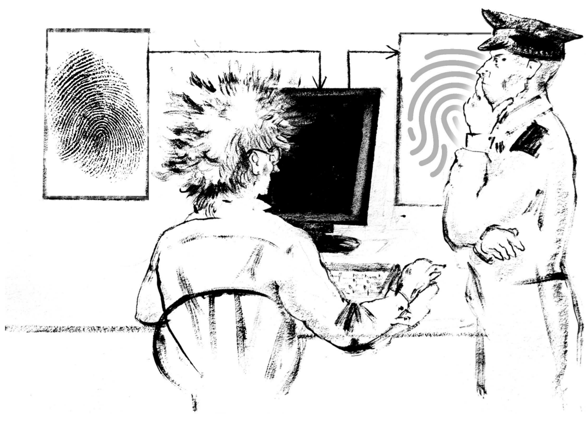
1. Зачем обрабатывать сигналы?

Каждый раз, когда мы работаем на компьютере или говорим по мобильному телефону, мы запускаем мощные механизмы обработки и хранения сигналов. Механизмы эти основаны на достижениях современной математики — теории функций, оптимизации, теории алгоритмов и еще десятка других областей. Какие-то из них родились в позапрошлом веке, другие (их большинство) — буквально вчера. Теория обработки и хранения сигналов (signal processing) является одним из наиболее востребованных направлений математики.

Любой сигнал, будь то звук, изображение или другая функция, почти никогда не хранится в компьютере по точкам. Это дорого и неэффективно. Сигнал раскладывается в сумму других, «базовых» функций, и хранятся его коэффициенты разложения. Главный вопрос — какую систему базовых функций использовать? И как построить хорошую систему, чтобы сигнал быстро и качественно воспроизводился и при этом занимал мало памяти? За это отвечает красивая математическая теория. В течение десятилетий базовыми функциями были целые сжатия синуса и косинуса, что естественно, учитывая природу звука. Это ряды Фурье, изобретенные более 200 лет назад. Однако к середине XX века стало ясно, что они не отвечают современным запросам. Поиск новых конструкций, превосходящих ряды Фурье, оказался непростой задачей. Над



этим трудилось не одно поколение математиков: функции Хаара, система Шеннона—Котельникова, всплески Мейера и Добеши и т. д. Новые функции уже не задаются явными формулами, а строятся как решения специальных уравнений. Они не являются гладкими, а, напротив, имеют свойства фракталов и самоподобных фигур. Сейчас они используются повсеместно при работе с фотографиями, аудио- и видеофайлами, в компьютерной томографии и т. д. — и даже в криминалистике, при обработке отпечатков пальцев[1]! Но математическая теория не стоит на месте.

Данный курс является введением в современную теорию обработки сигналов. Не во всю, а только в ее «функциональную» составляющую. Мы построим и исследуем различные базисные системы — от систем Фурье до всплесков.

Автор постарался сделать курс доступным младшим студентам, поэтому мы будем избегать понятий, которые не проходятся на 1—2 курсах математических отделений университетов. Мы не употребляем интеграл Лебега и не пользуемся символикой и строгим определением пространств L_2, ограничиваясь интуитивным понятием гильбертова пространства, а также не обосновываем сходимость некоторых интегралов, в частности преобразования Фурье. Более подготовленный читатель легко экстраполирует материал на свой уровень и подтвердит, что весь курс написан без потери математической строгости.

Автор глубоко признателен Н. Н. Андрееву, В. А. Клепцыну, Г. А. Мерзону, М. А. Раскину, Н. Гульельми и Х. М. Домингесу-Бегинесу за помощь в публикации этого курса, за внимательное чтение и массу ценных замечаний. Художественные иллюстрации выполнены М. К. Кашиным.

1.1. Непрерывное и дискретное

Любую функцию будем называть сигналом. Функцию будем обозначать $f(t)$, а переменную t будем всегда считать принадлежащей отрезку $[0, 1]$. Таким образом, $f : [0, 1] \to \mathbb{R}$.

Конечно, непрерывные сигналы в памяти компьютера не хранятся, а хранятся их дискретные значения на какой-нибудь сетке,

[1] Началось это, кстати, еще в прошлом веке, когда этим занялось ФБР [2, p. 24—25]: отсканированный набор отпечатков занимал порядка 10 мегабайт — размер, малопригодный к передаче по тогдашним модемам. Применение кодирования с помощью всплесков позволило уменьшить размер в 20 раз, до ~ 500 килобайт.

