случае, это утверждение верно в пределах одного языка. Однако и межъязыковые различия невелики. На каком бы языке текст ни был написан, форма кривой Зипфа останется неизменной. Могут немного отличаться лишь коэффициенты, отвечающие за наклон гиперболы. Пусть N — количество слов в тексте, $W(A)$ — частота появления слова A . Тогда второй закон Зипфа имеет вид $N \cdot W(A) = C$, где C есть величина постоянная.

Законы Зипфа универсальны. В принципе, они применимы не только к текстам. В аналогичную форму выливается, например, зависимость количества городов от числа проживающих в них жителей. Характеристики популярности узлов в сети Интернет — тоже отвечают законам Зипфа. Такая же закономерность действует для доли рынка программного обеспечения, безалкогольных напитков, автомобилей, конфет. Неоднократно публиковались работы, из которых следует, что подобный тип распределения, помимо лингвистики, характерен для объектов в области: биологии, экономики, социологии, науковедения и т. п.

Закон Зипфа был открыт экспериментально в 1949 году. Пять лет спустя знаменитый математик Бенуа Мандельброт внес небольшие изменения в формулы Зипфа, добившись более точного соответствия теории практике. Позднее Б. Мандельброт предложил его теоретическое обоснова-

ние. Мандельброт полагал, что можно сравнивать письменный язык с кодированием, причем все знаки должны иметь определенную «стоимость». Исходя из требований минимальной стоимости сообщений, Б. Мандельброт математическим путем пришел к аналогичной закону Зипфа зависимости $fr^\gamma = c$, где γ — величина (близкая к единице), которая может изменяться в зависимости от свойств текста.

При этом коэффициент γ — величина переменная. Постоянство коэффициента γ сохраняется только на среднем участке графика распределения. Этот участок принимает форму прямой, если график вышеприведенной закономерности построить в логарифмических координатах. Участок распределения с $\gamma = \text{const}$ называется центральной зоной рангового распределения (значение аргумента на этом участке изменяется от $\ln r_1$ до $\ln r_2$). Значениям аргумента от 0 до $\ln r_1$ соответствует зона ядра рангового распределения, а значениям аргумента от $\ln r_2$ до $\ln r_3$ — так называемая зона усечения. Какой же смысл заложен в существовании трех явно различаемых зон ранговых распределений? Если последнее относится к терминам, составляющим какую-либо область знания, то ядерная зона, или зона ядра рангового распределения, содержит наиболее общепотребительные, общенаучные термины. Центральная зона содержит термины, наиболее характерные для данной области знаний, которые в совокупности выражают ее специфичность, отличие от других наук, «охватывают ее основное содержание». В зоне усечения же сосредоточены термины, сравнительно редко употребляющиеся в данной области знаний. График рангового распределения наполнен глубоким смыслом: ведь по относительной величине той или иной зоны на графике можно судить о характеристиках всей области знаний. График с обширной ядерной зоной и малой зоной усечения относится к достаточно широкой и скорее всего консервативной области знаний. Для динамичных отраслей науки характерна увеличенная зона усечения. Малая величина ядерной зоны может

говорить об оригинальности области знаний, к которой относится построенное ранговое распределение и т. д. Так, на основании анализа рангового распределения оказалось возможным дать качественные оценки документальным информационным потокам в соответствии с теми отраслями науки, где они формировались.

Важное эмпирическое исследование массива публикаций в целях анализа научной продуктивности, принадлежащее А. Лотке, появилось в 1926 году. Распределение носит название *закона А. Лотки* или *закона обратных квадратов*: число ученых, написавших данное число статей, обратно пропорционально квадрату этого числа статей

$$n(x) = \frac{A}{x^2},$$

где $n(x)$ — число ученых, написавших x статей, A — число ученых, написавших всего одну статью; $x = 1, 2, \dots, x_{\max}$ (максимальная продуктивность ученого). Закон многократно проверялся на различных информационных массивах и оказался справедливым для всего диапазона научной продуктивности, кроме самых низких ($n = 1$) и самых больших значений («краевые эффекты»). Распределение научной продуктивности носит асимметрический, неравномерный характер. Инвариантность и устойчивость этого закона во множестве научных областей позволяет говорить о нем как об одном из важнейших библиометрических законов.

Частным случаем закона Зипфа является *закон Брэдфорда*, непосредственно связанный уже не с распределением слов в тексте, а с распределением документов внутри какой-либо тематической области. Английский химик и библиограф С. Брэдфорд, исследуя статьи по прикладной геофизике и смазке, заметил, что распределения научных журналов, содержащих статьи по смазке, и журналов, содержащих статьи по прикладной геофизике, имеют общий вид. На основа-

нии установленного факта С. Брэдфорд сформулировал закономерность распределения публикаций по изданиям. Основной смысл закономерности состоит в следующем: если научные журналы расположить в порядке убывания числа статей по конкретному вопросу, то журналы в полученном списке можно разбить на три зоны таким образом, чтобы количество статей в каждой зоне по заданному предмету было одинаковым. При этом в первую зону, так называемую зону ядра, входят профильные журналы, непосредственно посвященные рассматриваемой тематике. Количество профильных журналов в зоне ядра невелико. Вторую зону образуют журналы, частично посвященные заданной области, причем число их существенно возрастает по сравнению с числом журналов в ядре. Третья зона, самая большая по количеству изданий, объединяет журналы, тематика которых весьма далека от рассматриваемого предмета. При равном числе публикаций по определенной тематике в каждой зоне число наименований журналов резко возрастает при переходе от одной зоны к другой. С. Брэдфорд установил, что количество журналов в третьей зоне будет примерно во столько раз больше, чем во второй зоне, во сколько раз число наименований во второй зоне больше, чем в первой. Обозначим P_1 — число журналов в 1-й зоне, P_2 — во 2-й, P_3 — число журналов в 3-й зоне. Если a — отношение количества журналов 2-й зоны к числу журналов 1-й зоны, то закономерность, вскрытая С. Брэдфордом, может быть записана так:

$$P_1 : P_2 : P_3 = 1 : a : a^2 \text{ или } P_3 : P_2 = P_2 : P_1 = a.$$

Б. Викери уточнил модель С. Брэдфорда. Он выяснил, что журналы, проранжированные (выстроенные) в порядке уменьшения в них статей по конкретному вопросу, можно разбить не на три зоны, а на любое нужное число зон. Если периодические издания расположить в порядке уменьшения в них количества статей по конкретному вопросу, то в полученном списке можно выделить ряд зон, каждая из которых содержит одинаковое

количество статей. Введем обозначения: x — количество статей в каждой зоне, T_x — количество журналов, содержащих x статей, T_{2x} — количество журналов, содержащих $2x$ статей, т. е. сумма наименований журналов в 1-й и во 2-й зонах, T_{3x} — количество журналов, содержащих $3x$ статей, т. е. сумма наименований журналов в 1-й, 2-й и в 3-й зонах, T_{4x} — количество журналов, содержащих $4x$ статей. Тогда эта закономерность будет иметь вид

$$T_x : T_{2x} : T_{3x} : T_{4x} : \dots = 1 : a : a^2 : a^3 : \dots$$

Данное выражение называют *законом Брэдфорда в толковании Б. Викери*.

