Математический анализ

Дима Трушин

Семинар 2

Последовательности

Пусть S – множество всех последовательностей на множестве натуральных чисел, то есть $S = \{a\colon \mathbb{N} \to \mathbb{R}\}$. То есть, когда я хочу сказать, что a_n – последовательность определенная для всех $n \in \mathbb{N}$, я могу написать, что $a \in S$. Замечу, что S является векторным пространством над $\mathbb{R}$, потому что последовательности можно складывать и умножать на числа.

Обратите внимание, что значение a на элементе n можно обозначать двумя способами a_n или $a(n)$. Я буду пользоваться и той и другой записью в зависимости от удобства. Так же для удобства можно считать, что $a_n = 0$ при $n < 0$.[1]

Дискретное дифференцирование и интегрирование

Дискретным дифференцированием будем называть следующую операцию $\Delta\colon S \to S$ по правилу

$$\Delta a_n = a_{n+1} - a_n$$

Про это отображение нужно думать, как про дифференцирование в точке n. Дискретным интегрированием назовем отображение $I\colon S \to S$ по правилу

$$Ia_n = \sum_{k=0}^{n-1} a_k$$

Про это отображение надо думать, как про интегрирование от 0 до n. В табличке эти операции можно изобразить следующим образом

a	a_0	a_1	$\ldots$	a_n	$\ldots$
Δa	$a_1 - a_0$	$a_2 - a_1$	$\ldots$	$a_{n+1} - a_n$	$\ldots$
Ia	0	a_0	$\ldots$	$\sum_{k=0}^{n-1} a_k$	$\ldots$

Прямая проверка показывает, что

1. Отображение $\Delta\colon S \to S$ является линейным оператором.

2. Отображение $I\colon S \to S$ является линейным оператором.

3. Дискретное правило Лейбница

$$\Delta(ab)_n = (\Delta a_n)b_n + a_{n+1}\Delta b_n = (\Delta a_n)b_{n+1} + a_n\Delta b_n$$

   Обратите внимание, что в дискретном случае получаются два правила Лейбница из-за того, что можно сдвинуть один из двух аргументов.

4. Для любой последовательности $a \in S$ верно $(\Delta I a)_n = a_n$ для любого $n \in \mathbb{N}$. То есть если мы сначала проинтегрируем последовательность, а потом продифференцируем, то получим исходную последовательность.

5. Для любой последовательности $a \in S$ верно $(I\Delta a)_n = a_n - a_0$ для любого $n \in \mathbb{N}$. То есть если мы сначала продифференцируем последовательность, а потом проинтегрируем, то мы из исходной последовательности вычтем постоянную последовательность равную a_0 в каждом члене.

[1] Я считаю, что натуральные числа $\mathbb{N}$ начинаются с нуля.

Полиномиальные последовательности

Обозначим через P множество последовательностей, задаваемых полиномами, то есть $P = \{p\colon \mathbb{N} \to \mathbb{R} \mid p \in \mathbb{R}[x]\}$. То есть последовательность $a \in P$ тогда и только тогда, когда найдется многочлен $p \in \mathbb{R}[x]$ такой, что $a_n = p(n)$. Так как полином степени d однозначно задается по $d+1$ точке, то из равенства $p(n) = q(n)$ во всех точках $n \in \mathbb{N}$ для некоторых полиномов p и q, следует, что $p = q$ как многочлен. То есть для последовательности $a_n = p(n)$ многочлен p определен однозначно. Если d – степень многочлена p, то мы будем говорить, что последовательность a имеет степень d. Множество полиномиальных последовательностей степени не выше d будем обозначать через P_d.

Давайте посчитаем дискретную производную от последовательности n^d, получим

$$\Delta(n^d)_n = (n+1)^d - n^d = \sum_{k=0}^{d} C_d^k n^k - n^d = \sum_{k=0}^{d-1} C_d^k n^d = C_d^1 n^{d-1} + \ldots$$

Ядро и образ дискретного дифференцирования

Таким образом Δ задает отображение $\Delta\colon P_d \to P_{d-1}$. Я утверждаю, что

1. Если $a_n \in P_d$ задается в виде $a_n = cn^d + \ldots$, то $\Delta a_n = cdn^{d-1} + \ldots$.
2. $\ker \Delta = \mathbb{R}$, то есть $\dim \ker \Delta = 1$.
3. $\operatorname{Im} \Delta = P_{d-1}$.

