Алексей Буримов
Андрей Елисеев

Введение в (t, m, s)-сети и алгоритм их генерации

2020

Оглавление

Введение

Конец XX века открыл человечеству двери в цифровую эпоху. Компьютеры проникли во все сферы жизни, их быстрота и многозадачность подкупили своим удобством, а постоянно растущие вычислительные мощности дали новые перспективы развития численных методов – специального раздела математики, рассматривающего вопросы нахождения приближённых ответов на самые разнообразные математические задачи. На практике часто встречаются случаи, в которых аналитическое решение либо невозможно, либо крайне трудоёмко, поэтому, когда стоит задача относительно быстро найти примерный ответ, численные методы со своим внушительным инструментарием вкупе с быстрыми современными вычислителями становятся настоящими спасителями инженеров, физиков, архитекторов и многих других специалистов. Одной из компонент этого инструментария и являются (t, m, s)-сети, о которых пойдёт речь далее.

Первое точное определение (t, m, s)-сетей было дано австрийским алгебраистом Гаральдом Нидеррайтером в 1987 году и являлось логическим продолжением и обобщением результатов, полученных в начале и середине XX века Ильёй Соболем, Йоханнесом ван дер Корпутом, Анри Фором и другими именитыми математиками [1, 2, 4, 7]. Главной мотивацией для изучения (t, m, s)-сетей послужила их непревзойдённая эффективность при решении, в первую очередь, задач многомерного численного интегрирования методом квази-Монте-Карло, однако сегодня их так же успешно применяют и при решении многих других вопросов [11, 13].

Целью данной брошюры является наиболее краткое, полное и, вместе с тем, максимально понятное изложение материала, разделённого структурно на три отдельных главы. В первой главе даны основы теории о (t, m, s)-сетях: наиболее важные термины, определения и базовые свойства. Во второй главе разбирается понятие цифровых (t, m, s)-сетей – самого популярного класса

сетей при вычислениях на компьютере. Третья глава целиком посвящена двум алгоритмам генерации цифровых (t, m, s)-сетей с основанием 2.

Стоит отдельно и заранее отметить, что в настоящем тексте применяется авторская номенклатура, зачастую отличная от оригинальной. Такое обыкновенно непопулярное решение было принято исключительно для того, чтобы сделать обозначения используемых в работе математических объектов как можно более интуитивно понятными. Полная характеристика введённой нотации представлена в разделе «Обозначения», который рекомендован к тщательному ознакомлению перед началом чтения основной части документа. Вводимые в тексте нотацию авторы также использовали при создании программной реализации генератора (t, m, s)-сетей Нидеррайтера с основанием 2, расположенной по адресу

https://github.com/jointpoints/tms-nets

и находящейся в свободном доступе. Техническую документацию к программе можно найти по той же ссылке.

Обозначения

В основе используемой в данном тексте номенклатуры лежит следующий главный принцип:

- латинские буквы используются исключительно для обозначения объектов, связанных с числами;
- греческие буквы используются исключительно для обозначения объектов, связанных с конечными полями.

Под объектами, связанными с числами, понимаются непосредственно числа, векторы и матрицы, состоящие чисел, а также функции, возвращающие числа или числовые векторы и матрицы. Объекты, связанные с конечными полями, определяются аналогично.

Все итерируемые сущности, то есть такие сущности, компоненты которых можно пронумеровать, индексируются, начиная с нуля, а не с единицы, как это диктуется математической традицией. Единственным исключением из этого правила являются оси пространства, в котором происходит непосредственное построение (t, m, s)-сети, – они нумеруются с единицы. Так же нумеруются и координаты векторов, лежащих в этом пространстве, или индексы иных объектов, напрямую привязанных к данным осям. Сами такие номера вдобавок указываются особенным образом – справа от имени соответствующего объекта в квадратных скобках. Номера координат остальных векторов, а также индексы элементов матриц указываются справа от их имён подстрочным шрифтом. Например:

- i-я координата вектора $\boldsymbol{x} \in \mathcal{B}$ обозначается как $x[i] \in \mathcal{B}$, если $\mathcal{B}$ – это пространство, в котором строится сеть. Здесь $i \in \{1,2,\dots,\dim \mathcal{B}\}$;
- i-я координата вектора $\boldsymbol{x} \in \mathcal{B}$ обозначается подстрочным шрифтом как $x_i \in \mathcal{B}$, если $\mathcal{B}$ это не пространство, в котором строится сеть. Здесь $i \in \{0, 1, \dots, \dim \mathcal{B} - 1\}$.

Помимо указанных выше обозначений, в дальнейшем тексте применяются следующие:

- $\mathbb{Z}$ – множество всех целых чисел.
- $\mathbb{Z}_{\geq a}$ – множество всех целых чисел, больших либо равных a.
- $\mathbb{N} = \mathbb{Z}_{\geq 1}$ – множество всех *натуральных чисел*.
- $\mathbb{N}_0 = \mathbb{Z}_{\geq 0}$ – множество *натуральных чисел с нулём*.
- $\mathcal{I} = [0; 1)$ – единичный полуинтервал с открытой правой границей.
- $[a..b]$, где $a, b \in \mathbb{Z}, a \leq b$ – *целочисленный отрезок*, множество всех целых чисел c таких, что $a \leq c \leq b$. На основании данного обозначения также используются альтернативные обозначения для отрезков:
  - $[a\,..\,b) = [a\,..\,b-1]$,
  - $(a\,..\,b] = [a+1\,..\,b]$,
  - $(a\,..\,b) = [a+1\,..\,b-1]$.
- $(n)_b$, где $n \in \mathbb{N}_0$ – представление числа n в позиционной системе счисления с основанием $b \in \mathbb{Z}_{\geq 2}$.
- $(n)_{b,k}$, где $n \in \mathbb{N}_0$ и $k \in \mathbb{N}_0$ – k-й разряд числа n в позиционной системе счисления с основанием b.
- $\mathrm{div}(a, b)$, $\mathrm{mod}(a, b)$, где $a \in \mathbb{N}_0$, $b \in \mathbb{N}$ – операции целочисленного деления a на b и взятия остатка от деления a на b, соответственно.
- $\bigoplus\colon \mathbb{N}_0 \times \mathbb{N}_0 \longrightarrow \mathbb{N}_0$ – оператор *исключающей дизъюнкции* такой, что
$$a \oplus b = c \quad \Leftrightarrow \quad (a)_{2,k} \oplus (b)_{2,k} = (c)_{2,k} \;\; \forall k \in \mathbb{N}_0$$
- Под *последовательностью* элементов множества $\mathcal{A}$ будет пониматься функция из $\mathbb{N}_0$ в $\mathcal{A}$.
- Под *мультимножеством точек*, принадлежащих $\mathcal{A}$, будет пониматься функция из $\mathcal{A}$ в $\mathbb{N}$, сопоставляющая каждой точке из $\mathcal{A}$ количество её включений в *мультимножество точек*.
- Векторные величины, обозначаемые строчными буквами, выделяются полужирным начертанием. Например:

- $\boldsymbol{x} \in \mathcal{I}^s$ – вектор, каждая координата которого принадлежит $\mathcal{I}$,
- $\boldsymbol{\xi} \in \mathbb{F}_2^m$ – вектор, каждая координата которого принадлежит $\mathbb{F}_2$.

Как синоним слова *вектор*, мы также будем использовать слово *точка*. Любой вектор в тексте рассматривается, как вектор-столбец.

- Матрицы обозначаются заглавными буквами с обычным начертанием. Например:
  - $\Gamma \in \mathbb{F}_2^{m \times n}$ – матрица из m строк и n столбцов, каждый элемент которой принадлежит $\mathbb{F}_2$.
- Γ^k, Γ_j – k-й столбец и j-я строка матрицы Γ, соответственно.

1. (t, m, s)-сети

Суть численного метода Монте-Карло можно вкратце изложить следующим образом. Пусть есть интегрируемая функция f, зависящая от s независимых переменных. Тогда определённый интеграл f по некоторой ограниченной области $\mathcal{B} \subset \mathbb{R}^s$ можно приближённо найти, заполнив $\mathcal{B}$ равномерно распределёнными по ней точками $\boldsymbol{x}_n$ и рассчитав среднее арифметическое значений $f(\boldsymbol{x}_n)$. Интуитивно понятно, что чем большим числом точек заполняется данная область, тем ближе расчётное приближение к истинному ответу, однако существует значительно более экономный метод достижения высокой точности. Этот метод, получивший название «квази-Монте-Карло», состоит в наложении на точки $\boldsymbol{x}_n$ дополнительных ограничений, которым, в частности, удовлетворяют (t, m, s)-сети.

Фундаментальная теория (t, m, s)-сетей рассматривается при $\mathcal{B} = \mathcal{I}^s$, однако с помощью аффинных преобразований легко обобщается на любые кубоиды в $\mathbb{R}^s$, что позволяет применять её к обширному кругу практических задач. Существуют и более сложные случаи, где в качестве $\mathcal{B}$ принимаются сферы, шары или симплексы, которые рассмотрены, например, в [9, 10].

Основой понятия (t, m, s)-сети является элементарный интервал.