Если закон Зипфа характеризует многие явления социального и биологического характера, то закон Брэдфорда — это специфический случай распределения Зипфа для системы периодических изданий по науке и технике. Из этих закономерностей можно извлечь выводы огромной практической пользы. Закономерности концентрации и рассеяния научно-технической информации во множестве документов позволяют выбирать именно те издания, которые с наибольшей вероятностью содержат публикации, соответствующие определенному профилю знаний. Существующее рассеяние публикаций нельзя оценивать только как вредное явление. В условиях рассеяния улучшаются возможности для межотраслевого обмена информацией.

Формула Брукса уточняет закон Брэдфорда и состоит из двух уравнений:

$$R_n = an^b \quad (1 \leq n \leq c),$$

$$R_n = k \log_b \frac{n}{s} \quad (c < n \leq N),$$

где R_n — суммарное число публикаций по данной отрасли, содержащееся в n периодических изданиях, проранжированных в порядке уменьшения «продуктивности»; n — ранг (порядковый номер) периодического издания в списке по убывающей

продуктивности журнала; a — количество релевантных публикаций в самом продуктивном журнале; c — число периодических изданий в «ядре»; N — общее количество периодических изданий в ранговом списке; b, k, s — эмпирические коэффициенты. Первое уравнение описывает начальный участок кривой, который имеет экспоненциальный характер, а второе — прямолинейную часть кривой Брэдфорда. В зависимости от соотношения коэффициентов делают заключение о состоянии научного направления. Установлено, что для узкоспециализированных областей науки $s \leq 1$, а величина $c \geq 3$ и связана с s прямой зависимостью.

Оказывается, можно определять не только количественные параметры совокупностей научных документов, но и совокупностей элементов признаков научных документов: авторов, терминов, индексов классификационных систем, наименований изданий, т. е. наименований элементов, характеризующих содержание научных документов. Например, можно расположить журналы в порядке убывания числа печатающихся в них авторов, в порядке убывания средней величины публикуемых в них статей или упорядочить совокупность документов по любому ее элементу. В. И. Горькова доказала, что одним из критериев, позволяющих оценить развитие отдельных научных дисциплин и их составляющих, может являться средняя скорость роста количества публикаций за интервал времени Δt . В. И. Горьковой предложена следующая формула:

$$V = \frac{\Delta N}{N \cdot \Delta t},$$

где N — число публикаций в базисном году; ΔN — прирост за период Δt ; V — скорость роста числа публикаций, определяемая их приростом (ΔN) во времени (Δt) по отношению к базисному году (N) по данным экспериментальных исследований за ряд лет (более 5).

В лекции использованы материалы [14,15,16].

Лекция № 12. Фрактальные свойства информационного пространства

В настоящее время информационное пространство в целом, ввиду его объемов и динамики изменения, принято рассматривать как стохастическое. Во многих моделях информационного пространства изучаются структурные связи между тематическими множествами, входящими в это пространство. При этом численные характеристики этих множеств подчиняются гиперболическому закону (с возможными степенными поправками). Сегодня в моделировании информационного пространства все чаще используется фрактальный подход, базирующийся на свойстве самоподобия информационного пространства, т. е. сохранение внутренней структуры множеств при изменении их размеров или масштабов их рассмотрения извне.

Информационный поиск — наука, изучающая поиск информации в документах, расположенных в базе данных. При этом база данных может быть как автономной, так и сетевой, такой как «Всемирная паутина» (World Wide Web). Сущность всех задач информационного поиска сводится к поиску информации по некоторому запросу с дальнейшей обработкой результатов. В настоящее время в связи с широким распространением сети Интернет задачи подобного рода стали особенно актуальными. Самый наглядный пример систем, решающих эти задачи, — информационно-поисковые системы, позволяющие пользователю ввести запрос на естественном языке и получить более или менее релевантные результаты, т. е. документы, семантически связанные с запросом.

Кластеризация в информационном поиске является одним из способов повысить эффективность поиска. Кластер (англ. cluster скопление) — объединение нескольких однородных элементов, которое может рассматриваться как самостоятельная единица, обладающая определёнными свойствами. Поскольку ни один из существующих алгоритмов семантического анализа

текстовой информации не может работать безошибочно, особую важность приобретает возможность распределить результаты поиска по группам (кластерам), чтобы облегчить для пользователя процесс интерпретации результатов.

Несколько лет назад рядом исследователей было предложено применить для исследования кластерных систем фрактальный подход. В 2002 г. российский исследователь С. А. Иванов, обобщив результаты опубликованных работ и собственных исследований, предложил теорию стохастических фракталов для кластерных тематических образований в мировой системе периодических и продолжающихся изданий.

Самоподобие информационного пространства выражается, прежде всего в том, что при его лавинообразном росте в последние десятилетия, частотные и ранговые распределения, получаемые в таких разрезах, как источники, авторы, тематика практически не меняют своей формы. Поэтому применение теории фракталов при анализе информационного пространства позволяет с общей позиции взглянуть на закономерности, составляющие основы информатики. Например, тематические информационные массивы сегодня представляют развивающиеся самоподобные структуры, которые по своей сути являются стохастическими фракталами, т. к. их самоподобие справедливо лишь на уровне математических ожиданий, например, распределения кластеров по размерам. В информационном пространстве возникают, формируются, растут и размножаются кластеры — группы взаимосвязанных документов.

Не так давно в Интернет появился сервис Touchgraph (www.touchgraph.com), который наглядно демонстрирует появление кластерных образований, сформированных подобием информационных объектов, в частности, Web-сайтов (Touchgraph Google Browser). Пример такой визуализации см. на рис. 13.

Чем же определяется природа фрактальной структуры информационного пространства, порождаемым такими кластерными структурами? С одной стороны, параметрами ранговых

распределений, а, с другой стороны, механизмом развития информационных кластеров, который отражает природу информационного пространства. Появление новых публикаций увеличивает размерность уже существующих кластеров и является причиной образования новых. Фрактальные свойства характерны для кластеров информационных Web-сайтов, на которых публикуются документы, соответствующие определенным тематикам. Эти кластеры, как наборы тематических документов, представляют собой фрактальные структуры, обладающие рядом уникальных свойств. Например, российскими исследователями (С. Иванов и др.), определена фрактальная размерность подобных информационных массивов, изменяющаяся в пределах от 1.05 до 1.50, что свидетельствует о небольшой плотности заполнения кластеров документами по одной теме. Как один из основных законов отражающих самоподобие информационного пространства можно назвать закон Зипфа. Существуют также закономерности, открытые другими учеными (прежде всего, Брэдфордом — для периодических изданий и Лотки — для распределения авторов), являющиеся уточняющими следствиями закономерностей Зипфа, и также свидетельствующими о самоподобии информационного пространства.