т. е. в конечном множестве точек. Пусть это будут N точек $t_0, \ldots, t_{N-1}$. Чаще всего используют равномерную сетку: $0, \frac{1}{N}, \ldots, \frac{N-1}{N}$, хотя в некоторых задачах лучше работают другие сетки. Мы будем иметь дело только с равномерной: $t_k = \frac{k}{N}$, $k = 0, \ldots, N-1$. Заметим, что мы начинаем всегда с точки $t_0 = 0$, а заканчиваем не в точке 1, а чуть раньше, в точке $t_{N-1} = \frac{N-1}{N}$. Так делать удобнее, а почему, понятно станет потом.

Таким образом, берется сетка из N точек $t_k = \frac{k}{N}$, $k = 0, \ldots, N-1$. Каждой функции $f(t)$ поставим в соответствие числа $x(k) = f\left(\frac{k}{N}\right)$, $k = 0, \ldots, N-1$. Этот набор из N чисел можно понимать как вектор в N-мерном пространстве: $x = (x(0), \ldots, x(N-1)) \in \mathbb{R}^N$.

Итак, непрерывный сигнал — это функция $f\colon [0, 1] \to \mathbb{R}$, а дискретный — это вектор $x \in \mathbb{R}^N$. Из непрерывного сигнала f получается дискретный по его значениям в узлах сетки: $x(k) = f\left(\frac{k}{N}\right)$, $k = 0, \ldots, N-1$.

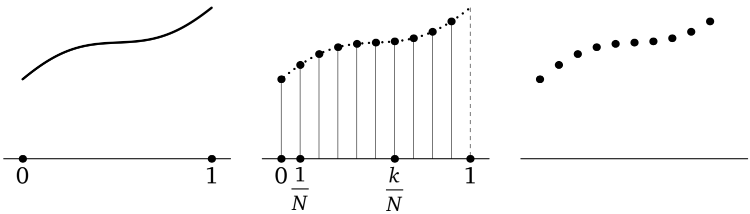


Рис. 1. Слева — непрерывный сигнал $f(t)$, $t \in [0, 1]$; в центре — его дискретизация $x(k) = f\left(\frac{k}{N}\right)$; справа — дискретизованный сигнал $x(k)$

Переходя от непрерывных сигналов к дискретным, т. е. от функций к векторам, и обратно, мы сможем понять свойства сигналов и предложить способы их хранения и передачи.

Для хранения качественного аудио- или видеосигнала, скажем фотографии, требуется число значений N порядка 10^6—10^8. Например, фото среднего качества — это 1 мегапиксель (Мп), т. е. 1 миллион пикселей. В каждом пикселе нужно хранить несколько чисел — яркость данной точки, ее цвет (определяющийся яркостью трех составляющих его цветов). Так что, округляя, будем считать, что нам нужно $N = 10^7$ чисел. С пространством столь огромного числа измерений приходится иметь дело. Подчеркнем, что речь идет о неподвижном изображении. Видеопоток содержит на порядки больше байтов.

Прикинем, сколько места потребуется для хранения 1 Мп. Будем считать, что для каждого числа нужно 10 цифр в двоичной системе счисления, т. е. числа хранятся с точностью до $\frac{1}{1000}$. Тогда получается 10^8 бит, или примерно 10^7 байт, т. е. 10 Мб. Не много ли для одного фото среднего качества? Конечно, изображение надо упаковать и сжать. Раз в десять. Тогда фото будет весить 1 Мб, что вполне приемлемо. Но что значит «упаковать» и «сжать»? Это означает по-другому разместить, уложить более экономно. Для этого существуют разные подходы, например архивация. Но их часто бывает недостаточно. Как можно разместить, пусть даже экономно, 10 Мб информации на пространстве в 1 Мб? Это все равно, что пытаться засунуть 10 кубометров груза в багажник машины объемом в 1 кубометр. Как ни упаковывай, все равно не влезет. Поэтому, когда мы говорим про хранение сигнала, термины «упаковать» и «сжать» не всегда точно отражают реальность. Сигнал надо уменьшить, выбросить 90 % входящих в него чисел. И сделать это без потери (а иногда и с улучшением) качества. Как такое возможно, нам и предстоит разобраться.