ОПРЕДЕЛЕНИЕ 1.1

Пусть $b \in \mathbb{Z}_{\geq 2}$, $s \in \mathbb{N}$, а также $\boldsymbol{d} \in \mathbb{N}_0^s$, $\boldsymbol{a} \in \mathbb{N}_0^s$ и $a[i] < b^{d[i]}$ для всех $i \in [1..s]$. Тогда *элементарным интервалом с основанием* b называется множество

$$\mathcal{E} = E(b, \boldsymbol{a}, \boldsymbol{d}) = \bigtimes_{i=1}^{s} \left[\frac{a[i]}{b^{d[i]}}, \frac{a[i]+1}{b^{d[i]}} \right) \subset \mathcal{I}^s.$$

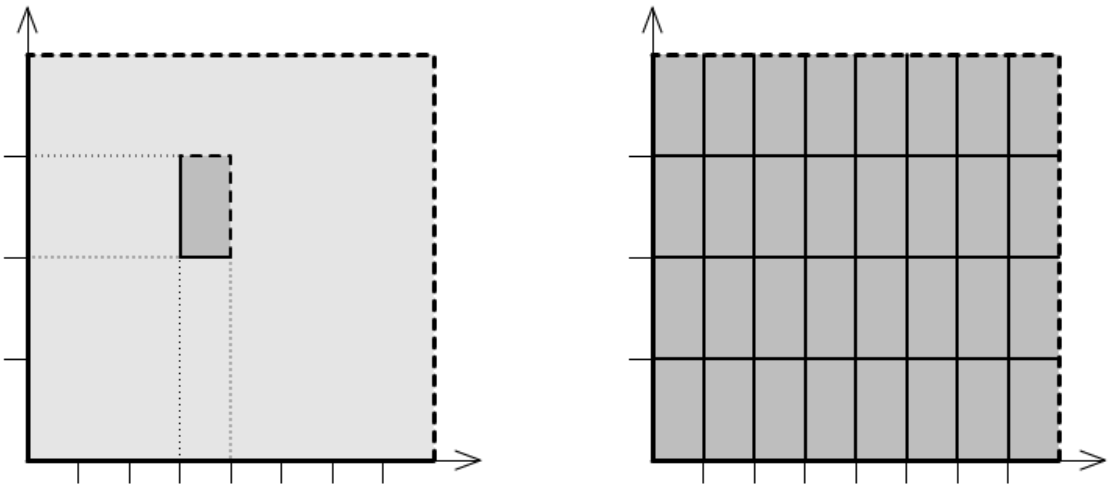
Из этого определения можно понять, что элементарный интервал представляет собой прямоугольный параллелепипед в s-мерном пространстве, для которого верны следующие свойства:

1. его длина в смысле Лебега вдоль каждой i-й компоненты равна $\frac{1}{b^{d[i]}}$;

2. любые два различных элементарных интервала с равными параметрами b и $\boldsymbol{d}$ не имеют общих точек;

3. объединение всех существующих элементарных интервалов с равными параметрами b и $\boldsymbol{d}$ совпадает с $\mathcal{I}^s$.

Из свойств 2 и 3 непосредственно следует, что все существующие элементарные интервалы с равными параметрами b и $\boldsymbol{d}$ образуют *разбиение* $\mathcal{I}^s$ на «клеточки» – непересекающиеся кубоидные области одинаковой формы, что показано на рисунке 1.1(б).



(а) (б)

Рисунок 1.1. (а) Двумерный элементарный интервал с основанием 2 и параметрами $\boldsymbol{a} = \boldsymbol{d} = (3, 2)$, расположенный внутри $\mathcal{I}^2$; (б) $\mathcal{I}^2$, разбитый на части элементарными интервалами с основанием 2 и фиксированным параметром $\boldsymbol{d} = (3, 2)$.

Зная свойство 1, не составит труда также определить s-мерный *объём* элементарного интервала $\mathcal{E}$ в смысле Лебега:

$$V(\mathcal{E}) = \prod_{i=1}^{s} \frac{1}{b^{d[i]}} = \prod_{i=1}^{s} b^{-d[i]} = b^{-d[1]-d[2]-\cdots-d[s]} = b^{-\sum_{i=1}^{s} d[i]}. \qquad (1.1)$$

Заметим, что $V(\mathcal{E})$ зависит от параметров интервала b и $\boldsymbol{d}$, но не зависит от параметра $\boldsymbol{a}$, следовательно, в каждом упомянутом только что разбиении $\mathcal{I}^s$ все элементарные интервалы имеют одинаковый объём.

Зная всё вышеописанное, перейдём к рассмотрению главного объекта в данной теме, а именно (t, m, s)-сети.

ОПРЕДЕЛЕНИЕ 1.2

Пусть $t \in \mathbb{N}_0$, $m \in \mathbb{N}_0$, $s \in \mathbb{N}$, $b \in \mathbb{Z}_{\geq 2}$ причём $t \leq m$. Тогда мультимножество точек $\mathcal{P} \subset \mathcal{I}^s$ является (t, m, s)-*сетью с основанием* b, если в любом элементарном интервале $\mathcal{E}$ объёма $V(\mathcal{E}) = b^{t-m}$ содержится ровно b^t точек из $\mathcal{P}$ (включая повторы).

Сразу отметим несколько важных свойств (t, m, s)-сетей:

1. каждая (t, m, s)-сеть состоит ровно из b^m точек (включая повторы);
2. каждая (t, m, s)-сеть является (u, m, s)-сетью с тем же основанием, если $t < u \leq m$ (истинность первых двух утверждений легко доказать, воспользовавшись определением $V(\mathcal{E})$, а также свойствами 2 и 3 элементарных интервалов с предыдущей страницы);
3. каждая (t, m, s)-сеть обладает свойством *однородности* (low discrepancy).

Однородность, говоря простым языком, гарантирует то, что между элементами множества (или мультимножества) нет «больших пустот», что, в свою очередь, и обуславливает быструю сходимость метода квази-Монте-Карло. В случае (t, m, s)-сетей данное поведение достигается за счёт равенства числа точек в каждом элементарном интервале фиксированного объёма. Нидеррайтер, Пилихсхэммер и другие учёные вводили специальные функции, с помощью которых можно определить «уровень однородности» для каждого конкретного набора точек. Такие функции, называемые *нормами неравномерности* (discrepancies) обладают, в первую очередь, важным теоретическим значением. Так, например, именно благодаря им было

установлено, что меньшие значения параметра t соответствуют более однородным сетям (чем меньше t, тем «качественнее» точки рассредоточены по $\mathcal{I}^s$). На интуитивном уровне в этом можно убедиться самостоятельно, рассмотрев определение (t, m, s)-сети и зафиксировав в нём все значения, кроме t. За более подробной информацией на тему популярных норм неравномерности D, D^*, L_p и L_p^* можно обратиться, например, в [11].

Ниже, на рисунке 1.2, показано несколько примеров $(t, m, 2)$-сетей с разными параметрами t, m и b.

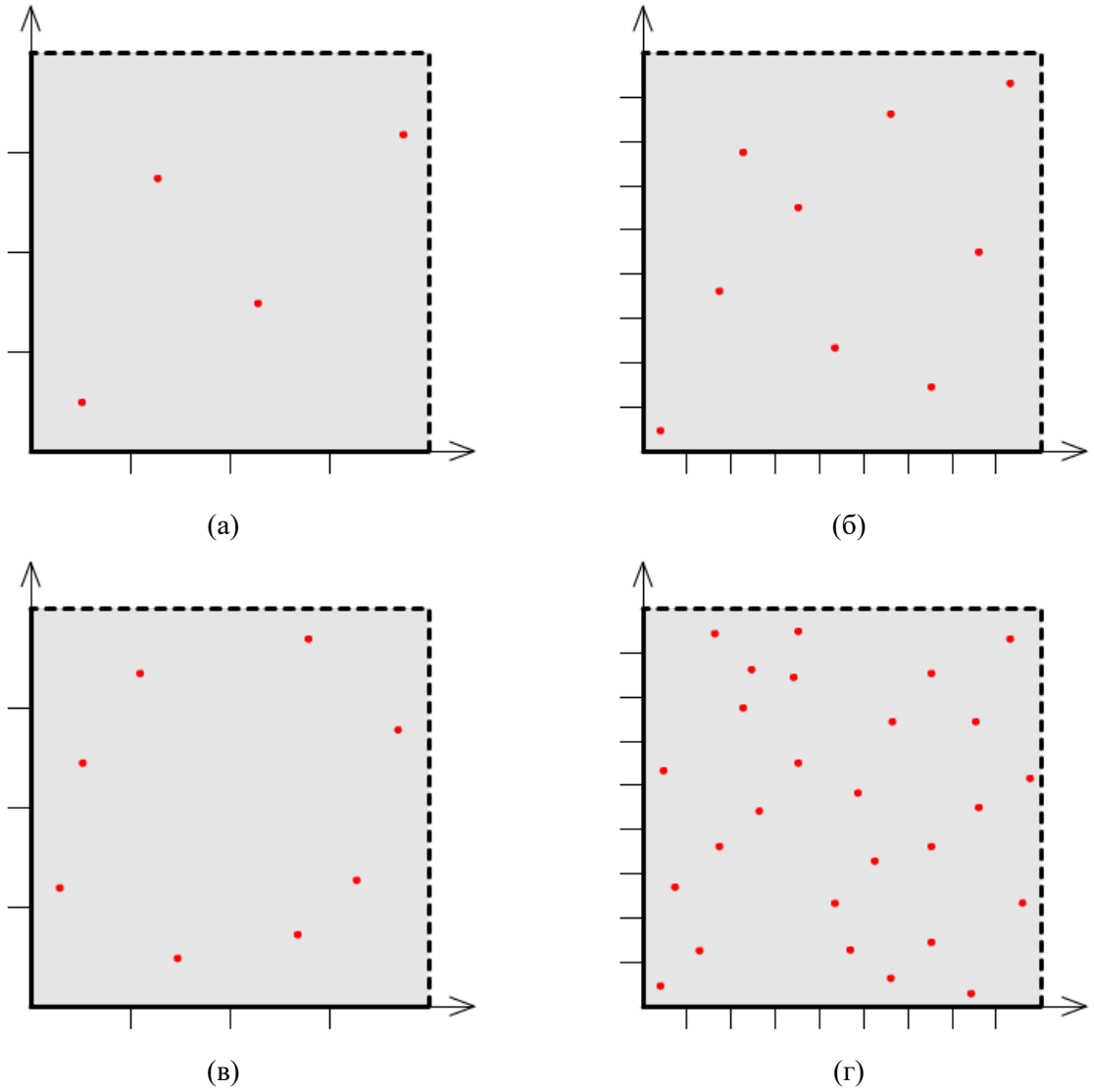


Рисунок 1.2. Примеры $(t, m, 2)$-сетей в двумерном кубе $\mathcal{I}^2$ с разными параметрами:
(а) $(0,2,2), b = 2$, (б) $(0,2,2), b = 3$,
(в) $(1,3,2), b = 2$, (г) $(1,3,2), b = 3$.