Первое следует из предыдущего примера. Второе следует из явной формулы. Давайте объясним последнее. Мы знаем, что $\dim \ker \Delta + \dim \operatorname{Im} \Delta = \dim P_d = d + 1$. Значит, $\dim \operatorname{Im} \Delta = d$. Но $\operatorname{Im} \Delta$ является подпространством в P_{d-1}, чья размерность тоже d. А значит $\operatorname{Im} \Delta = P_{d-1}$.

Последнее утверждение означает, что для любой полиномиальной последовательности a_n степени d найдется последовательность b_n степени $d+1$ такая, что $\Delta b = a$, причем все такие последовательности отличаются друг от друга на константу.

Ядро и образ дискретного интегрирования

Я утверждаю, что I задает инъективное отображение $I\colon P_{d-1} \to P_d$, то есть $\ker I = 0$ при этом образ I совпадает со всеми полиномиальными последовательностями зануляющимися в нулевом члене, то есть последовательности задаваемые многочленами вида $xp(x)$, где $\deg p = d - 1$. Давайте явно перечислим эти свойства:

1. Если $a_n \in P_{d-1}$ имеет вид $a_n = cn^{d-1} + \ldots$, то $Ia_n = \frac{c}{d} n^d + \ldots$.
2. $\ker I = 0$.
3. $\operatorname{Im} I = \{a \in P_d \mid a_0 = 0\}$.

Действительно, пусть $a \in P$ – последовательность степени d. Из-за того, что $\Delta I a = a$, то I инъективно.

Теперь, $b = Ia$ – это такая последовательность, что $\Delta b = a$, а из предыдущего раздела мы знаем, что степень b на единичку больше, чем степень a. Значит I действительно задает отображение $I\colon P_{d-1} \to P_d$. По явной формуле для I мы видим, что $Ia_0 = 0$ всегда. Потому образ лежит в подпространстве $\{a \in P_d \mid a_0 = 0\}$ и размерность этого подпространства равна $d = \dim P_d - 1$. С другой стороны, так как I инъективно, то размерность образа равна размерности P_{d-1}, которая есть d. А значит $\operatorname{Im} I = \{a \in P_d \mid a_0 = 0\}$, что и требовалось.

Теперь покажем первое свойство. Мы знаем, что $\Delta(n^d) = dn^{d-1} + \ldots$, а значит $\Delta(\frac{c}{d} n^d) = cn^{d-1} + g(n)$, где $\deg g < d - 1$. Так как мы знаем, что найдется многочлен h степени $d - 1$ такой, что $\Delta h = g$, то мы видим, что $\Delta(\frac{c}{d} n^d + h(n)) = cn^d$. А теперь воспользуемся тем, что $I(cn^d)$ отличается от $\frac{c}{d} n^d + h(n)$ на константу и получим требуемое.

Применение для вычисления асимптотик

Пусть $z \in \mathbb{R}$ – произвольное число, рассмотрим последовательность

$$a_n = \sum_{k=0}^{n-1} (k+z)^m$$

Заметим, что $a = Ib$, где $b_n = (n+z)^m$. Тогда по свойствам дискретного интегрирования, мы видим, что $a_n = \frac{1}{m+1} n^{m+1} + g(n)$, где g – какой-то многочлен степени не больше m. Кроме того, тут можно заметить, что g будет многочленом от n и от z. В частности это будет непрерывная функция по z.

Равномерная сходимость и интегралы

Пусть $f_n(x)$ – последовательность непрерывных функций на отрезке $[0,1]$.[2] И пусть она поточечно сходится к некоторой непрерывной функции $f(x)$ на этом отрезке. Тогда оказывается, что f_n сходится равномерно к f. Это очень полезное утверждение. Главный бонус от него заключается вот в чем. Если на отрезке $[0,1]$ последовательность непрерывных функций f_n равномерно сходится к функции f, то функция f тоже непрерывны и верно равенство

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = \int_0^1 \lim_{n\to\infty} f_n(x)\,dx = \int_0^1 f(x)\,dx$$

Пример задачи

Давайте теперь найдем следующий интеграл

$$\int_0^1 e^{\{nx\}} x^m\,dx$$

А потом посчитаем его предел при $n \to \infty$. Здесь $\{x\}$ – дробная часть числа x, то есть разница между x и наибольшим целым не превосходящим x, то есть $\{x\} = x - [x]$.