1.2. Приближения гладких сигналов

Если бы мы хранили все возможные сигналы $x \in \mathbb{R}^N$, меньшим количеством чисел мы бы, конечно, не обошлись. Главная хитрость заключается в том, что все сигналы хранить не нужно. Все звуки, изображения, физические процессы и другие сигналы, взятые из жизни, обладают некоторой регулярностью. Это значит, что они не могут слишком резко меняться. А хранить плавно меняющиеся сигналы можно, используя значительно меньше места.

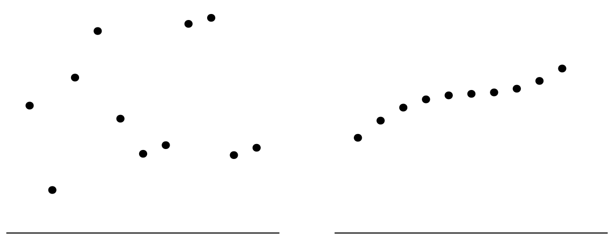


Рис. 2. Сигналы: хаотичный (слева) и регулярный (справа)

Для того чтобы строго сформулировать сказанное, мы перейдем от дискретных сигналов обратно к непрерывным, т. е. к функциям. Если значения $x(k)$ меняются плавно с изменением k, то функция f должна быть непрерывна. На самом деле даже непрерывности недостаточно, и мы потребуем, чтобы она обладала более сильным свойством.

Определение 1. Функция называется *липшицевой*, если существует такая положительная константа L, что для любых точек $t_1, t_2 \in [0, 1]$ выполнено неравенство $|f(t_1) - f(t_2)| \leqslant L|t_1 - t_2|$.

Таким образом, расстояние между значениями функции в любых двух точках не более чем в L раз превосходит расстояние между точками. Липшицеву функцию с константой L будем для краткости называть L-липшицевой[1]. Для множеств гладких функций мы используем стандартные обозначения: функция f принадлежит классу $C^k[a, b]$, если она k раз дифференцируема и ее производная $f^{(k)}$ непрерывна на отрезке $[a, b]$. При этом мы считаем, что $C^0 = C$ — класс непрерывных функций. Так же определяется класс $C^k(\mathbb{R})$. Класс $C^{(\infty)}$ состоит из бесконечно дифференцируемых функций, т. е. функций, у которых существуют производные всех порядков.

Легко доказывается (с использованием теоремы о среднем), что если у функции $f \in C^1$ норма производной не превосходит константы L: $\|f'\| \leqslant L$, то f является липшицевой с константой L. Поэтому любая функция из класса $C^1[a, b]$, а значит, и из $C^k[a, b]$ при любом $k \geqslant 1$ липшицева. То же верно и для периодических функций на всей прямой $\mathbb{R}$. Вывод, к которому мы пришли, кратко запишем так.

Все функции класса C^k, $k \geqslant 1$ (на отрезке или периодические на $\mathbb{R}$), липшицевы.

Поэтому все, что верно для липшицевых функций, верно и для всех функций из C^k при всех $k \geqslant 1$. Если функция f липшицева, то соответствующий функции вектор $x(k) = f\left(\frac{k}{N}\right)$ также будем называть L-липшицевым. Он определяется свойством $|x(k+1) - x(k)| \leqslant \frac{L}{N}$ для всех $k = 0, \ldots, N-2$. Таким образом, соседние компоненты вектора x отличаются не более чем на число $\frac{L}{N}$. Много ли N-мерных липшицевых векторов? Чтобы дать определенный ответ, надо ограничить множество векторов. Поэтому мы рассмотрим множество векторов, координаты которых неотрицательны и не превосходят 1.

[1] Рудольф Липшиц (Rudolf Otto Sigismund Lipschitz, 1832—1903) — немецкий математик. Был учителем знаменитого Феликса Клейна. Потомкам известен главным образом по названию класса функций.

Все такие векторы образуют N-мерный куб со стороной 1. Мы будем обозначать этот куб I_N, а иногда просто I. Его объем, конечно же, равен 1. Возьмем для определенности $L=1$ и построим явно множество всех 1-липшицевых векторов, лежащих в этом кубе. Обозначим через Λ_N множество всех 1-липшицевых векторов, у которых $x(0)\in[0,1]$. Его пересечение с кубом I_N и будет множеством, которое нас интересует. Чему равен объем множества Λ_N? Начнем с малых N.