Для наглядности покажем, что множество $\mathcal{P}$, изображённое красными точками, например, на рисунке 1.2(в), действительно является $(1,3,2)$-сетью с основанием 2. Здесь даны параметры:

$$t = 1 \quad m = 3 \quad s = 2 \quad b = 2.$$

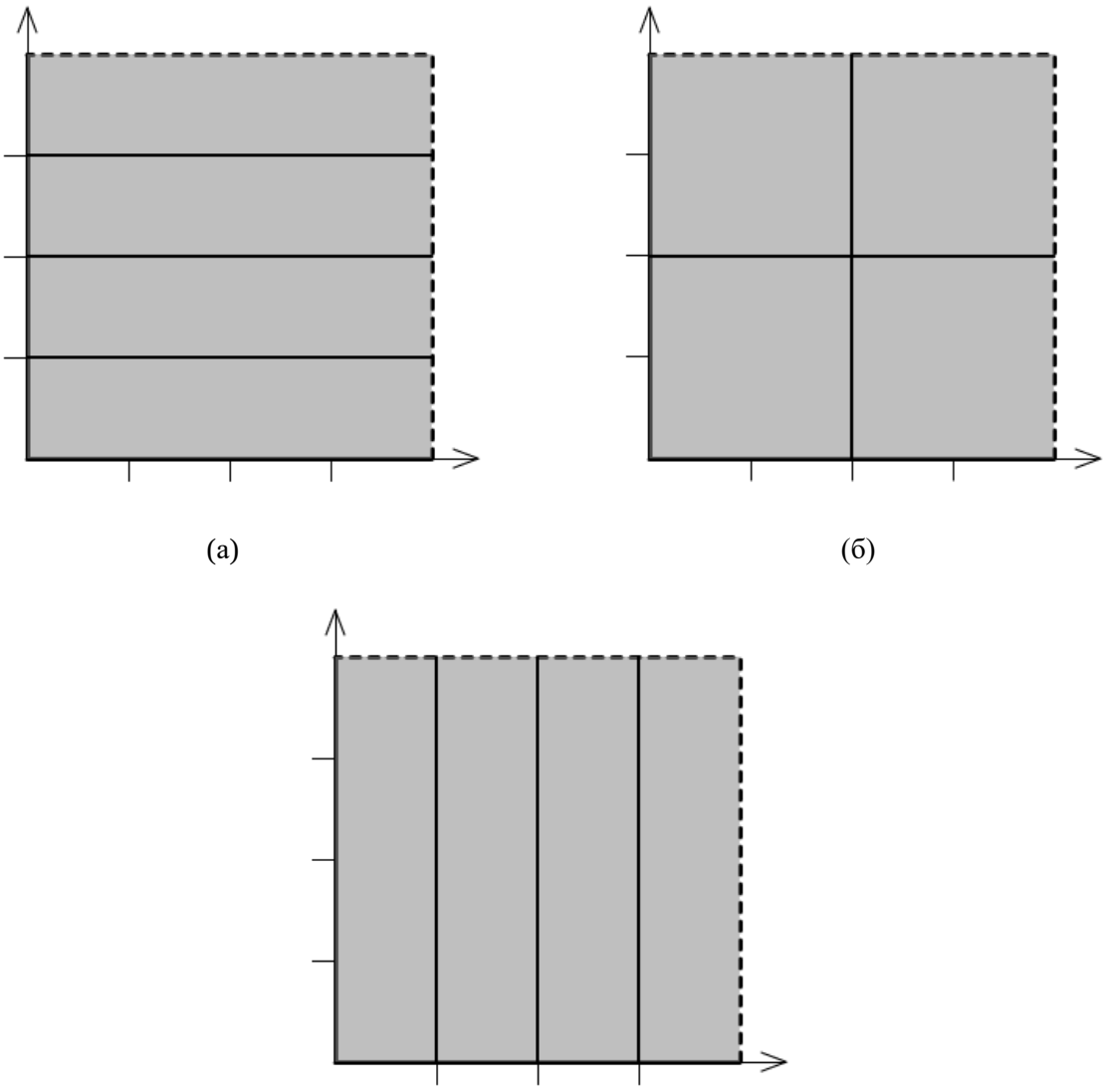
По определению, чтобы множество $\mathcal{P}$ было (t, m, s)-сетью, необходимо, чтобы в каждом элементарном интервале объёма b^{t-m} содержалось ровно b^t точек множества $\mathcal{P}$. Подставим указанные выше числовые значения в определение (t, m, s)-сети. Получим, что для рассматриваемого частного случая каждый элементарный интервал объёма $2^{1-3} = 2^{-2} = 1/4$ должен содержать ровно $2^1 = 2$ точки.

Рассмотрим все существующие элементарные интервалы объёма $1/4$ и проверим, что для них указанные ограничения выполняются. Вспомним, что каждый элементарный интервал однозначно определяется тремя параметрами: b, $\boldsymbol{d}$ и $\boldsymbol{a}$. Параметр b уже известен: $b = 2$, следовательно, остаётся определить все $\boldsymbol{d}$ и $\boldsymbol{a}$ такие, что $V\big(E(2, \boldsymbol{a}, \boldsymbol{d})\big) = 1/4$. Вспомним также, что объём элементарного интервала зависит только от параметров b и $\boldsymbol{d}$. Подставляя известное b в формулу (1.1) (стр. 9), нетрудно видеть, что выражение $V(\mathcal{E}) = 1/4$ эквивалентно выражению $\sum_{i=1}^{2} d[i] = 2$. Отсюда получаем три возможных значения для параметра $\boldsymbol{d}$:

$$\boldsymbol{d} = [0, 2]^T,$$
$$\boldsymbol{d} = [1, 1]^T,$$
$$\boldsymbol{d} = [2, 0]^T.$$

Чтобы избежать явного нахождения всех допустимых $\boldsymbol{a}$ для каждого из трёх случаев и рассмотрения каждого элементарного интервала по отдельности, воспользуемся тем, что все элементарные интервалы с равными параметрами b и $\boldsymbol{d}$ образуют разбиение $\mathcal{I}^s$.

На рисунке 1.3 показаны разбиения двумерного куба $\mathcal{I}^2$ элементарными интервалами с фиксированными параметрами $b = 2$ и $\boldsymbol{d}$.

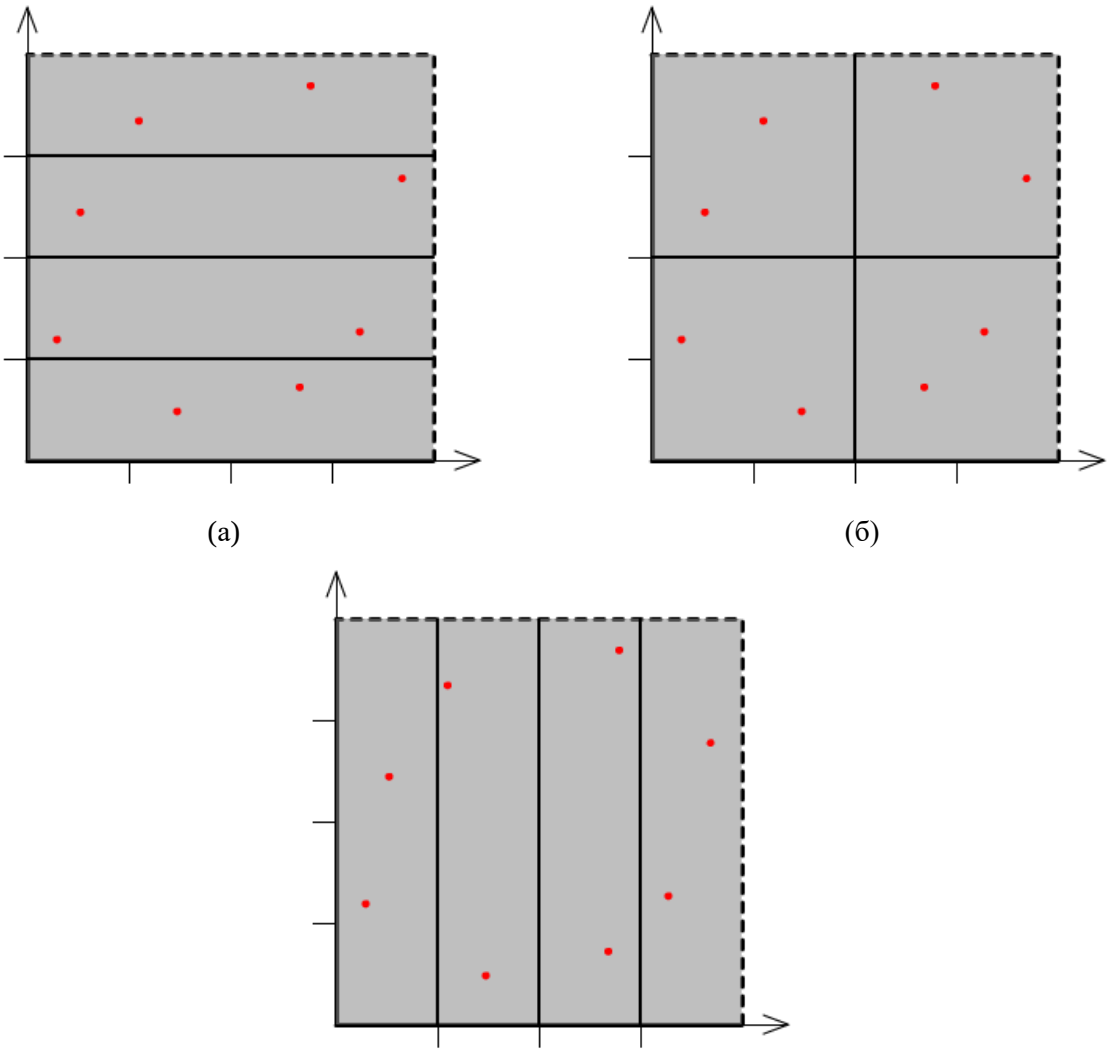


(а) (б)

(в)

Рисунок 1.3. Разбиения двумерного куба $\mathcal{I}^2$ элементарными интервалами объёма $1/4$ с фиксированными параметрами $b = 2$ и $\boldsymbol{d}$, где
(а) $\boldsymbol{d} = [0, 2]^T$, (б) $\boldsymbol{d} = [1, 1]^T$, (в) $\boldsymbol{d} = [2, 0]^T$.

Наложим точки множества $\mathcal{P}$ на полученные разбиения. Результат представлен на рисунке 1.4.



(а) (б)

(в)

Рисунок 1.4. Множество $\mathcal{P}$ и разбиения двумерного куба $\mathcal{I}^2$ элементарными интервалами объёма $1/4$ с фиксированными параметрами $b = 2$ и $\boldsymbol{d}$, где
(а) $\boldsymbol{d} = [0, 2]^T$, (б) $\boldsymbol{d} = [1, 1]^T$, (в) $\boldsymbol{d} = [2, 0]^T$.

Как видно, во всех элементарных интервалах оказалось ровно две точки, следовательно, множество $\mathcal{P}$ – $(1,3,2)$-сеть с основанием 2 по определению.