В начале разобьем интеграл в сумму

$$\int_0^1 e^{\{nx\}} x^m\,dx = \sum_{k=0}^{n-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} e^{\{nx\}} x^m\,dx = \sum_{k=0}^{n-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} e^{nx-k} x^m\,dx$$

Теперь сделаем замену $nx - k = z$ в каждом из этих интегралов:

$$\sum_{k=0}^{n-1} \int_0^1 e^z \left(\frac{z+k}{n}\right)^m d\left(\frac{z+k}{n}\right) = \frac{1}{n^{m+1}} \int_0^1 e^z \sum_{k=0}^{n-1} (k+z)^m\,dz$$

А мы уже посчитали, что выражение $\sum_{k=0}^{n-1}(k+z)^m$ имеет вид $\frac{1}{m+1}n^{m+1} + g(n,z)$, где $g(n,z)$ – многочлен по n и по z, причем по n его степень не превосходит m. Значит мы имеем интеграл

$$\int_0^1 e^z \frac{\frac{1}{m+1}n^{m+1} + g(n,z)}{n^{m+1}}\,dz = \int_0^1 e^z \left(\frac{1}{m+1} + \frac{g(n,z)}{n^{m+1}}\right) dz$$

Теперь мы видим, что $\frac{g(n,z)}{n^{m+1}}$ является непрерывной функцией на отрезке $[0,1]$, которая поточечно сходится на этом отрезке к 0 при $n \to \infty$. А значит из-за компактности отрезка эта функция сходится к 0 равномерно. А равномерную сходимость можно интегрировать. Значит

$$\lim_{n\to\infty} \int_0^1 e^z \left(\frac{1}{m+1} + \frac{g(n,z)}{n^{m+1}}\right) dz = \int_0^1 e^z \lim_{n\to\infty} \left(\frac{1}{m+1} + \frac{g(n,z)}{n^{m+1}}\right) dz = \int_0^1 e^z \left(\frac{1}{m+1}\right) dz = \frac{e-1}{m+1}$$

[2] Можно взять любой отрезок, $[0,1]$ выбран исключительно из эстетических соображений.

Оптимизационная задача

Давайте рассмотрим следующую задачу: нам даны функции $f, g_1, \ldots, g_k \colon \mathbb{R}^n \to \mathbb{R}$ и мы хотим найти минимум или максимум функции f при условии, что все остальные функции g_i равны нулю.[3] То есть мы хотим решить следующую задачу

$$\begin{cases} f(x_1, \ldots, x_n) \to \min \\ g_1(x_1, \ldots, x_n) = 0 \\ \ldots \\ g_k(x_1, \ldots, x_n) = 0 \end{cases} \quad \text{или} \quad \begin{cases} f(x_1, \ldots, x_n) \to \max \\ g_1(x_1, \ldots, x_n) = 0 \\ \ldots \\ g_k(x_1, \ldots, x_n) = 0 \end{cases}$$

В качестве решения мы хотим предъявить точку $x \in \mathbb{R}^n$ такую, что $f(x) \leqslant f(y)$ (для второй задачи $f(x) \geqslant f(y)$) для любой $y \in \mathbb{R}^n$ такой, что $g_1(y) = \ldots = g_k(y) = 0$. Стандартный рецепт решения такой задачи следующий:

1. Надо составить так называемый Лагранжиан

   $$\mathcal{L} = \alpha f(x) - \lambda_1 g_1(x) - \ldots - \lambda_k g_k(x)$$

   где $\alpha, \lambda_1, \ldots, \lambda_k \in \mathbb{R}$ – какие-то неизвестные константы.