Теория фракталов тесно связана с кластерным анализом, решающим задачу выделения компактных групп объектов с близкими свойствами. Кластеризация сегодня применяется при реферировании больших документальных массивов, определении взаимосвязанных групп документов, для упрощения процесса просмотра при поиске необходимой информации, нахождения уникальных документов из коллекции, выявления дубликатов или близких по содержанию документов.

Фрактальный принцип самоподобия предполагает бесконечное дробление набора объектов с сохранением их свойств. В тематических информационных потоках, например, можно наблюдать подобие сюжетных цепочек, получаемых при

уточнении запроса (конечно, в определенных рамках). Вместе с тем, сегодня многими исследователями рассматривается не дробление, а естественный рост размеров информационного пространства. Свойства самоподобия фрагментов информационного пространства наглядно демонстрирует новый интерфейс представленный на веб-сайте службы News Is Free (<http://newsisfree.com>). На этом сайте отображается состояние информационного пространства в виде ссылок на источники и отдельные сообщения (см. рис. 14). При этом учитывается два основных параметра отображения — ранг популярности и «свежесть» информации. В рамках этой модели можно наблюдать «дробление» групп источников при увеличении ранга популярности и «свежести» изданий. Когда этот ранг становится достаточно высоким, дробление не позволяет без особых усилий читать названия источников и идентифицировать отдельные документы.

Web-пространство, являясь, пожалуй, самой динамичной частью информационного пространства, характеризуется большим количестве скрытых в нем неявных экспертных оценок, реализованных в виде гиперссылок. В ноябре 1999 года один из руководителей института поиска и анализа текстов, входящего в исследовательское подразделение IBM, Андрей Брёдер (Andrei Broder) и его соавторы из компаний AltaVista, IBM и Compaq математически описали «карту» ресурсов и гиперсвязей. Исследования опровергли расхожее мнение, будто Internet — это единое густое пространство. Проследив с помощью поискового механизма AltaVista свыше 200 млн. Web-страниц и несколько миллиардов ссылок, размещенных на этих страницах, ученые пришли к выводу о структуре Web-пространства как ориентированного графа, в котором вершины соответствуют Web-страницам, а ребра — соединяющим эти страницы гиперссылкам.

Новостная составляющая информационного пространства Интернет сегодня настолько значительна по своим объему и

динамике, что может рассматриваться как мощный информационный поток. Пусть существует система, сканирующая новостную информацию с веб-сайтов сети Интернет (либо любой другой информационной среды) по мере публикации этой информации. Другими словами, на входе такой системы — веб-сайты Интернет, а на выходе — последовательность сообщений, следующих одно за другим по мере публикации. Под информационным потоком понимается дискретный числовой ряд, члены которого соответствуют количеству публикаций за единицу времени, выдаваемых такой системой. Другими словами, анализируется поток элементарных частиц контента. В качестве такой единицы рассматривается документ. Причем поток достаточно неоднородный, который может характеризоваться большим количеством параметров, среди которых выделяются такие, как источники информации (веб-сайт) и тематики. Именно их можно рассматривать, как лежащие на поверхности основы для кластеризации.

Объемы сообщений в тематических информационных потоках образуют временные ряды. Временные ряды, порождаемые тематическими информационными потоками, также обладают фрактальными свойствами и могут рассматриваться как стохастические фракталы. Этот подход расширяет область применения теории фракталов на информационные потоки, динамика которых описывается средствами теории случайных процессов.

Как показано в работах С. А. Иванова, для последовательности сообщений тематических информационных потоков в соответствии с принципом самоподобия, количество сообщений, резонансов на события реального мира пропорционально некоторой степени количества источников информации (кластеров) и итерационно продолжается в течение определенного времени. Точно так же, как и в традиционных научных коммуникациях, растущее множество сообщений в Интернет по одной тематике во времени представляет собой динамическую кластерную систему, возникающую в результате итерационных процессов.

Фрактальная размерность в кластерной системе, соответствующей тематическим информационным потокам, показывает степень заполнения информационного пространства сообщений в течение определенного времени:

$$N = (\varepsilon N_k(t))^D,$$

где N — размер кластерной системы (общее число электронных публикаций в информационном потоке); N_k — размер (число) кластеров (тематик или источников), D — фрактальная размерность информационного массива; ε — коэффициент масштабирования. В приведенном соотношении между количеством сообщений и кластеров проявляется свойство сохранения внутренней структуры множества при изменении масштабов его внешнего рассмотрения. По мнению Иванова, все основные законы научной коммуникации, такие как законы Парето, Лотки, Бредфорда, Зипфа, могут быть обобщены именно в рамках теории стохастических фракталов.

В лекции использованы материалы [6,15].

Задачи и упражнения

1. Докажите свойства фрактала Кантора (см. Лекцию № 2).
2. Найдите фрактальную размерность: А) кривой Коха, снежинки и креста Коха; Б) ковра и решета Серпинского; В) пирамиды Серпинского и губки Менгера; Г) кривой Пеано; Д) кривой и острова Леви; Е) кривой Минковского.
3. Определите размерность фракталов, которые строятся, как указано на рис. 15 Приложения.
4. Определите размерность фракталов, которые строятся, как указано на рис. 16 Приложения.
5. Покажите, что сумма площадей треугольников, «выкинутых» при построении решета Серпинского, равняется площади исходного треугольника.

6. Покажите, что сумма площадей квадратов, «выкинутых» при построении ковра Серпинского, равняется площади исходного квадрата.

7. Являются ли ковер Серпинского и решето Серпинского совершенными множествами? Обоснуйте ответ.

8. Чему равен периметр каркаса ковра и решета Серпинского (т.е. сумма периметров квадратов или треугольников, входящих в множество)?

9. Докажите, что граница снежинки Коха имеет бесконечную длину.

10. Найдите фрактальную размерность модифицированного множества Кантора, в котором на каждом шаге выбрасывается центральная пятая часть каждого интервала.

11. Определите фрактальную размерность обобщенного множества Кантора, в котором на каждом шаге выбрасывается интервал длины $1 - 2r$ с центром в точке $\frac{1}{2}$, где $0 < r < \frac{1}{2}$.

12. Рассмотрите фрактал, который строится, как указано на рис. 17 Приложения. Запишите бесконечный ряд для суммы площадей частей, которые были удалены при построении. Найдите сумму этого ряда. Определите фрактальную размерность данного фрактала.

13. Придумайте пример фрактала, фрактальная размерность которого равна единице.