***N* = 1**. В этом случае $\Lambda_1=[0,1]$, $\operatorname{Vol}\Lambda_1=1$.

***N* = 2**. Имеем

$$\Lambda_2=\left\{(x(0),x(1))\in\mathbb{R}^2 \mid 0\leqslant x(0)\leqslant 1,\ -\frac{1}{2}\leqslant x(1)-x(0)\leqslant\frac{1}{2}\right\}.$$

Таким образом, Λ_2 — параллелограмм с вершинами

$$\left(0,-\frac{1}{2}\right),\ \left(0,\frac{1}{2}\right),\ \left(1,\frac{1}{2}\right),\ \left(1,\frac{3}{2}\right).$$

Его объем $\operatorname{Vol}\Lambda_2$ равен 1.

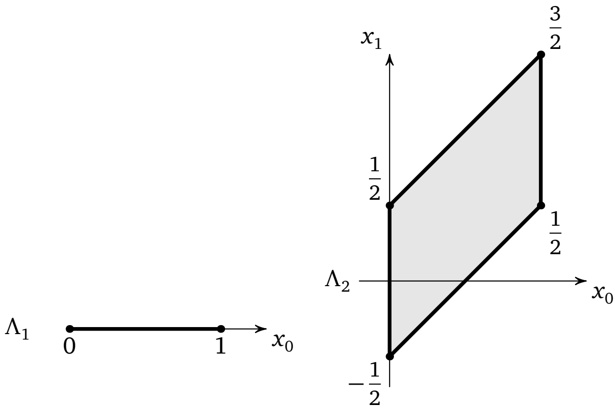


Рис. 3. Множество липшицевых сигналов для $N=1,2$

***N* = 3**. Имеем

$$\Lambda_3=\left\{(x(0),x(1),x(2))\in\mathbb{R}^3 \mid 0\leqslant x(0)\leqslant 1,\ -\frac{1}{3}\leqslant x(1)-x(0)\leqslant\frac{1}{3},\right.$$
$$\left. -\frac{1}{3}\leqslant x(2)-x(1)\leqslant\frac{1}{3}\right\}.$$

Тогда Λ_3 — параллелепипед с вершинами

$$\left(0,-\frac{1}{3},-\frac{2}{3}\right),\left(0,-\frac{1}{3},0\right),\left(0,\frac{1}{3},0\right),\left(0,\frac{1}{3},\frac{2}{3}\right),$$
$$\left(1,\frac{2}{3},\frac{1}{3}\right),\left(1,\frac{2}{3},1\right),\left(1,\frac{4}{3},1\right),\left(1,\frac{4}{3},\frac{5}{3}\right).$$

Его объем Vol Λ_3 равен $\dfrac{4}{9}$.

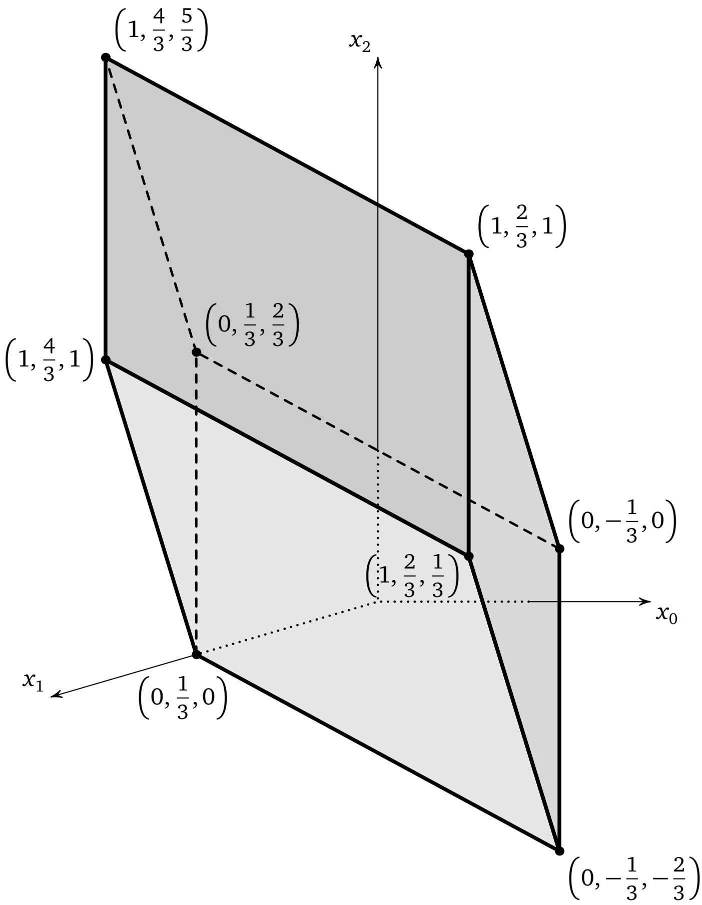


Рис. 4. Множество липшицевых сигналов для $N=3$

Таким образом, Λ_N — параллелепипед, который строится по индукции: сначала берется одномерный параллелепипед — отрезок