2. Далее говорится, что если точка $x \in \mathbb{R}^n$ является точкой минимума или максимума f при условии, что все g_i равны нулю, то найдутся такие константы $\alpha, \lambda_1, \ldots, \lambda_k \in \mathbb{R}$, что точка x будет критической для $\mathcal{L}$.

3. Это значит, что мы перебираем все возможные значения констант $\alpha, \lambda_1, \ldots, \lambda_k$ и для них решаем уравнения:

   $$\frac{\partial \mathcal{L}}{\partial x_1} = \ldots = \frac{\partial \mathcal{L}}{\partial x_n} = 0$$

Самое удивительное во всем этом – оно работает! Но весь вопрос в том, а почему оно работает? Давайте сделаю пару замечаний, от которых у вас должны закрасться обоснованные сомнения. Например, что если мы рассмотрим случай $\alpha = 0$? Тогда Лагранжиан вообще не зависит от функции f и мы получаем какие-то странные условия на функции g_i. Особенно это звучит странно, если среди g_i есть повторяющиеся функции, мы получим очень странное условие, что все $x \in \mathbb{R}^n$ нам подойдут в качестве подозрительных точек. Давайте обсудим, что тут происходит, какой у всего этого геометрический смысл и как не запутаться в куче странных условий. Надеюсь, что после геометрического описания, станет понятно, почему эта штука вообще работает.

Наглядный частный случай

Движение по касательному градиенту Давайте для простоты объяснения рассмотрим случай всего двух переменных и одной вспомогательной функции, а именно

$$\begin{cases} f(x, y) \to \min \\ g(x, y) = 0 \end{cases} \quad \text{или} \quad \begin{cases} f(x, y) \to \max \\ g(x, y) = 0 \end{cases}$$

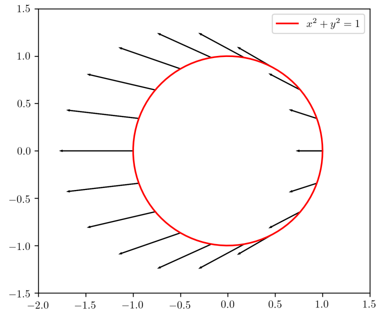
Более того, давайте рассмотрим конкретный пример

$$f(x, y) = (x-2)^2 + y^2 \quad \text{и} \quad g(x, y) = x^2 + y^2 - 1$$

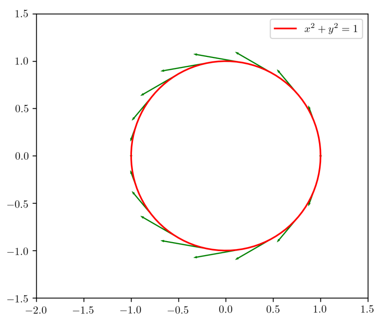
Тогда условие $g(x, y) = 0$ задает нам окружность радиуса 1 с центром в начале координат, а функция $f(x, y)$ имеет глобальный минимум в точке $(2, 0)$, которая не лежит на этой окружности. Давайте думать про нашу функцию $f(x, y)$ так: ее минимум – это источник жидкости, а жидкость из него вытекает и бежит вдоль градиента в сторону максимума (в реальной жизни жидкость бежит вниз к земле, а у нас она будет тянуться вверх). Давайте нарисуем поле градиента для функции $f(x, y)$, но не на всей плоскости а только на окржуности $g(x, y) = 0$.[4]

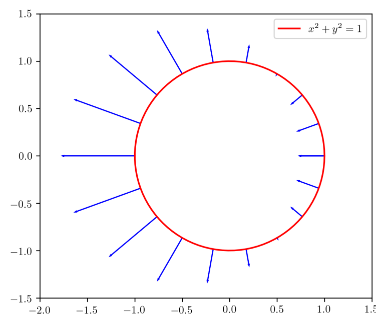
[3] Я для простоты предположил, что все функции определены в $\mathbb{R}^n$, но вообще говоря достаточно какой-то области $U \subseteq \mathbb{R}^n$. Технически задачи ничем не отличаются, однако, изложение становится проще и яснее.

[4] Для читаемости на картинке изображено поле градиента для функции $\frac{1}{8} f(x, y)$, иначе векторы слишком длиннющие получаются.