14. Придумайте множество и найдите его размерность Хаусдорфа а) совпадающую, б) не совпадающую с топологической размерностью.

15. Постройте L-систему для острова Коха, ориентированного внутри треугольника (рис. 18.)

16. Постройте L-систему для острова Леви.

17. Постройте L-систему для кривой Пеано.

18. Докажите утверждение: Если $E \subset F \subset \mathbb{R}^n$, то мера Хаусдорфа удовлетворяет $H^s(E) \leq H^s(F)$, в частности, если $\{F_i\}$ — любой счетный набор непересекающихся борелевских

подмножеств, то $H^s(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} H^s(F_i)$.

19. Пусть F — плоский диск единичного радиуса в пространстве \mathbb{R}^3 . Найдите размерность Хаусдорфа $\dim_H F$.

20. Найдите размерность Хаусдорфа $\dim_H F$, где

а) $F = \left\{ (x, y) : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \right\}$; б) $F = \left\{ (x, y) : \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1 \right\}$;

в) $F = \left\{ (x, y, z) : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, z = 0 \right\}$;

г) $F = \left\{ (x, y, z) : \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1, z = 0 \right\}$;

д) $F = \left\{ (x, y, z) : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \right\}$;

е) $F = \left\{ (x, y, z) : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1 \right\}$.

21. Пусть функция $f(z) = z^3 + c$. Покажите, что если $|c| > 2$, то область притяжения стремится к бесконечности.

22. Что же произойдет в случае увеличения степени в отображении $f(z) = z^k - 1$ при использовании итерационного процесса Ньютона?

Лабораторный практикум

Задание 1. Построение динамических фракталов.

1. Используя компьютер, постройте множество Мандельброта для функции $f(z) = z^k + c$, $k = 3, 4, 5$ и т. д.

2. Используя компьютер, постройте фрактал Жулия для функции $f(z) = z^k + c$, $k = 3, 4, 5$ и т. д.

3. Используя компьютер, постройте фракталы Ньютона для функции а) $f(z) = z^k - c$, $k = 4, 5, 6$; б) $g(z) = z^3 - z$.

4. Постройте динамический фрактал для Вашей функции $f(z)$.

Задание 2. Построение конструктивных фракталов. Напишите программу для построения конструктивных фракталов, используя один из рассмотренных способов.

Задание 3. Вычисление фрактальной размерности. Напишите программу для вычисления фрактальной размерности, используя один из методов для:

- а) конструктивных фракталов;
- б) динамических фракталов;
- в) фрактальных кластеров;
- г) временных рядов.

Задание 4. Исследование временных рядов. Найдите фрактальную размерность некоторого временного ряда (температура, осадки, курсы валют и акций, и др.), используя метод Херста;

метод исключения трендов;

один из методов, рассмотренных в Лекции № 8.

Задание 5. Придумайте и реализуйте на компьютере L-системы, результатом работы которых были бы Ваши собственные варианты фигур:

снежинка или остров (с границей без разрывов);

мозаика или острова (с разрывной границей);

куст или сорняк.

Исследовательские работы

1. Найдите фрактальную размерность границы Удмуртии, повторив эксперимент Ричардсона и используя различные карты и масштабы.

2. Проведите исследование фрактальных свойств кластера какого-нибудь информационного Web-сайта.

3. Проведите исследование какого-нибудь текста, используя закон Дж. Зипфа.

4. Исследование самоподобия информационного пространства с помощью закона Лотки (для распределения авторов) или с помощью закона Брэдфорда (для периодических изданий).

5. Использование закона Парето и распределения Парето.

6. Исследование временных рядов с помощью метода Херста (нормированного размаха).

7. Проверьте, удовлетворяют ли города России по численности населения распределению Дж. Зипфа.

8. Проведите исследование с точки зрения фрактальных свойств: а) растения с перекрестным опылением (соцветия); б) восточного орнамента; в) мозаики; г) фрактальной музыки.

9. Сравнительный анализ способов вычисления фрактальной размерности.

10. Сжатие изображений с помощью фракталов.

11. Клеточные автоматы.

12. Методы исследования документальных информационных потоков.

Темы докладов и рефератов

1. Фракталы Леви и Минковского.
2. Кривые Пеано и Гильберта.
3. Фракталы «Драконы».
4. Фракталы: спирали, звезды, деревья.
5. Построение фракталов с помощью L-систем.
6. Построение фракталов с помощью аффинных преобразований.
7. Связь фракталов и систем счисления.
8. Фракталы Мандельброта и Жулиа.
9. Проблема Кэли и фракталы Ньютона.
10. Системы итерируемых функций.
11. Классическое броуновское движение.
12. Фрактальное броуновское движение.
13. Фракталы в природе.
14. Закон Парето.
15. Распределение Парето.
16. Фрактальный подход в моделировании информационного пространства.
17. Фрактальные свойства для кластеров информационных Web-сайтов.

18. Закон Зипфа как один из основных законов отражающих самоподобие информационного пространства.
19. Законы (закономерности) Лотки и Брэдфорда как уточняющие следствия закона Зипфа.
20. Методы исследования документальных информационных потоков.
21. Приложения фракталов.

Вопросы к зачету

1. Конструктивные фракталы. Способы построения и свойства.
2. Динамические фракталы.
3. Случайные фракталы. Классическое и фрактальное броуновское движение, их свойства.
4. Хаусдорфова мера и размерность Хаусдорфа–Безиковича, их свойства.
5. Различные определения фрактальной размерности: клеточная размерность (емкость), информационная и корреляционная размерности, размерность Ляпунова и др.
6. Свойство самоподобия. Размерность подобия и ее связь с размерностью Хаусдорфа.
7. Метод R/S -анализа Херста и показатель Херста, его связь с фрактальной размерностью.
8. Методы определения фрактальной размерности.
9. Массовые фракталы. Фрактальные кластеры. Модели ограниченной диффузией агрегации.
10. Принцип и распределение Парето.
11. Методы исследования документальных информационных потоков.