Отметим, что согласно свойству 2 (t, m, s)-сетей (стр. 10), множество $\mathcal{P}$ с таким же успехом будет и $(2,3,2)$-сетью, и $(3,3,2)$-сетью, но оно, однако, не будет $(0,3,2)$-сетью, в чём можно убедиться, проведя аналогичную проверку для $t = 0$. Зная о наличии обратной зависимости между значением t и уровнем однородности (t, m, s)-сети, можно заключить, что

параметр $t = 1$ «наилучшим» образом среди всех $t \leq m$ (то есть среди всех $t \leq 3$) описывает однородность данной сети.

ОПРЕДЕЛЕНИЕ 1.3

Наименьшее $t \in \mathbb{N}_0$, при котором данное мультимножество $\mathcal{P}$ является (t, m, s)-сетью с некоторым основанием, называется *дефектом* этой сети.

Таким образом, $t = 1$ является дефектом сети $\mathcal{P}$ из только что разобранного случая. Выполнив несколько проверок определения для каждого примера из рисунка 1.2 (стр. 11), можно увидеть, что все обозначенные в нём значения t на самом деле будут и дефектами для соответствующих сетей.

2. Цифровые (t, m, s)-сети над конечными полями

Обратим внимание на тот факт, что до сих пор не было никаких правил, которые обязывали бы элементы (t, m, s)-сетей иметь какую-либо аналитическую взаимосвязь друг с другом. Так, например, обратившись ещё раз к рисунку 1.2 (стр. 11), можно заметить, что точки во всех приведённых примерах расположены достаточно хаотично – при подготовке иллюстраций их расположение подбиралось «на глаз». Точно так же можно было сгенерировать некоторое множество случайных или псевдослучайных s-мерных точек и небезосновательно ожидать, что с какой-то ненулевой вероятностью это множество окажется (t, m, s)-сетью для каких-нибудь t, m и b.

Очевидно, что подходы «на глаз» и «с какой-то ненулевой вероятностью» далеко не самые привлекательные, когда речь заходит о решении конкретных прикладных задач. К счастью, для (t, m, s)-сетей уже были разработаны способы их детерминированной генерации, в частности, при помощи рассмотрения особого класса сетей, характеризующегося наличием функциональных выражений координат для каждой точки. Такие сети называются *цифровыми* (t, m, s)-сетями, и их определение основывается на теории колец. Расчёт точек цифровых сетей по большей части производится над кольцами, мощность которых совпадает с величиной основания (t, m, s)-сети. На практике производить вычисления над кольцами общего вида не удобно, ввиду чего обычно используется частный подкласс цифровых сетей, основные расчёты точек которых производятся в конечных полях. Сети из этого подкласса называются *цифровыми* (t, m, s)*-сетями над конечными полями*, и в дальнейшем мы будем описывать именно их. Подробное описание цифровых сетей в общем виде можно найти в [8].

Прежде чем перейти к строгому определению, опишем некоторые общие идеи и замечания:

- Точки любой цифровой (t, m, s)-сети с основанием b полагаются пронумерованными, то есть каждой точке такой сети однозначно сопоставлен номер $n \in [0\,..\,b^m)$. Таким образом, мультимножество точек $\mathcal{P}$, являющееся сетью, имеет вид $\{\boldsymbol{x}_n\}_{n\in[0..b^m)}$, где $\boldsymbol{x}_n \in \mathcal{I}^s$;
- Основание цифровой сети над конечным полем может равняться только степени простого числа. Это обусловлено аналогичным свойством порядка конечных полей;
- Любое $n \in [0\,..\,b^m)$ однозначно представимо в b-ичной системе счисления. Следовательно, любому n можно сопоставить упорядоченный набор разрядов из его выражения в определённой системе счисления.

Исходя из последнего замечания, введём два вспомогательных понятия.

ОПРЕДЕЛЕНИЕ 2.1

а. *m-разрядной векторизацией числа по основанию b* называется отображение, ставящее числу $n \in \mathbb{N}_0$ в соответствие m-мерный вектор, k-я координата которого равна k-му разряду в разложении n по основанию b:

$$\mathbf{vec}_{b,m}(n) = \left[(n)_{b,0}, (n)_{b,1}, (n)_{b,2}, \ldots, (n)_{b,m-1}\right]^T \in [0\,..\,b)^m;$$

б. *Реверсивной нумеризацией вектора $\boldsymbol{v} \in [0\,..\,b)^m$ по основанию b* называется операция построения целого числа $n \in [0\,..\,b^m)$ такого, что в разложении в b-ичной системе счисления k-й разряд $(n)_{b,k}$ равен $(m-k-1)$-й координате v_k, а разряды большие m полагаются равными нулю:

$$\mathrm{rnum}_b(\boldsymbol{v}) = \sum_{k=0}^{m-1} v_{m-k-1} \cdot b^k \in [0\,..\,b^m).$$

Например, $32 = (1012)_3$ и $64 = (2101)_3$, следовательно, $\mathbf{vec}_{3,4}(32) = = [2,1,0,1]^T$, $\mathbf{vec}_{3,6}(32) = [2,1,0,1,0,0]^T$, $\mathbf{vec}_{3,2}(32) = [2,1]^T$, а $\mathrm{rnum}_3([2,1,0,1]^T) = 64$.

Наконец, сформулируем определяющий критерий цифровой (t, m, s)-сети над конечным полем $\mathbb{F}_b$.

ОПРЕДЕЛЕНИЕ 2.2

(t, m, s)-сеть с основанием b равным степени простого числа является *цифровой* (t, m, s)*-сетью над конечным полем* $\mathbb{F}_b$, если существуют

1. Биекции

$$\phi : [0 \,..\, b) \leftrightarrow \mathbb{F}_b,$$
$$f : \mathbb{F}_b \leftrightarrow [0 \,..\, b),$$

порождающие функции

$$\boldsymbol{\phi} : [0 \,..\, b)^m \leftrightarrow \mathbb{F}_b^m,$$
$$\boldsymbol{f} : \mathbb{F}_b^m \leftrightarrow [0 \,..\, b)^m$$

следующим образом:

$$\boldsymbol{\phi}(\boldsymbol{v}) = [\phi(v_0), \phi(v_1), \dots, \phi(v_{m-1})]^T,$$
$$\boldsymbol{f}(\boldsymbol{\xi}) = [f(\xi_0), f(\xi_1), \dots, f(\xi_{m-1})]^T;$$

2. Матрицы $\Gamma[i] \in \mathbb{F}_b^{m \times m}$, где $i \in [1..s]$, называемые *генерирующими*, такие, что любая координата произвольной точки сети с номером n может быть задана выражением:

$$x_n[i] = \mathrm{rnum}_b \left(\boldsymbol{f} \left(\Gamma[i] \cdot \boldsymbol{\phi} \left(\mathbf{vec}_{b,m}(n) \right) \right) \right) \cdot b^{-m}.$$

Несмотря на частность данного определения относительно общего понятия (t, m, s)-сети, оно всё равно может показаться довольно громоздким. Однако главное, что следует извлечь из него, – это общие представления о ранее не раскрывавшейся математической основе, позволяющей строить (t, m, s)-сети.

На практике при программировании различных алгоритмов построения сетей как правило рассматриваются ещё более частные виды (t, m, s)-сетей, а именно цифровые (t, m, s)-сети над *простыми* конечными полями $\mathbb{F}_b$, то есть над полями с простым основанием b. Такие сети

замечательны тем, что конечные поля, ассоциированные с ними, по построению являются множествами целых чисел $[0\,..\,b) = \{0, 1, 2, \dots, b-1\}$, арифметические операции сложения и умножения над которыми производятся по модулю b. Данное обстоятельство серьёзно упрощает задачу тем, что требуемыми в определении цифровой (t, m, s)-сети (стр. 18, 2.2) биекциями ϕ и f становится допустимо пренебречь, так как они, по сути, оказываются перестановками целых чисел от 0 до $b-1$:

$$\phi : [0..b) \leftrightarrow \mathbb{F}_b \quad \Leftrightarrow \quad \phi : [0..b) \leftrightarrow [0..b),$$
$$f : \mathbb{F}_b \leftrightarrow [0..b) \quad \Leftrightarrow \quad f : [0..b) \leftrightarrow [0..b),$$

которые, в свою очередь, для удобства можно принять тривиальными:

$$\phi(v_k) = v_k,$$
$$f(\xi_k) = \xi_k.$$

Таким образом, для задания сетей над простыми конечными полями достаточно определить генерирующие матрицы $\Gamma[i] \in \mathbb{F}_b^{m\times m}$, $i \in [1\,..\,s]$, с помощью которых любая координата точки сети с номером $n \in [0\,..\,b^m)$ выражалась бы упрощённой формулой:

$$x_n[i] = \mathrm{rnum}_b\left(\Gamma[i] \cdot \mathbf{vec}_{b,m}(n)\right) \cdot b^{-m}. \qquad (2.1)$$

Отсюда, описания разнообразных способов построения сетей сводятся к описанию разнообразных способов построения соответствующих генерирующих матриц.

Нидеррайтер в статье [5] представил один из таких способов, который мы далее рассмотрим в частном случае для $b = 2$. Ввиду этого договоримся, что в дальнейшем под цифровой (t, m, s)-сетью будет пониматься цифровая (t, m, s)-сеть над полем $\mathbb{F}_2 = (\{0, 1\} \subset \mathbb{Z},\ \oplus,\ \cdot\ )$.

3. (t, m, s)-сети Нидеррайтера с основанием 2

3.1. Алгоритм генерации

Алгоритм построения генерирующих матриц, предложенный Нидеррайтером, основывается на нетривиальной математической теории, связанной с алгеброй многочленов, линейными рекуррентными последовательностями и формальными рядами Лорана над конечными полями. Для полного понимания алгоритма необходимо ознакомиться с теорией, которую можно найти в [5, 6]. Мы же, ставя целью настоящего документа строгое объяснение минимальное по объёму и максимальное по доступности, сосредоточимся на описании алгоритма и необходимого теоретического минимума.

Напомним некоторые определения из алгебры, которые будут использоваться в дальнейшем.