Теперь думать надо так. Наша невесомая жидкость бежит из истока по градиентному полю на плоскости. А теперь мы поставим границы для нашей жидкости в виде равенства $g(x, y) = 0$ и разрешим бежать только по ней. В этом случае градиент $\operatorname{grad} f$ будет иметь две составляющие: вдоль окружности (касательная компонента отмечена зеленым цветом) и перпендикулярно окружности (ортогональная компонента отмечена синим цветом). Вторая часть будет гаситься тем, что мы запрещаем жидкости покидать окружность, значит наша жидкость будет плыть только за счет касательной составляющей.



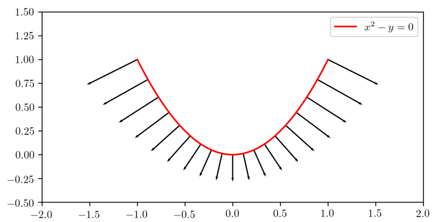


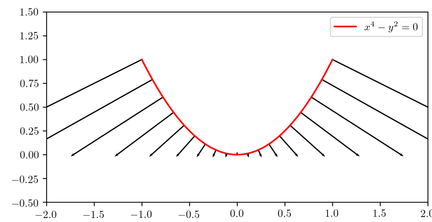
Как видно из рисунка справа, минимум и максимум достигаются там, где градиент функции f ортогонален окружности. Но с другой стороны, $\operatorname{grad} g$ – это вектор, который тоже ортогонален окружности в каждой точке по его геометрическому смыслу (окружность – это линия уровня для g). А значит $f = \lambda \operatorname{grad} g$ для некоторого числа λ.

Плохая параметризация Однако, может случиться неприятность. Что если мы плохо параметризовали нашу кривую и градиент g не всегда является ненулевым вектором? Давайте попробуем минимизировать функцию $f(x, y) = x^2 + (y + 1)^2$ на параболе заданной в первом случае с помощью $g_1(x, y) = x^2 - y$, а во втором случае с помощью $g_2(x, y) = x^4 - y^2$. То есть получаем две оптимизационные задачи

$$\begin{cases} f(x,y) \to \min \\ g_1(x,y) = 0 \end{cases} \quad \text{и} \quad \begin{cases} f(x,y) \to \min \\ g_2(x,y) = 0 \end{cases}$$

Заметим, что функция f, при ограничении на параболу, принимает минимум в начале координат. Но при этом $\operatorname{grad} f$ не равен нулю в начале координат и смотрит вниз. Теперь давайте посмотрим, а что же происходит с градиентами $\operatorname{grad} g_1$ и $\operatorname{grad} g_2$. Они изображены на рисунках ниже:





В частности мы видим, что для g_1 градиент в точке $(0,0)$ не равен нулю и мы можем любой вектор ортогональный параболе в этой точке представить в виде $\lambda \operatorname{grad} g_1$. С другой стороны, если поглядеть на правый рисунок, то видно, что $\operatorname{grad} g_2$ в точке $(0,0)$ равен нулю. Потому мы никак не сможем представить $\operatorname{grad} f$ в виде $\lambda \operatorname{grad} g_2$ в начале координат.

Теперь давайте составим лагранжианы этих задач в наивном виде $\mathcal{L}_i = f - \lambda g_i$. Получим

$$\begin{cases} \mathcal{L}_1(x,y) = x^2 + (y+1)^2 - \lambda(x^2 - y) \\ \dfrac{\partial \mathcal{L}}{\partial x}(x,y) = 2x - 2\lambda x = 0 \\ \dfrac{\partial \mathcal{L}}{\partial y}(x,y) = 2(y+1) + \lambda = 0 \end{cases} \quad \text{и} \quad \begin{cases} \mathcal{L}_2(x,y) = x^2 + (y+1)^2 - \lambda(x^4 - y^2) \\ \dfrac{\partial \mathcal{L}}{\partial x}(x,y) = 2x - 4\lambda x^3 = 0 \\ \dfrac{\partial \mathcal{L}}{\partial y}(x,y) = 2(y+1) + 2\lambda y = 0 \end{cases}$$