Список рекомендуемой литературы

- [1] Анищенко В. С., Вадивасова Т. Е. Лекции по нелинейной динамике. М.-Ижевск: НИЦ «РХД», 2011. 516 с.
- [2] Ватолин Д. С. Алгоритмы сжатия изображений. Методическое пособие: Москва, 1999. 76 с.
- [3] Ватолин Д. С. Использование ДКП для ускорения фрактального сжатия изображений // Программирование. № 3. 1999. С. 51–57.
- [4] Жиков В. В. Фракталы // Соросовский образовательный журнал. Серия «Математика». № 12. 1996. С. 109–117.
- [5] Кроновер Р. М. Фракталы и хаос в динамических системах. М.: Техносфера, 2006. 488 с.
- [6] Ландэ Д. В. Фрактальные свойства тематических информационных потоков из Интернет // Регистрация, хранение и обработка данных. Киев, 2006. Т.8, № 2. С. 93–99. (<http://dwl.kiev.ua/art/frakip/>)
- [7] Мандельброт Б. Фракталы и хаос. Множество Мандельброта и другие чудеса. М.-Ижевск: НИЦ «РХД», 2009. 392 с.
- [8] Мандельброт Б. Фракталы, случай и финансы. М.-Ижевск: НИЦ «РХД», 2004. 256 с.
- [9] Морозов А. Д. Введение в теорию фракталов. Н. Новгород: Изд-во Нижегородского ун-та, 1999. 140 с.
- [10] Шишкин Е. И. Моделирование и анализ пространственных и временных фрактальных объектов. Екатеринбург: Изд-во УрГУ, 2004. 89 с.
- [11] Сэломон Д. Сжатие данных, изображений и звука. М.: Техносфера, 2004. 368 с.
- [12] Уэлстид С. Фракталы и вейвлеты для сжатия изображений в действии. М: Триумф, 2003. 320 с.
- [13] Gerald A.E. Measure, Topology and Fractal Geometry. N.Y.: Springer-Verlag, 1990. 293 p.
- [14] <http://n-t.ru/ri/ch/pi05.htm>
- [15] <http://download.yandex.ru/class/lande/frakt-lecture.pdf>
- [16] <http://www.spsl.nsc.ru/win/Bibliosfera/redkina.pdf>
- [17] <http://fractals.nsu.ru/>
- [18] <http://www.fractalus.com>
- [19] <http://fractalworld.xaoc.ru/>
- [20] <http://www.xaoc.ru>