ОПРЕДЕЛЕНИЕ 3.1

Последовательность $\{\alpha_0, \alpha_1, \alpha_2, \dots\}$ над конечным полем $\mathbb{F}_b$ называется *линейной рекуррентной последовательностью порядка* k, если существуют k элементов $\mu_l \in \mathbb{F}_b$, $l \in [0\,..\,k)$ таких, что для $n \in \mathbb{Z}_{\geq k}$ выполняется *линейное рекуррентное соотношение*:

$$\alpha_n = \mu_{k-1}\alpha_{n-1} + \mu_{k-2}\alpha_{n-2} + \cdots + \mu_1\alpha_{n-k+1} + \mu_0\alpha_{n-k}.$$

Исходя из данного определения, установим, что для задания линейной рекуррентной последовательности порядка k помимо линейного рекуррентного соотношения необходимо задать первые k элементов α_n, где $n \in [0\,..\,k)$, которые называются *инициализирующими* (или *начальными*) элементами последовательности $\{\alpha_n\}$.

Введём также важнейшее понятие, непосредственно связанное с линейными рекуррентными соотношениями.

ОПРЕДЕЛЕНИЕ 3.2

Характеристическим многочленом линейной рекуррентной последовательности $\{\alpha_n\}$ порядка k над полем $\mathbb{F}_b$ называется такой $\mu(\omega) \in \mathbb{F}_b[\omega]$, что

$$\mu(\omega) = \omega^k - \mu_{k-1}\omega^{k-1} - \mu_{k-2}\omega^{k-2} - \cdots - \mu_1\omega - \mu_0,$$

и коэффициенты $\mu_l \in \mathbb{F}_b$, $l \in [0\,..\,k)$ которого задают линейное рекуррентное соотношение последовательности:

$$\alpha_n = \mu_{k-1}\alpha_{n-1} + \mu_{k-2}\alpha_{n-2} + \cdots + \mu_1\alpha_{n-k+1} + \mu_0\alpha_{n-k}, \qquad n \in \mathbb{Z}_{\geq k}.$$

Таким образом, линейная рекуррентная последовательность вполне задаётся набором инициализирующих значений и характеристическим многочленом, порождающим линейное рекуррентное соотношение.

Перейдём непосредственно к алгоритму. Пусть известны $t, m \in \mathbb{N}_0$ и $s \in \mathbb{N}$. Тогда для каждой i-й размерности необходимо выбрать неприводимый над $\mathbb{F}_2$ многочлен $\pi[i]$ так, чтобы все многочлены $\pi[1], \ldots, \pi[s]$ были различны, а их степени $e[i] \coloneqq \deg \pi[i]$ удовлетворяли условию $t = \sum_{i=1}^{s}(e[i] - 1) \leq m$. Если такие многочлены выбрать не удалось, это означает, что построить (t, m, s)-сеть с заданными параметрами t, m и s по данному алгоритму невозможно, и на этом его действие заканчивается. На практике обычно оказывается удобнее задавать только часть параметров, а потом, исходя из известных ограничений, выражать оставшиеся.

В случае подходящих под ограничение параметров t, m и s, начинает выполняться основная часть алгоритма, определяющая создание генерирующих матриц $\Gamma[i]$. Так как оно происходит аналогично для любой размерности, мы условимся для лучшей читаемости опускать в нашем описании индекс $[i]$.

Положим $\pi \coloneqq \pi[i]$, $e \coloneqq \deg \pi$. Тогда разделим построчно матрицу $\Gamma \coloneqq \Gamma[i] \in \mathbb{F}_2^{m \times m}$, начиная с нулевой строки, на *секции* так, что

- к первой секции относятся строки с номерами $[0\,..\,e)$,

- ко второй – с номерами $[e\,..\,2e)$,
- и так далее вплоть до последней секции, которой будут отвечать строки с номерами $[a \cdot e\,..\,m)$, где $a \in \mathbb{N}$.

Для наглядности проиллюстрируем такое разделение на примере матрицы размера 8×8 ($m = 8$), для которой $e = 3$.

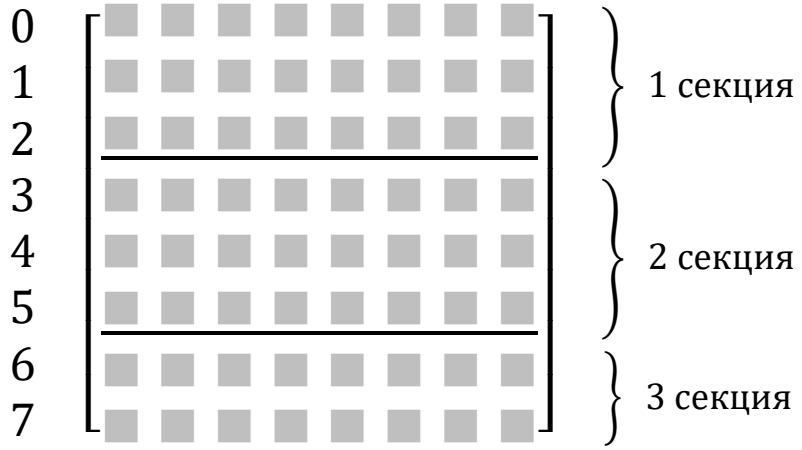


Рисунок 3.1.1. Пример разделения матрицы 8×8 на секции при $e = 3$.

Каждой u-й секции, где $u \in [1\,..\,\mathrm{div}(m-1, e) + 1]$, ставится в соответствие линейная рекуррентная последовательность $\{\alpha_l(u)\}$ порядка $e \cdot u$ с характеристическим многочленом π^u, начальные элементы которой задаются по следующему правилу:

- $\alpha_l(u)$ полагаются равными 0 для $l \in \big[0\,..\,e \cdot (u-1)\big)$;
- Среди $\alpha_l(u)$ с номерами $l \in [e \cdot (u-1)\,..\,e \cdot u)$ хотя бы один элемент назначается не равным нулю.

Оставшиеся элементы последовательности для $l \in \mathbb{Z}_{\geq e \cdot u}$ вычисляются по линейному рекуррентному соотношению:

$$\alpha_l(u) = \mu_{e \cdot u-1}(u)\alpha_{l-1}(u) \oplus \mu_{e \cdot u-2}(u)\alpha_{l-2}(u) \oplus \ldots \oplus \mu_0(u)\alpha_{l-e \cdot u}(u),$$

где коэффициенты $\mu_k(u)$ определяются из выражения $\pi^u(\omega)$:

$$\pi^u(\omega) = \omega^{e \cdot u} \oplus \mu_{e \cdot u-1}(u)\omega^{e \cdot u-1} \oplus \mu_{e \cdot u-2}(u)\omega^{e \cdot u-2} \oplus \ldots \oplus \mu_1(u)\omega \oplus \mu_0(u).$$

Элементами построенной последовательности $\{\alpha_l(u)\}$ затем заполняются, начиная с нулевого элемента, строки u-й секции по следующему принципу:

- нулевая строка в секции заполняется $\alpha_l(u)$ с номерами $l \in [0\,..\,m)$,
- первая – $\alpha_l(u)$ с номерами $l \in [1\,..\,m+1)$,
- вторая – $\alpha_l(u)$ с номерами $l \in [2\,..\,m+2)$
- и так далее, вплоть до последней строки в секции.

Изобразим данный способ заполнения на рассмотренном ранее примере с $m = 8$ и $e = 3$, обозначив серым цветом нулевые элементы, оранжевым цветом возможно ненулевые начальные элементы последовательностей, а синим – элементы, вычисляемые по линейному рекуррентному соотношению.

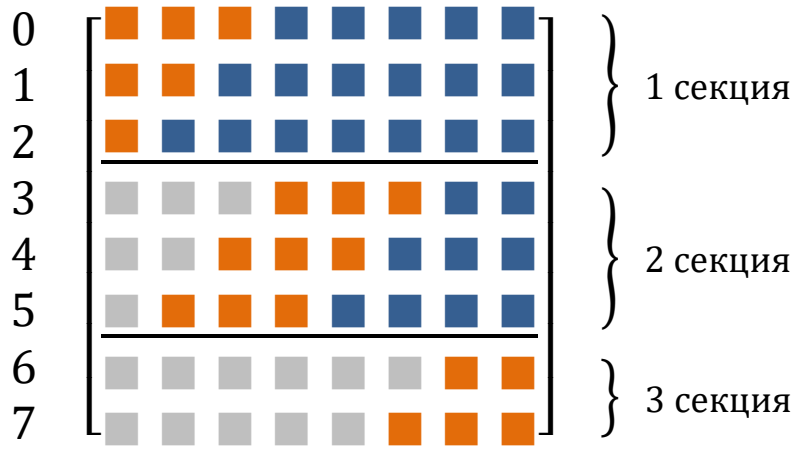


Рисунок 3.1.2. Пример разделения матрицы 8×8 на секции при $e = 3$. Здесь: серые клетки – нулевые начальные элементы последовательностей, оранжевые клетки – возможно ненулевые начальные элементы последовательностей, синие клетки – элементы, вычисляемые по рекуррентному соотношению.

На языке строгих математических формул соответствие между элементами γ_{jk} матрицы Γ и членами последовательностей $\{\alpha_l(u)\}$ устанавливается следующим образом:

$$\gamma_{jk} = \alpha_{r_j+k}(q_j + 1), \qquad j,k \in [0\,..\,m),$$

где $r_j := \mathrm{mod}(j,e)$, $q_j := \mathrm{div}(j,e)$.

Обобщая полученную формулу на произвольные $i \in [1\,..\,s]$, получаем:

$$\gamma_{jk}[i] = \alpha_{r_j[i]+k}[i]\big(q_j[i]+1\big), \qquad j,k \in [0\,..\,m),$$

где

- $\gamma_{jk}[i]$ – это элементы матрицы $\Gamma[i]$,
- $\alpha_l[i](u)$ – элементы последовательностей, строящихся для заполнения матриц $\Gamma[i]$,
- $e[i]$ – степени многочленов $\pi[i]$,
- $r_j[i]$ – остаток от деления j на $e[i]$,
- $q_j[i]$ – целая часть от деления j на $e[i]$.

На этом основная часть алгоритма заканчивается. Задав описанным способом генерирующие матрицы, мы имеем возможность рассчитать любую точку цифровой (t, m, s)-сети по формуле (2.1) (стр. 19). В совокупности всех этих действий и состоит алгоритм Нидеррайтера.