$[0, 1]$, затем двумерный: каждой вершине a отрезка (где a — это 0 или 1) ставятся в соответствие две вершины нового параллелепипеда $\left(a, a - \frac{1}{N}\right)$ и $\left(a, a + \frac{1}{N}\right)$ и т. д. На последнем шаге каждая вершина $(N-1)$-мерного параллелепипеда $(a_0, \ldots, a_{N-2})$ порождает две вершины $\left(a_0, \ldots, a_{N-2}, a_{N-2} - \frac{1}{N}\right)$ и $\left(a_0, \ldots, a_{N-2}, a_{N-2} + \frac{1}{N}\right)$ параллелепипеда Λ_N. Объем параллелепипеда Λ_N легче всего найти, используя известный принцип.

Принцип Кавальери. *Если два N-мерных тела таковы, что любая прямая, параллельная заданной прямой, пересекает их по отрезкам одинаковой длины, то эти тела имеют равный объем.*

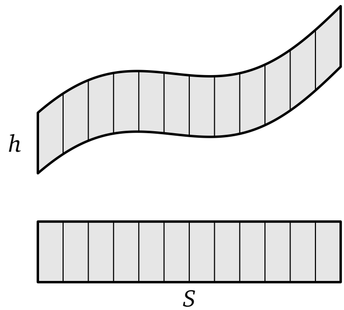


Рис. 5. Принцип Кавальери: $\mathrm{Vol} = S \cdot h$

Мы применим принцип Кавальери к параллелепипеду Λ_N и к N-мерному параллелепипеду с основанием, равным проекции Λ_N на плоскость первых $N-1$ координат, и боковым ребром, равным $\frac{2}{N}$ и перпендикулярным основанию. Ясно, что объем второго параллелепипеда равен $(N-1)$-мерному объему основания, умноженному на $\frac{2}{N}$. Любая прямая, параллельная боковому ребру и пересекающая параллелепипед, пересекает его по отрезку длины $\frac{2}{N}$. С другой стороны, эта прямая пересекает Λ_N также по отрезку длины $\frac{2}{N}$. В самом деле, если $(a_0, \ldots, a_{N-2}, a_{N-1})$ — любая точка параллелепипеда Λ_N, то прямая $\{(a_0, \ldots, a_{N-2}, t) \in \mathbb{R}^N \mid t \in \mathbb{R}\}$ пересекает Λ_N по отрезку с концами $\left(a_0, \ldots, a_{N-2}, a_{N-2} - \frac{1}{N}\right)$ и $\left(a_0, \ldots, a_{N-2}, a_{N-2} + \frac{1}{N}\right)$. Длина этого отрезка равна $\frac{2}{N}$. Следовательно, объем параллелепипеда Λ_N равен объему его проекции, умноженному на $\frac{2}{N}$. Отсюда по индукции заключаем, что при любом $j = 1, \ldots, N-1$ объем проекции множества Λ_N на первые j координат равен $\left(\frac{2}{N}\right)^j$ (естественно, имеется в виду j-мерный объем). При $j = N-1$ получаем

$$\mathrm{Vol}(\Lambda_N) = \left(\frac{2}{N}\right)^{N-1}.$$

Итак, множество всех N-мерных векторов с координатами от 0 до 1 является единичным кубом I_N и имеет объем 1, в то время как множество 1-липшицевых векторов $I \cap \Lambda_N$ имеет объем, меньший чем $\mathrm{Vol}(\Lambda_N) = \left(\frac{2}{N}\right)^{N-1}$.

Уже при $N = 10$, как мы видим, 1-липшицевы векторы составляют примерно одну двухмиллионную объема всех векторов, а при $N = 20$ эта доля равна 10^{-19}. Напомним, что нас интересуют размерности больше $N = 10^6$, при которых этот мизерный объем уже невозможно представить. Мы пришли к главному выводу.