ПРИЛОЖЕНИЕ

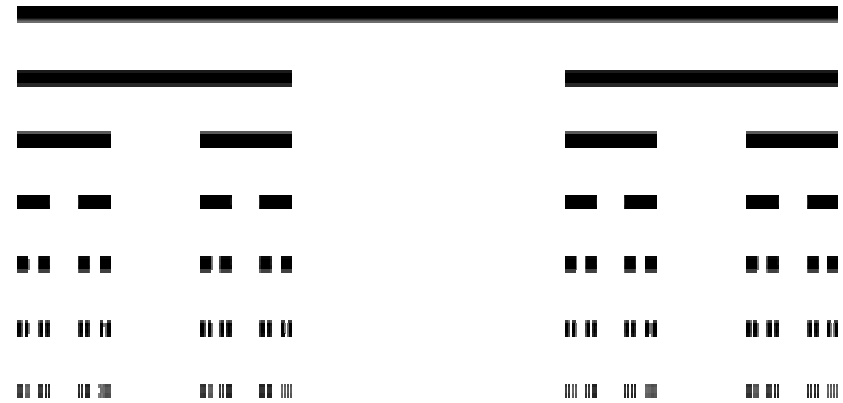


Рис. 1. Множество Кантора

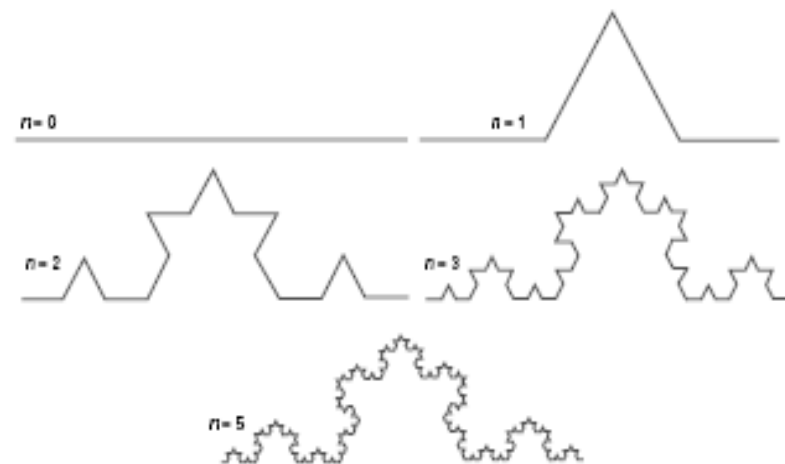


Рис. 2. Построение кривой Коха

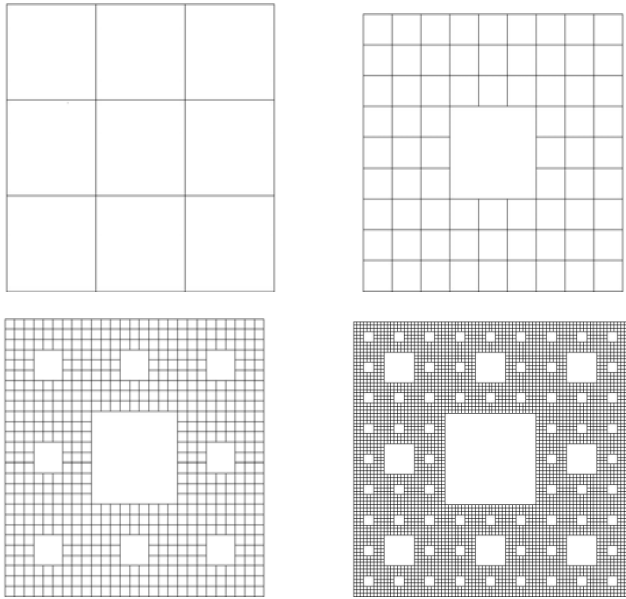


Рис. 3. Построение ковра Серпинского

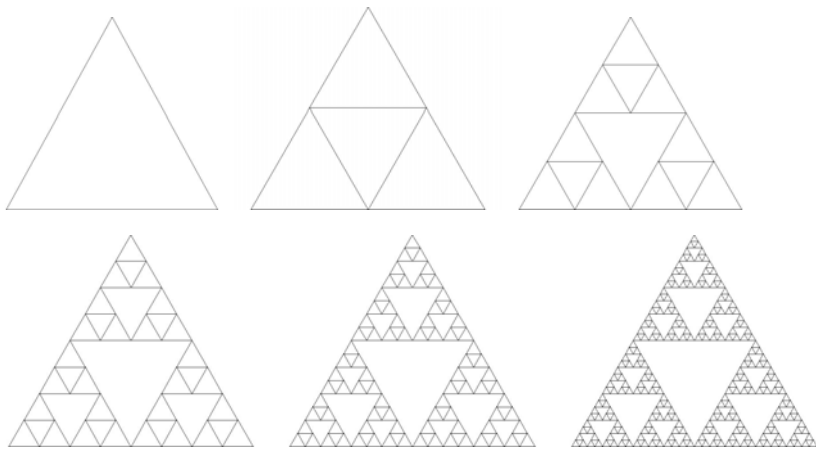


Рис. 4. Построение решета Серпинского

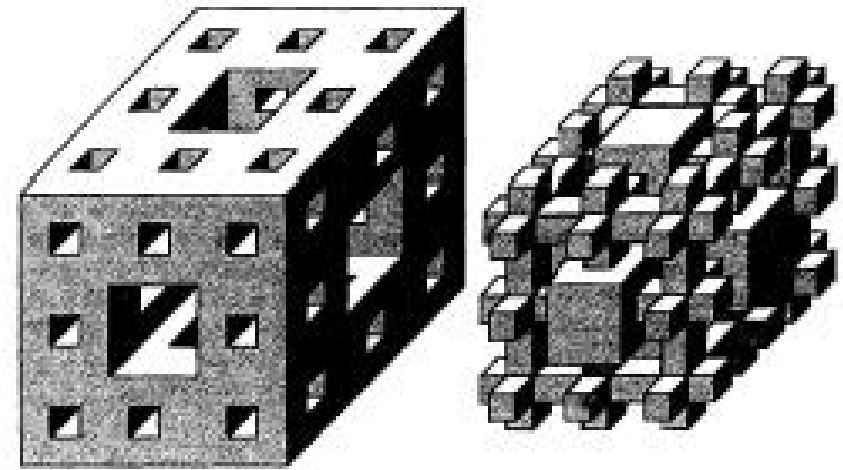


Рис. 5. Губка Менгера

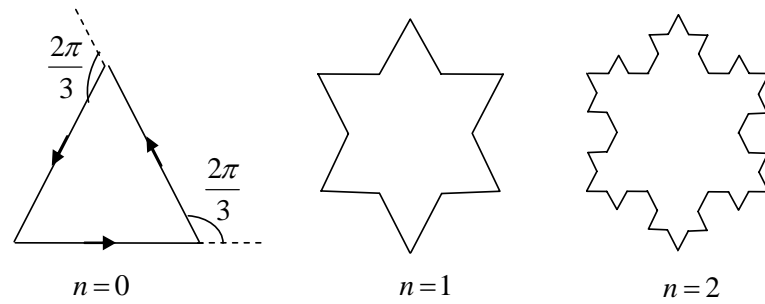


Рис. 6. Построение снежинки Коха с помощью L-системы

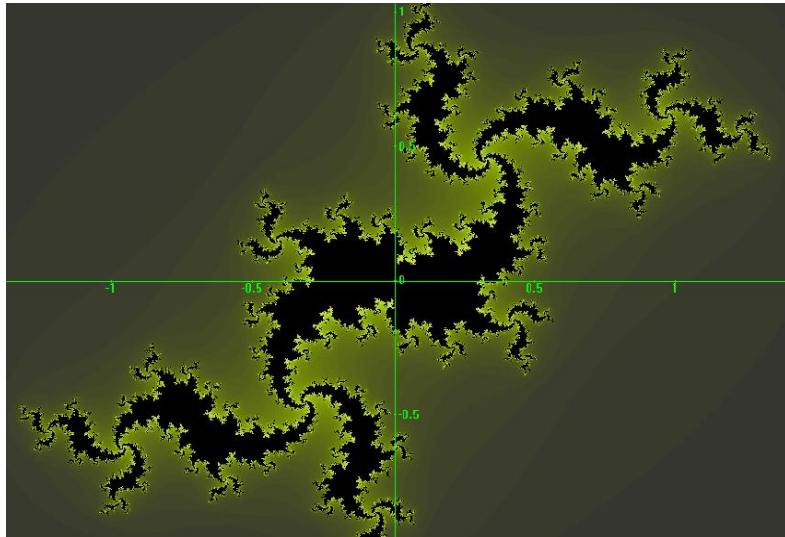


Рис. 7. Фрактал Жулиа

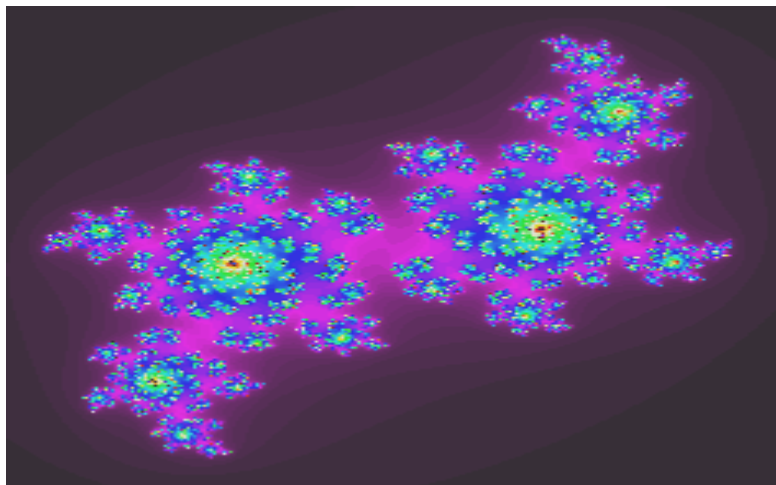


Рис. 8. Фрактал Жулиа (пыль Фату)