Обозначим ещё раз основную часть алгоритма, предполагая, что заданы приемлемые параметры t, m и s:

1. Выбираем s различных неприводимых многочленов $\pi[i]$ со степенями $e[i] \coloneqq \deg \pi[i]$ такими, что $t = \sum_{i=1}^{s}(e[i]-1)$;
2. Для $i \in [1\,..\,s]$
   2.1. $q_m[i] \coloneqq \mathrm{div}(m-1, e[i])$, $r_m[i] \coloneqq \mathrm{mod}(m-1, e[i])$;
      2.1.1. Для $u \in [1\,..\,q_m[i]+1]$ находим $\mu(\omega) \coloneqq \pi^u[i](\omega)$, а также
      2.1.2. Определяем начальные элементы последовательности $\{\alpha_l[i](u)\}$ с характеристическим многочленом $\mu(\omega)$ так, что
         - $\alpha_l[i](u) \coloneqq 0$ для $l \in \big[0\ ..\ e[i]\cdot(u-1)\big)$, а
         - среди $\alpha_l[i](u)$, $l \in [e[i]\cdot(u-1)\,..\,e[i]\cdot u)$, хотя бы один равен 1;
      2.1.3. Если $u \neq q_m[i]+1$, то $r_{hi} \coloneqq\ e[i]-1$, иначе: $r_{hi} \coloneqq r_m[i]$;
      2.1.4. Рекуррентно вычисляем $\alpha_l[i](u)$ для $l \in [e[i]\cdot u\,..\,m+r_{hi})$;
      2.1.5. Для $k \in [0\,..\,m)$ и для $r \in [0\,..\,r_{hi}]$
         Вычисляем $\gamma_{e\cdot u+r,k}[i] \coloneqq \alpha_{r+k}[i](u)$.

3.2. Пример работы алгоритма

Рассмотрим работу классического алгоритма генерации на конкретном простом примере построения $(1,3,2)$-сети с основанием 2. Иными словами, известны параметры:

$$t = 1 \quad m = 3 \quad s = 2 \quad b = 2.$$

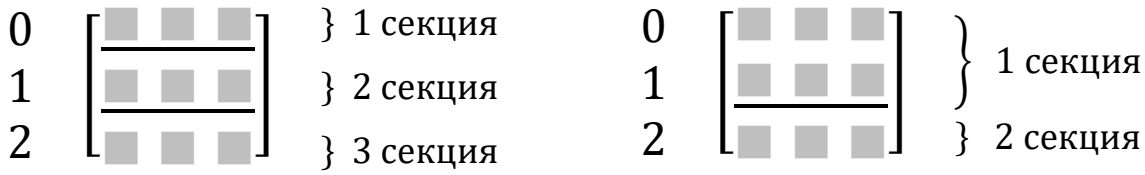
По описанной в предыдущем разделе схеме, перед началом непосредственных расчётов необходимо выбрать для каждой из двух компонент пространства неприводимые над $\mathbb{F}_2$ многочлены $\pi[1]$ и $\pi[2]$ такие, что $\pi[1] \neq \pi[2]$ и $t = \sum_{i=1}^{s}(e[i] - 1)$, или, в нашем случае, такие, что $1 = \sum_{i=1}^{2}(e[i] - 1) = e[1] + e[2] - 2$. В качестве искомых многочленов подойдут, скажем,

$$\pi[1] = \omega \oplus 1,$$
$$\pi[2] = \omega^2 \oplus \omega \oplus 1.$$

Действительно, $e[1] = 1$, $e[2] = 2$ и $e[1] + e[2] - 2 = 1 + 2 - 2 = 1$. Обратим внимание, что ограничения на сам вид неприводимых многочленов или порядок их назначения компонентам, вообще говоря, не установлены. С таким же успехом можно было вместо $\omega \oplus 1$ выбрать многочлен ω или, например, поменять местами $\pi[1]$ с $\pi[2]$ – единственное условие, которое должно выполняться, – это условие $t = \sum_{i=1}^{s}(e[i] - 1)$. Иными словами, любая пара многочленов $(\pi[1], \pi[2])$ среди, например, таких пар как $(\omega,\ \omega^2 \oplus \omega \oplus 1)$, $(\omega^2 \oplus \omega \oplus 1,\ \omega)$ или $(\omega^2 \oplus \omega \oplus 1,\ \omega + 1)$ точно так же подойдёт для генерации $(1,3,2)$-сети, как и обозначенная выше пара $(\omega \oplus 1,\ \omega^2 \oplus \omega \oplus 1)$. Тем не менее, результаты, получаемые с помощью различных многочленов, всё-таки, будут отличаться. Подробнее об этом можно узнать, ознакомившись с понятием $(t, m, \boldsymbol{e}, s)$-сетей в работах японского математика Сю Тезуки [12].

Определившись с многочленами, перейдём к построению матриц $\Gamma[1]$ и $\Gamma[2]$. Известно, что обе эти матрицы должны иметь размер $m \times m$, то есть

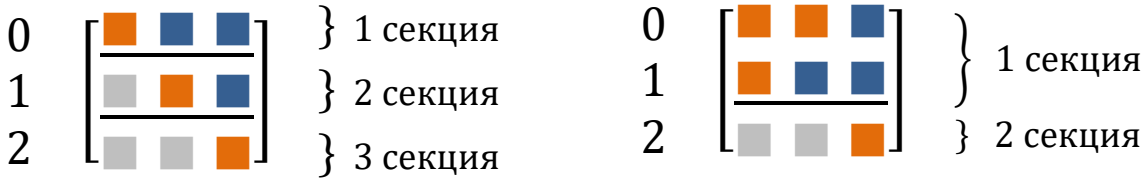
3×3. Для удобства разделим сразу каждую из них на секции, как это было сделано в предыдущем разделе.



(а) (б)

Рисунок 3.2.1. (а) $\Gamma[1]$, разбитая на секции при $e[1] = 1$;
(б) $\Gamma[2]$, разбитая на секции при $e[2] = 2$.

Обе матрицы Γ необходимо заполнить членами линейных рекуррентных последовательностей, порождаемых степенями соответствующих многочленов π. Изобразим схему расположения элементов последовательностей в матрицах аналогично тому, как это сделано на рисунке 3.1.2 (стр. 23).



(а) (б)

Рисунок 3.2.2. (а) $\Gamma[1]$, разбитая на секции при $e[1] = 1$; (б) $\Gamma[2]$, разбитая на секции при $e[2] = 2$. Здесь: серые клетки – нулевые начальные элементы последовательности, оранжевые клетки – возможно ненулевые начальные элементы последовательностей, синие клетки – элементы, вычисляемые по рекуррентному соотношению.

Найдём члены линейной рекуррентной последовательности, необходимые для заполнения первой секции матрицы $\Gamma[1]$. Здесь $u = 1$ и, следовательно, многочлен $\pi^u[1] = \pi^1[1] = \pi[1] = \omega \oplus 1$ будет характеристическим для $\{a_l[1](1)\}$, порождая собой линейное рекуррентное соотношение

$$\alpha_l[1](1) = \alpha_{l-1}[1](1).$$

Определим единственное необходимое начальное значение $\alpha_0[1](1)$ по правилам, перечисленным в пункте 2.1.3 схемы алгоритма (стр. 24). Согласно им, при $l \in \big[0..e[1]\cdot(u-1)\big)$ все $\alpha_l[1](1)$ должны быть равны нулю. Вместе с тем, $e[1]\cdot(u-1) = 1\cdot 0 = 0$, следовательно, $\big[0..e[1]\cdot(u-1)\big) = [0..0) = \emptyset$ и эту часть правил можно пропустить. Далее, при $l \in [e[1]\cdot(u-1)..e[1]\cdot u)$ хотя бы один $\alpha_l[1](1)$ должен быть равным единице. Определим, что $e[1]\cdot u = 1\cdot 1 = 1$ и, таким образом, $[e[1]\cdot(u-1)..e[1]\cdot u) = [0..1) = \{0\}$. Данные обстоятельства не оставляют нам особого выбора: единственный подходящий под все условия вариант – $\alpha_0[1](1) = 1$. Воспользовавшись выведенным ранее отношением $\alpha_l[1](1) = \alpha_{l-1}[1](1)$, получим необходимые элементы линейной рекуррентной последовательности для заполнения первой секции Г[1]:

$$\alpha_0[1](1) = 1,$$
$$\alpha_1[1](1) = 1,$$
$$\alpha_2[1](1) = 1.$$

Таким образом, матрица Г[1] приобретает вид, показанный ниже.

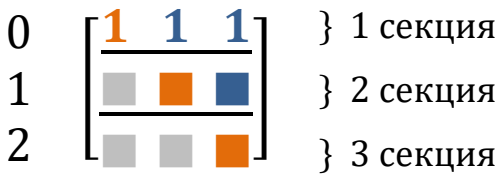


Рисунок 3.2.3. Матрица Г[1] с заполненной первой секцией.

Характеристическими многочленами для второй и третьей секции будут, соответственно, $\pi^2[1] = (\omega \oplus 1)^2 = \omega^2 \oplus 1$ и, аналогичным образом, $\pi^3[1] = (\omega \oplus 1)^3 = \omega^3 \oplus \omega^2 \oplus \omega \oplus 1$, которые определяют рекуррентные отношения, соответственно,

$$\alpha_l[1](2) = \alpha_{l-2}[1](2),$$
$$\alpha_l[1](3) = \alpha_{l-1}[1](3) \oplus \alpha_{l-2}[1](3) \oplus \alpha_{l-3}[1](3).$$

Задавая начальные значения по тем же принципам, получаем требуемые элементы последовательностей для заполнения второй и третьей секций матрицы:

$$\alpha_0[1](2) = 0, \qquad \alpha_0[1](3) = 0,$$
$$\alpha_1[1](2) = 1, \qquad \alpha_1[1](3) = 0,$$
$$\alpha_2[1](2) = 0, \qquad \alpha_2[1](3) = 1.$$

Проиллюстрируем полностью заполненную матрицу Г[1].

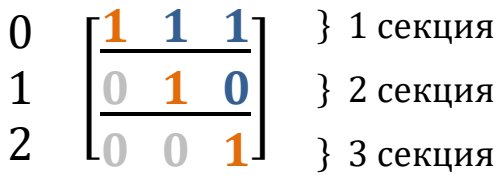


Рисунок 3.2.4. Заполненная матрица Г[1].