При больших N липшицевы сигналы составляют ничтожную часть от всех сигналов.

Поэтому можно ожидать, что для описания этой ничтожной части нужно гораздо меньше информации и меньше памяти. Математически это означает, что можно найти небольшое число сигналов (значительно меньшее, чем N), комбинациями которых можно приблизить любой липшицев сигнал. Тогда вместо того, чтобы хранить N координат вектора $x \in \mathbb{R}^N$, можно будет обойтись маленьким числом коэффициентов такой комбинации.

Остается только выбрать это небольшое множество сигналов, которыми можно приближать липшицевы сигналы. Задачей нахождения таких систем функций занимается специальная отрасль математики — *теория приближений*. Главный ее принцип: различные классы «хороших» функций (например, дифференцируемых, гладких, аналитических и т. д.) можно приближать с помощью небольшого числа специальных *базисных* функций. В качестве приближающих комбинаций, как правило, используются алгебраические многочлены (базисные функции $\{t^k\}_{k\geqslant 0}$) и тригонометрические многочлены (базисные функции $\{\sin 2\pi kx\}_{k\geqslant 1}$ и $\{\cos 2\pi kx\}_{k\geqslant 1}$). Сама возможность такого приближения уже нами осознана на примере липшицевых сигналов. Как мы видим, 1-липшицевы сигналы образуют параллелепипед очень малого объема в множестве всех сигналов. А дальше мы применяем следующий геометрический факт.

Выпуклое тело малого объема в $\mathbb{R}^N$ можно приблизить плоскостью меньшей размерности, причем чем меньше объем, тем лучше приближение.

Приблизить — значит провести плоскость так, чтобы она проходила рядом с каждой точкой выпуклого тела, т. е. расстояние от каждой точки тела до плоскости было мало. Плоскость можно задать, выбрав порождающую ее систему векторов — *базис*. Если размерность этой плоскости равна n (напомним, что $n < N$), то базис состоит из n векторов. Тогда каждую точку тела можно приближенно задать n числами — координатами близкой к ней точки плоскости. Так мы обходимся меньшим количеством чисел для описания точек N-мерного выпуклого тела.

Пример 1. Найти (приближенно) минимальное значение функции $f(x, y) = 2x^2 + 3y$ при условии

$$10000(x-y)^2 + (x+y)^2 \leqslant 4.$$

Решение. Множество пар чисел (x, y), удовлетворяющих этому неравенству, ограничено эллипсом, изображенным на рис. 6. Его центр находится в начале координат, а полуоси направлены в точки $(1, 1)$ и $\left(-\frac{1}{100}, \frac{1}{100}\right)$. Его объем очень мал (он равен $\pi\sqrt{2}\cdot\frac{\sqrt{2}}{100} = \frac{\pi}{50}$). Значит, его можно хорошо приблизить плоскостью меньшей размерности (в данном случае — приблизить прямой).

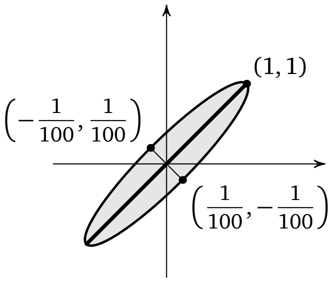


Рис. 6

В самом деле, приближающая прямая — биссектриса координатного угла. Расстояние от эллипса до этой прямой не превосходит ширины эллипса (т. е. длины его малой полуоси), равной $\frac{\sqrt{2}}{100}$, а сам эллипс практически совпадает с отрезком $\{(t, t) \mid t \in [-1, 1]\}$. Поэтому $f(x, y) \approx 2t^2 + 3t$ и минимум этой функции на отрезке $[-1, 1]$ равен $-\frac{9}{8}$. Приближенный ответ получен. Точный отличается от него меньше чем на 0,02.

В этом примере нам помогла картинка, на которой видно приближающую прямую. Если же размерность N большая, то мы не видим выпуклого тела. Мы лишь знаем, что у него маленький объем и, значит, его можно приблизить плоскостью. Как найти эту плоскость и как (что даже важнее!) выбрать в ней базис? Как, например, выбрать базис для приближений параллелепипеда 1-лишицевых векторов? Сколько в этом базисе должно быть элементов $\{\varphi_1, \ldots, \varphi_n\}$, чтобы порожденная ими плоскость приближала параллелепипед с точностью, скажем, $\varepsilon = 0{,}01$?