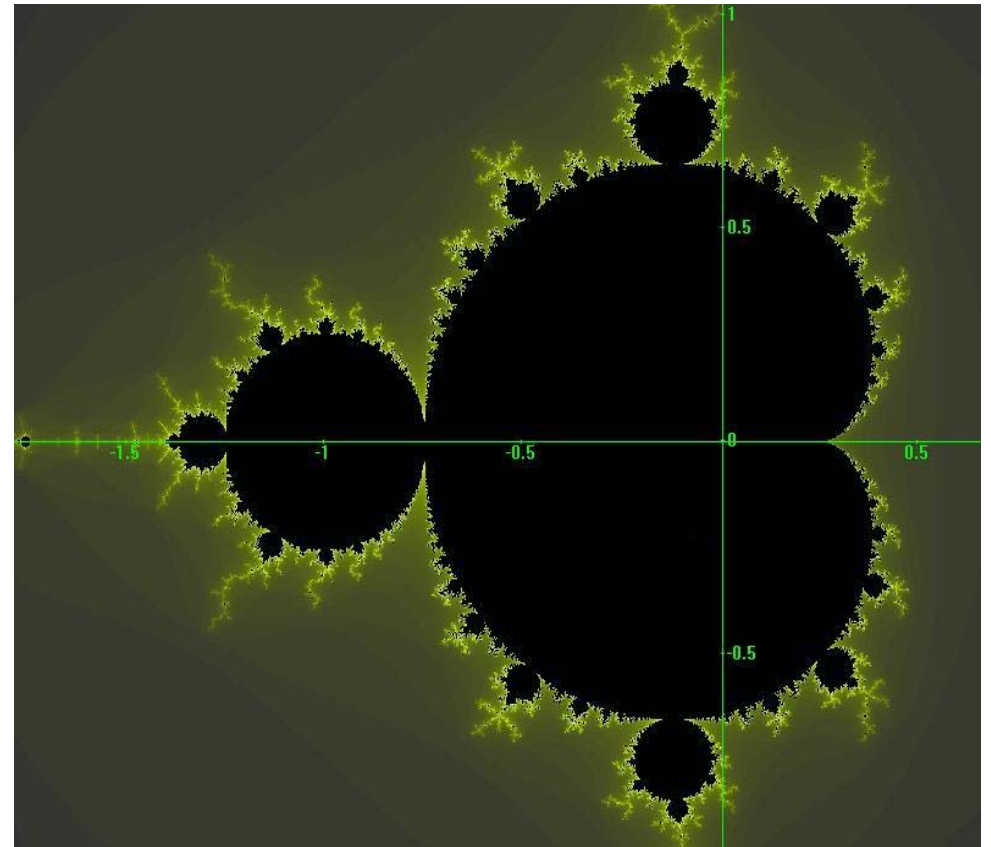


Рис. 9. Множество Мандельброта

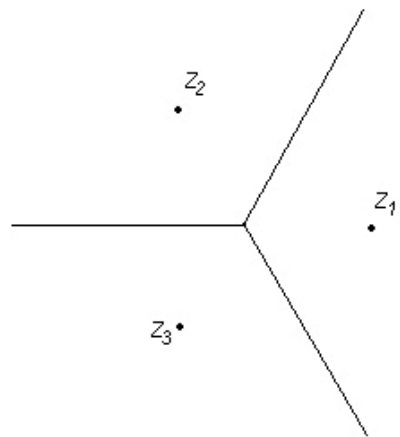


Рис. 10. Корни уравнения

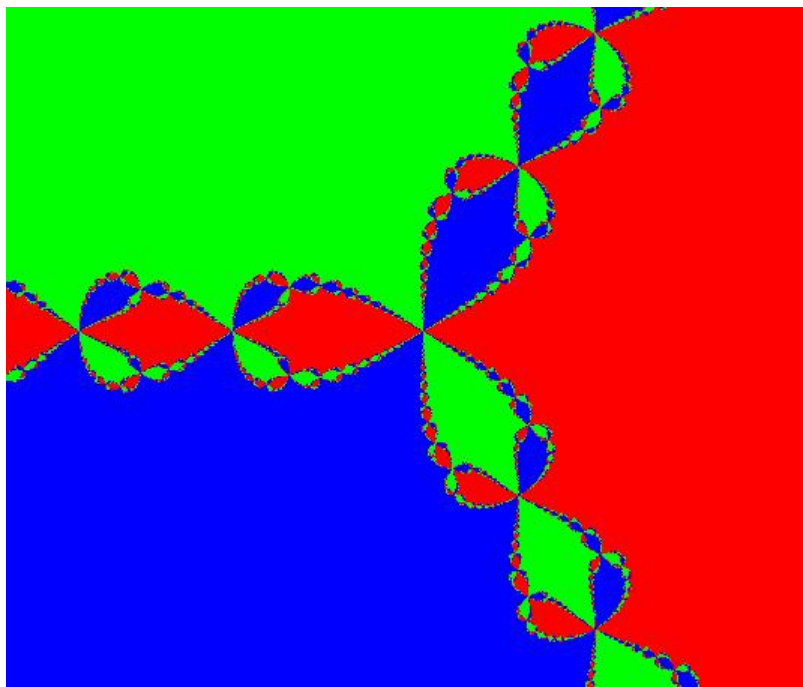


Рис. 11. Фрактал Ньютона



Рис. 12. Фрактальный кластер

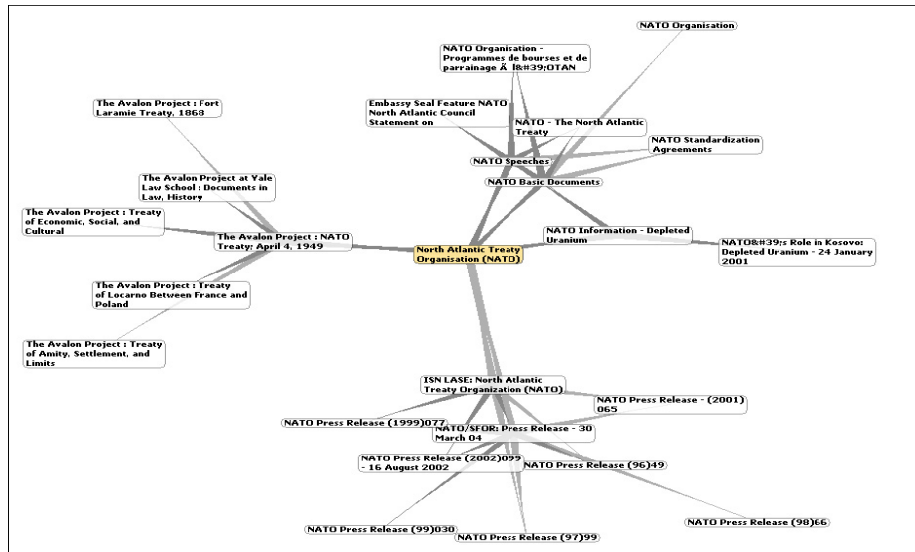


Рис. 13. Объединение Web-сайтов по признаку подобия

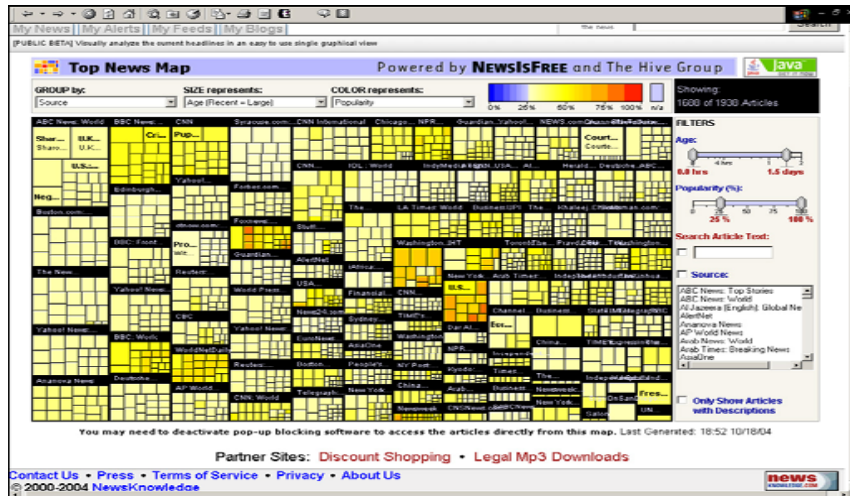


Рис. 14. Кластеры публикаций службы NewsIsFree

Рис. 15. Построения к упр. 3

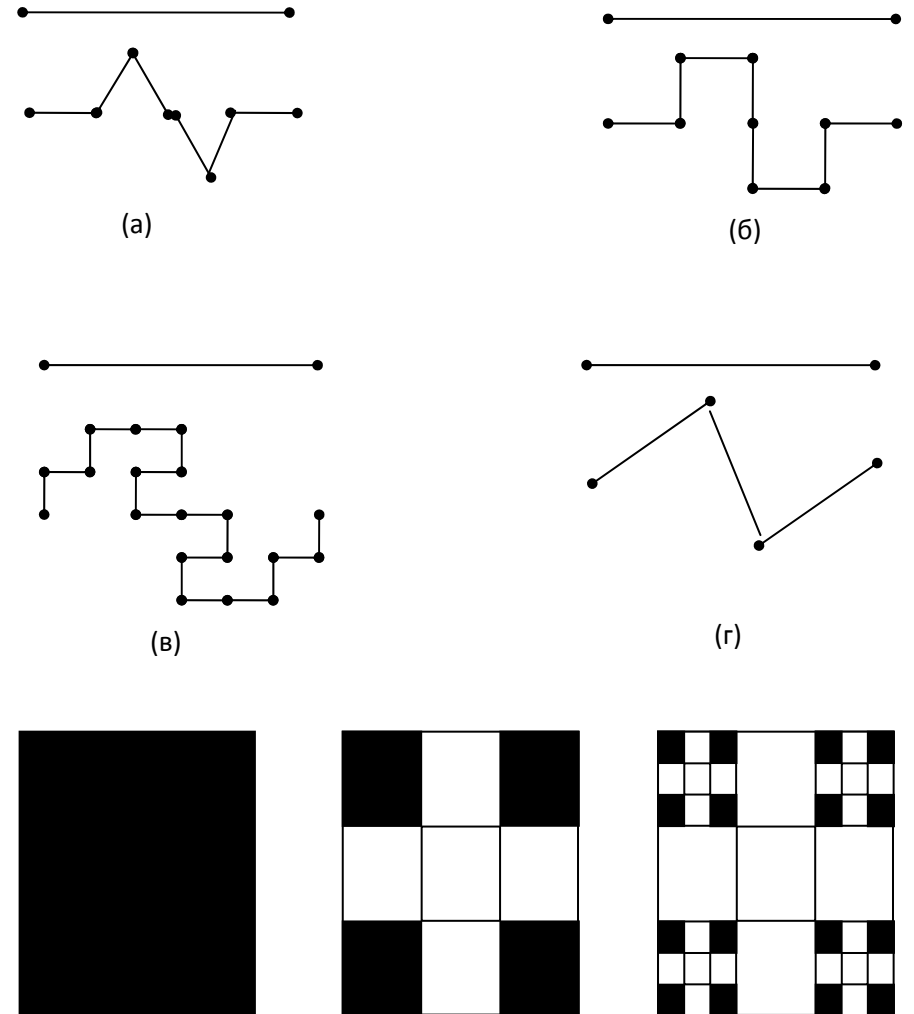


Рис. 16. Построения к упр. 4

Рис. 17. Построения к упр. 12

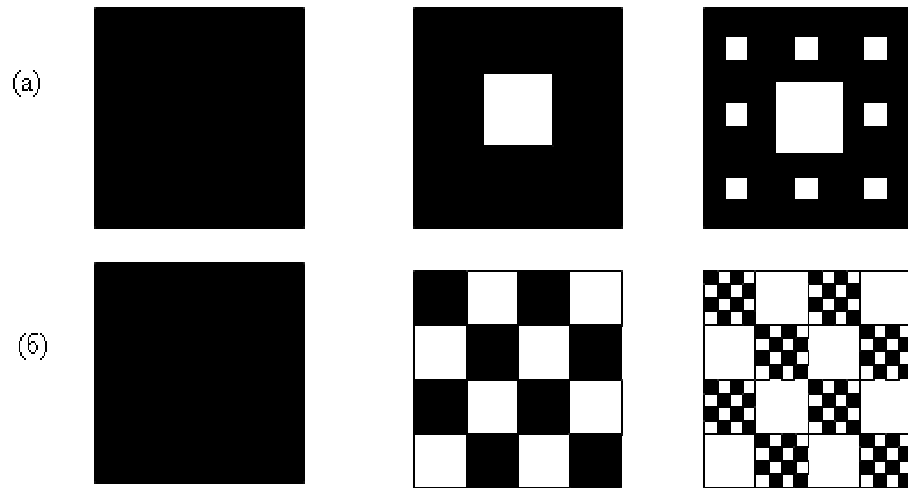
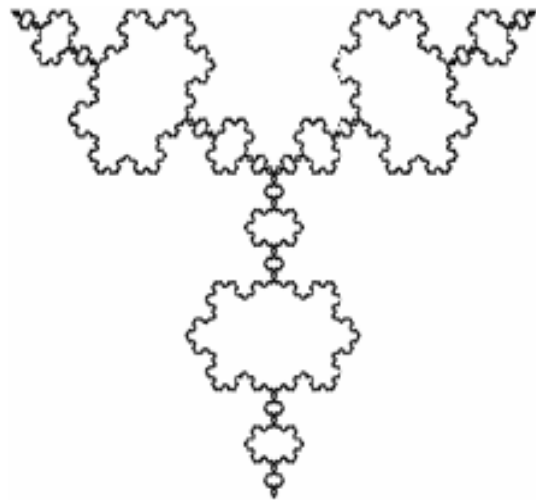


Рис.18. Остров Коха, ориентированный вовнутрь (упр. 15)



Параметры	Изображение
<p><i>Дракон Хартера-Хайтвея</i> axiom=FX порождающие правила: newf = F newx = X+YF+ newy = -FX-Y $\theta = \frac{\pi}{2}$</p>	
<p><i>Треугольник Серпинского</i> axiom = FXF- -FF- -FF порождающие правила: newf = FF newx = - -FXF++FXF++FXF - - $\theta = \frac{\pi}{3}$</p>	
<p><i>Кривая Гильберта</i> axiom = X порождающие правила: newf = F newx = -YF+XFX+FY- newy = +XFX-YFY-FX+ $\theta = \frac{\pi}{2}$</p>	
<p><i>Кривая Госпера</i> axiom = XF newf = F newx = X+YF++YF-FX- -FXFX-YF+ newy = -FX+YFYF++YF+FX- -FX-Y $\theta = \frac{\pi}{3}$</p>	

Учебное издание

Латыпова Наталья Владимировна

Компьютерная обработка данных

Фракталы

Учебное пособие

Компьютерный набор и верстка Н. В. Латыпова

Напечатано в авторской редакции

с оригинал-макета заказчика

Подписано в печать 00.11.12. Формат 60x84 1/16.

Усл.печ.л. 4,3. Уч.-изд.л. 3,9. Тираж 25 экз.

Печать офсетная. Заказ №

Издательство «Удмуртский университет»

426034, г. Ижевск, ул. Университетская, 1, корп. 4.