Приступим к построению матрицы Г[2], состоящей из двух секций. Рассмотрим первую из них с характеристическим многочленом $\pi^1[2] = \pi[2] = \omega^2 \oplus \omega \oplus 1$. В очередной раз подберём начальные значения для линейной рекуррентной последовательности. Для $l \in \big[0..e[2] \cdot (u-1)\big)$ все $\alpha_l[2](1)$ должны равняться нулю. Нетрудно убедиться, что $\big[0..e[2] \cdot (u-1)\big)$ здесь снова будет равняться пустому множеству. Далее, для $l \in [e[2] \cdot (u-1)..e[2] \cdot u)$, или, если конкретно, для $l \in [0..2)$ хотя бы один $\alpha_l[2](1)$ должен быть единицей. Как видно, в данном случае появляется бóльшая свобода действий: можно назначить единицей либо $\alpha_0[2](1)$, либо $\alpha_1[2](1)$, либо оба этих элемента сразу. Мы остановимся на варианте, когда $\alpha_0[2](1) = \alpha_1[2](1) = 1$. Выводя из характеристического многочлена общий вид рекуррентного отношения

$$\alpha_l[2](1) = \alpha_{l-1}[2](1) \oplus \alpha_{l-2}[2](1),$$

получаем список элементов линейной рекуррентной последовательности $\{\alpha_l[2](1)\}$, необходимых для заполнения первой секции матрицы Г[2]:

$$\alpha_0[2](1) = 1,$$
$$\alpha_1[2](1) = 1,$$
$$\alpha_2[2](1) = 0,$$
$$\alpha_3[2](1) = 1.$$

Заполним ими матрицу.

$$\begin{matrix} 0 \\ 1 \\ 2 \end{matrix} \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \blacksquare & \blacksquare & \blacksquare \end{bmatrix} \begin{matrix} \} \\ \\ \} \end{matrix} \begin{matrix} 1 \text{ секция} \\ \\ 2 \text{ секция} \end{matrix}$$

Рисунок 3.2.5. Матрица Γ[2] с заполненной первой секцией.

Проделав ещё раз аналогичные действия для второй секции матрицы Γ[2], получаем её окончательный вид.

$$\begin{matrix} 0 \\ 1 \\ 2 \end{matrix} \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{matrix} \} \\ \\ \} \end{matrix} \begin{matrix} 1 \text{ секция} \\ \\ 2 \text{ секция} \end{matrix}$$

Рисунок 3.2.6. Заполненная матрица Γ[2].

Таким образом, выражения для обеих матриц Γ получены, можно приступать к непосредственному вычислению координат точек конструируемой (1,3,2)-сети. Продемонстрируем данный процесс, рассчитав явно координаты, например, точки x_5.

Для точки x_5 значение $n = 5$ и первая координата, в соответствии с алгоритмом, будет рассчитана по формуле (2.1) (стр. 19):

$$x_5[1] = \mathrm{rnum}_b\left(\Gamma[i] \cdot \mathbf{vec}_{b,m}(n)\right) \cdot b^{-m} =$$
$$= \mathrm{rnum}_2\left(\Gamma[1] \cdot \mathbf{vec}_{2,3}(5)\right) \cdot 2^{-3} =$$
$$= \mathrm{rnum}_2\left(\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right) \cdot 2^{-3} =$$

$$= \mathrm{rnum}_2\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right) \cdot 2^{-3} = 4 \cdot 2^{-3} = 0.125.$$

Аналогично, вторая координата:

$$x_5[2] = \mathrm{rnum}_b\left(\Gamma[i] \cdot \mathbf{vec}_{b,m}(n)\right) \cdot b^{-m} =$$

$$= \mathrm{rnum}_2\left(\Gamma[2] \cdot \mathbf{vec}_{2,3}(5)\right) \cdot 2^{-3} =$$

$$= \mathrm{rnum}_2\left(\begin{bmatrix}1 & 1 & 0\\1 & 0 & 1\\0 & 0 & 1\end{bmatrix} \cdot \begin{bmatrix}1\\0\\1\end{bmatrix}\right) \cdot 2^{-3} =$$

$$= \mathrm{rnum}_2\left(\begin{bmatrix}1\\0\\1\end{bmatrix}\right) \cdot 2^{-3} = 5 \cdot 2^{-3} = 0.625.$$

Получаем, что $x_5 = [0.125,\ 0.625]^T$. Проделав эту же процедуру для всех $n \in [0..b^m) = [0..2^3) = [0..8)$, в итоге получаем координаты всех восьми точек, образующих $(1,3,2)$-сеть с основанием 2, полный перечень которых представлен в таблице 3.2.1.

Таблица 3.2.1. Сгенерированная $(1,3,2)$-сеть

n	$x_n[1]$	$x_n[2]$
0	0.0	0.0
1	0.5	0.75
2	0.75	0.5
3	0.25	0.25
4	0.625	0.375
5	0.125	0.625
6	0.375	0.875
7	0.875	0.125

Для большей наглядности продемонстрируем полученные результаты графически.

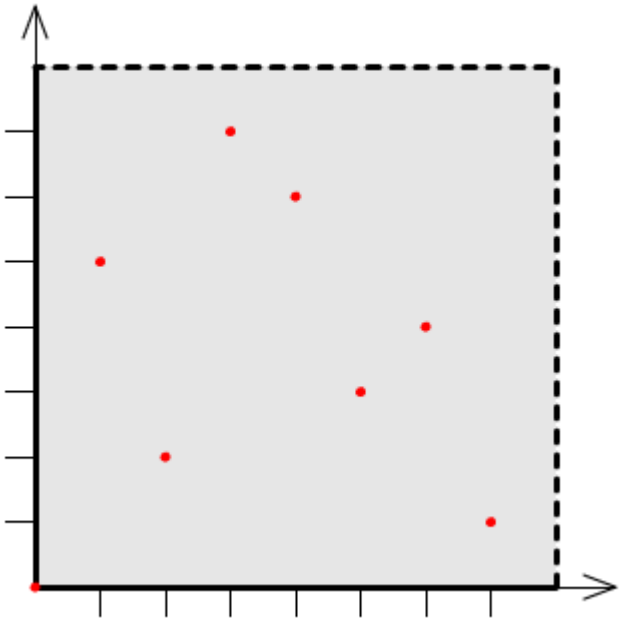


Рисунок 3.2.7. Сгенерированная $(1,3,2)$-сеть внутри единичного двумерного куба $\mathcal{I}^2$.

При желании можно провести проверку выполнимости определения, подобную той, что была проделана в главе 1, и убедиться в том, что полученное множество действительно удовлетворяет всем предъявляемым требованиям.

3.3. Оптимизация алгоритма генерации

Только что на простом примере было показано, как алгоритм, изложенный в разделе 3.1, позволяет строить (t, m, s)-сети с основанием 2. Попробуем теперь явно оценить вычислительную сложность этого алгоритма и модифицировать его таким образом, чтобы её уменьшить.

Для начала оговорим, что именно понимается здесь под вычислительной сложностью. Условно схему генерации цифровых (t, m, s)-сетей можно разделить на два последовательных этапа: этап нахождения генерирующих матриц Г и этап непосредственного расчёта точек. Учитывая то, что нахождение матриц Г достаточно провести один раз, можно назвать это своего рода инициализирующим этапом выполнения алгоритма и принять, что время, затрачиваемое на инициализацию, не столь важно – важно только то, чтобы каждая очередная точка генерировалась с максимальной скоростью. Таким образом, именно затраты на выполнение второго этапа алгоритма принимаются далее за вычислительную сложность, которая, очевидно, пропорциональна сложности вычисления одной координаты одной точки.

Нахождение координаты точки в схеме генерации цифровых (t, m, s)-сетей с основанием 2 производится по формуле (2.1) (стр. 19), в которой b принимается равной двум. Рассчитаем количество операций, необходимых для проведения расчётов по этой формуле. Для наглядности продублируем ещё раз исходный её вид с $b = 2$:

$$x_n[i] = \mathrm{rnum}_2\left(\Gamma[i] \cdot \mathbf{vec}_{2,m}(n)\right) \cdot 2^{-m}.$$

Выразим эквивалентным образом аргумент функции реверсивной нумеризации, воспользовавшись базовым свойством матричного умножения:

$$\Gamma[i] \cdot \mathbf{vec}_{2,m}(n) = \Gamma[i] \cdot \begin{bmatrix} (n)_{2,0} \\ (n)_{2,1} \\ \vdots \\ (n)_{2,m-1} \end{bmatrix} = \bigoplus_{k=0}^{m-1} (n)_{2,k} \cdot \Gamma^k[i].$$

Подставим получившееся выражение в формулу для $\boldsymbol{x}_n[i]$:

$$x_n[i] = \mathrm{rnum}_2\left(\Gamma[i] \cdot \mathbf{vec}_{2,m}(n)\right) \cdot 2^{-m} =$$

$$= \mathrm{rnum}_2\left(\bigoplus_{k=0}^{m-1} (n)_{2,k} \cdot \Gamma^k[i]\right) \cdot 2^{-m} =$$

$$= \bigoplus_{k=0}^{m-1} \left((n)_{2,k} \cdot \mathrm{rnum}_2(\Gamma^k[i])\right) \cdot 2^{-m}.$$

Последний переход в этом равенстве может показаться неочевидным, однако он довольно просто объясняется определением операции $\oplus$ исключающей дизъюнкции (см. «Обозначения»). Согласно нему, чтобы $c = a \oplus b$, требуется выполнимость равенства $(c)_{2,k} = (a)_{2,k} \oplus (b)_{2,k}$ для всех двоичных разрядов этих чисел. Помимо этого, отметим то, что существует взаимно-однозначное соответствие между разрядами чисел $\mathrm{rnum}_2(\boldsymbol{v}) \in [0\,..\,2^m)$ в двоичной системе счисления и векторами $\boldsymbol{v} \in [0\,..2)^m$. Следовательно, если для двоичных векторов выполняется равенство $\boldsymbol{w} = \boldsymbol{u} \oplus \boldsymbol{v}$, то будет выполняться и равенство $\mathrm{rnum}_2(\boldsymbol{w}) = \mathrm{rnum}_2(\boldsymbol{u}) \oplus \mathrm{rnum}_2(\boldsymbol{v})$. Полагая в нём вектор

$$\boldsymbol{w} = \bigoplus_{k=0}^{m-1} (n)_{2,k} \cdot \Gamma^k[i],$$

а векторы $\boldsymbol{u}$ и $\boldsymbol{v}$, скажем,

$$\boldsymbol{u} = (n)_{2,0} \cdot \Gamma^0[i],$$

$$\boldsymbol{v} = \bigoplus_{k=1}^{m-1} (n)_{2,k} \cdot \Gamma^k[i]$$

и продолжая такое разложение по индукции, получаем равенство между второй и третьей строками выражения выше.

Ввиду особой значимости чисел $\mathrm{rnum}_2(\Gamma^k[i])$, для них существует отдельное определение.

ОПРЕДЕЛЕНИЕ 3.3.1

Пусть $\Gamma[i]$ – генерирующие матрицы цифровой (t, m, s)-сети. Тогда целые числа $\mathrm{rnum}_2(\Gamma^k[i])$, называются *направляющими числами* соответствующей цифровой (t, m, s)-сети и обозначаются как $g_k[i]$.

Используя введённое определение, получим более удобную для практического применения формулу:

$$x_n[i] = \bigoplus_{k=0}^{m-1} \left(g_k[i] \cdot (n)_{2,k}\right) \cdot 2^{-m}, \tag{3.3.1}$$

из которой видно, что программное вычисление любой координаты $\boldsymbol{x}_n[i]$ произвольной точки цифровой (t, m, s)-сети по формуле (2.1) (стр. 19) подразумевает применение

- одной операции деления чисел с плавающей точкой,
- m операций выделения двоичных разрядов числа n,
- $(m-1)$ операции «$\oplus$» и
- m операций «$\cdot$».

Имеется возможность существенно повысить вычислительную эффективность генерации при поочерёдном расчёте точек с последовательными номерами [6]. Для того, чтобы объяснить такое улучшение, введём понятие кода Грея.

ОПРЕДЕЛЕНИЕ 2.3.2

Кодом Грея числа $n \in \mathbb{N}_0$ называется целое неотрицательное число $G(n) = n \oplus \left\lfloor \frac{n}{2} \right\rfloor$.

Код Грея обладает следующими замечательными свойствами:

1. $G(n)$ является биекцией на $\mathbb{N}_0$;
2. Коды $G(n)$ и $G(n+1)$ отличаются только в одном двоичном разряде.

Иначе говоря,

$$\forall\, n \in \mathbb{N}_0\ \exists!\ k_0 \mid \begin{cases} \big(G(n+1)\big)_{2,k} = \big(G(n)\big)_{2,k} & k \neq k_0 \\ \big(G(n+1)\big)_{2,k} = \big(G(n)\big)_{2,k} \oplus 1 & k = k_0 \end{cases}.$$

Аналитически номер изменяющегося разряда k_0 выражается с помощью формулы $k_0 = \log_2\big(G(n) \oplus G(n+1)\big)$. Также его можно выразить, как минимальный номер нулевого разряда в двоичном представлении n, иными словами $k_0 = \min\{k \mid (n)_{2,k} = 0\}$;

3. Для любого $m \in \mathbb{N}_0$, кодирующая функция G взаимно-однозначно отображает целочисленный отрезок вида $[0\,..\,2^m)$ в себя же:

$$G\big([0\,..\,2^m)\big) = [0\,..\,2^m);$$

4. Для любых $m \in \mathbb{N}_0$, $k_1 \in \mathbb{N}_0$ существует единственное $k_2 \in \mathbb{N}_0$ такое, что

$$G\big([k_1 2^m\,..\,(k_1+1)2^m)\big) = [k_2 2^m\,..\,(k_2+1)2^m).$$

Иначе говоря, функция G взаимно-однозначно отображает множество всех подмножеств $\mathbb{N}_0$ вида $[k2^m..\,(k+1)2^m)$, где $k, m \in \mathbb{N}_0$, в себя же.

Исходя из последних двух свойств, можно сделать вывод о том, что если $\{x_n\}_{n\in[0..2^m)}$ является (t,m,s)-сетью, то $\{x_{G(n)}\}_{n\in[0..2^m)}$ будет той же самой (t,m,s)-сетью с точностью до индексирования точек [3]. Этого, казалось бы, незначительного различия между $\{x_n\}_{n\in[0..2^m)}$ и $\{x_{G(n)}\}_{n\in[0..2^m)}$ становится достаточно для того, чтобы последовательный расчёт элементов цифровых (t,m,s)-сетей в порядке, задаваемом кодом Грея, оказался значительно проще и эффективнее. Продемонстрируем это далее.

Для удобства введём новое обозначение $y_n[i] \coloneqq x_n[i] \cdot 2^m$. Точки $\boldsymbol{y}_n$, де-факто, получаются выделением из формулы (3.3.1) (стр. 34) целочисленных расчётов:

$$y_n[i] = x_n[i] \cdot 2^m = \bigoplus_{k=0}^{m-1} \big(g_k[i] \cdot (n)_{2,k}\big) \cdot 2^{-m} \cdot 2^m =$$

$$= \bigoplus_{k=0}^{m-1} g_k[i] \cdot (n)_{2,k}.$$

Подставим в полученную формулу вместо индекса точки n его код Грея:

$$y_{G(n)}[i] = \bigoplus_{k=0}^{m-1} g_k[i] \cdot \big(G(n)\big)_{2,k}. \tag{3.3.2}$$

Координаты точки с номером $G(n+1)$, где $n < 2^m - 1$, выражаются как

$$y_{G(n+1)}[i] = \bigoplus_{k=0}^{m-1} g_k[i] \cdot \big(G(n+1)\big)_{2,k}.$$

Как известно, коды Грея чисел n и $n+1$ отличаются в единственном разряде. Это позволяет нам утверждать, что $\exists!\ k_0 \in [0\,..\lceil \log_2 G(n+1)\rceil]$ такое, что

$$y_{G(n+1)}[i] = \left(\bigoplus_{k=0}^{m-1} g_k[i] \cdot \big(G(n)\big)_{2,k} \right) \oplus g_{k_0}[i],$$

или же

$$y_{G(n+1)}[i] = y_{G(n)}[i] \oplus g_{k_0}[i].$$

Таким образом, зная номер $n < 2^m - 1$ и направляющие числа $g_k[i]$, возможно из любой координаты $y_{G(n)}[i]$ получить соответствующую координату $y_{G(n+1)}[i]$ за одну операцию $\oplus$ и за одну операцию отыскания соответствующего k_0, подразумевающую в наихудшем случае m операций выделения двоичных разрядов и m сравнений выделенных разрядов с нулём.

Как видно из приведённых рассуждений, рассчитывать координаты точек цифровой (t, m, s)-сети $\{\boldsymbol{x}_{G(n)}\}_{n\in[0..2^m)}$ оказывается даже в наихудшем случае эффективнее, чем координаты точек сети $\{\boldsymbol{x}_n\}_{n\in[0..2^m)}$, а ввиду того, что обе сети определяют одно и то же мультимножество точек, конечному пользователю зачастую не важен их порядок. Таким образом, в большинстве случаев оптимально на запрос генерации точки с номером n вычислять точку $\boldsymbol{x}_{G(n)}$, а процесс вычисления производить по следующей схеме:

Пусть имеются заданные генерирующие матрицы $\Gamma[i]$ и направляющие числа $g_k[i]$ для $i \in [1..s]$, и пусть требуется сгенерировать точки с номерами $[n_0 \,..\, n_0 + h)$, где $n_0 \in \mathbb{N}_0$ и $h \in \mathbb{N}$.

1. По формуле (3.3.2) (стр. 36) вычисляем целые числа $y_{prev}[i] \coloneqq y_{G(n_0)}[i]$ для $i \in [1..s]$;
2. Если $h \neq 1$, то для $l \in [1..h)$:
   2.1. Находим номер разряда k_0, отличающегося в бинарном представлении $G(n_0 + l - 1)$ от $G(n_0 + l)$;
   2.2. Для $i \in [1..s]$:
      2.2.1. Вычисляем $y_{next}[i] \coloneqq y_{prev}[i] \oplus g_{k_0}[i]$;
      2.2.2. Рассчитываем координаты $x_{G(n_0+l-1)}[i] \coloneqq y_{prev}[i] \cdot 2^{-m}$;
      2.2.3. Присваиваем $y_{prev}[i] \coloneqq y_{next}[i]$.
3. Для $i \in [1..s]$:
   3.1. Рассчитываем координаты $x_{G(n_0+h-1)}[i] \coloneqq y_{prev}[i] \cdot 2^{-m}$.

Список литературы

1. van der Corput J.G. Verteilungsfunktionen (Erste Mitteilung) // Proceedings of the Koninklijke Akademie van Wetenschappen te Amsterdam. — 1935. — Vol. 38. — P. 813-821.
2. Sobol I.M. Distribution of points in a cube and approximate evaluation of integrals // U.S.S.R Comput. Maths. Math. Phys. — 1967. — Vol. 7. — P. 784-802.
3. Антонов И.А., Салеев В.М. Экономичный способ вычисления ЛПτ-последовательностей // Ж. вычисл. матем. и матем. физ. — 1979. — Том 19, №1. — С. 243-245.
4. Niederreiter H. Point sets and sequences with small discrepancy // Monatshefte für Mathematik. — 1987. — Vol. 104, No. 4. — P. 273-337.
5. Niederreiter H. Low-Discrepancy and Low-Dispersion Sequences // Journal of Number Theory. — 1988. — Vol. 30. — P. 51-70.
6. Bratley P., Fox B.L., Niederreiter H. Implementations and Tests of Low-Discrepancy Sequences ACM Transactions on Modeling and Computer Simulation, Vol. 2, No. 3, July 1992, Pages 195-213
7. Faure H. Good permutations for extreme discrepancy // Journal of Number Theory. — 1992. — Vol. 42. — P. 45-56.
8. Niederreiter H. Random Number Generation and Quasi-Monte Carlo Methods. — Philadelphia: SIAM, 1992. — 241 p.
9. Cui J., Freeden W. Equidistribution on the sphere // SIAM Journal on Scientific Computing. — 1997. — Vol. 18, No. 2. — P. 595-609.
10. Pillards T., Cools R. A theoretical view on transforming low-discrepancy sequences from a cube to a simplex // Monte Carlo Methods and Applications. — 2004. — Vol. 10, No. 3-4. — P. 511-529.
11. Dick J., Pillichshammer F. Digital Nets and Sequences: Discrepancy Theory and Quasi–Monte Carlo Integration. — NY: Cambridge University Press, 2010. — 618 p.

12. Tezuka S. On the discrepancy of generalized Niederreiter sequences // Journal of Complexity. — 2013. — Vol. 29. — P. 240-247.
13. Decorrelation of low discrepancy sequences for progressive rendering // US Patent #10074212. 2016 / Waechter C., Binder N.