Тогда мы видим, что для точки $(0,0)$ можно подобрать $\lambda = -2$, чтобы она стала решением первой системы. С другой стороны, для второй системы при любом параметре λ точка $(0,0)$ не является решением. Таким образом, когда Лагранжиан записывают в виде $\mathcal{L} = \alpha f - \lambda g$, это делается для того, чтобы исключить плохие параметризации. Если $\alpha = 0$, $\operatorname{grad} \mathcal{L} = 0$ превращается в условие $\operatorname{grad} g = 0$ и это означает, что надо проверить точки, которые плохо запараметризованы с помощью g. А если $\alpha \neq 0$, то мы можем заменить его на 1 и рассматривать обычную функцию Лагранжа вида $\mathcal{L} = f - \lambda g$.

Интуиция для многомерного случая В общем виде, когда нам задана задача

$$\begin{cases} f(x) \to \min \\ g_1(x) = 0 \\ \dots \\ g_k(x) = 0 \end{cases} \quad \text{где } x \in \mathbb{R}^n$$

Условия $g_1(x) = \ldots = g_k(x) = 0$ задают некоторую «поверхность». Если в точке x, удовлетворяющей этой системе, градиенты $\operatorname{grad} g_1, \ldots, \operatorname{grad} g_k$ линейно независимы, то оказывается, в окрестности точки x эти уравнения задают поверхность размерности $n-k$ (гладко изогнутое подпространство размерности $n-k$). Последнее утверждение – это одна из форм теоремы о неявной функции. Потому, когда мы задаем Лагранжиан в виде $\mathcal{L} = \alpha f - \sum_{i=1}^k \lambda_i g_i$, мы подразумеваем, что у нас есть два случая:

- $\alpha = 0$. Тогда условие для критической точки функции Лагранжа $\operatorname{grad} \mathcal{L} = 0$ означает, что для некоторых λ_i имеет место равенство $\lambda_1 \operatorname{grad} g_1 + \ldots + \lambda_k \operatorname{grad} g_k = 0$. То есть это случай линейно зависимых градиентов, случай плохой параметризации.

- $\alpha \neq 0$, а значит можно считать $\alpha = 1$. В этом случае можно рассматривать задачу с функцией $\mathcal{L} = f - \lambda_1 g_1 + \ldots + \lambda_k g_k$. Тогда условие для критической точки функции Лагранжа $\operatorname{grad} \mathcal{L} = 0$ означает, что для некоторых λ_i имеет место равенство $\operatorname{grad} f = \lambda_1 \operatorname{grad} g_1 + \ldots + \lambda_k \operatorname{grad} g_k$. То есть это случай, когда $\operatorname{grad} f$ ортогонален нашей поверхности. Если думать в терминах потока, то в этой точке поток не двигается вдоль поверхности.

Полный набор уравнений для наглядности Пусть мы решаем задачу

$$\begin{cases} f(x) \to \min \\ g_1(x) = 0 \\ \dots \\ g_k(x) = 0 \end{cases} \quad \text{где } x \in \mathbb{R}^n$$

Тогда составим функцию $\mathcal{L} = \alpha f - \lambda_1 g_1 + \ldots + \lambda_k g_k$. В этом случае точку минимума надо искать среди решений следующей системы

$$\begin{cases} \operatorname{grad} \mathcal{L}(x) = 0 \\ g_1(x) = 0 \\ \dots \\ g_k(x) = 0 \end{cases} \quad \text{где } x \in \mathbb{R}^n$$

При этом мы перебираем все возможные значения параметров α, $\lambda_1, \ldots, \lambda_k$. Как говорилось выше, для параметра α можно считать, что его значение либо 0 либо 1.

Оптимизационная задача с неравенствами

Давайте рассмотрим следующую задачу: нам даны функции $f, g_1, \ldots, g_k, h_1, \ldots, h_s \colon \mathbb{R}^n \to \mathbb{R}$ и мы хотим найти минимум или максимум функции f при условии, что функции g_i равны нулю и выполнены неравенства $h_i \leqslant 0$. То есть мы хотим решить следующую задачу

$$\begin{cases} f(x_1, \ldots, x_n) \to \min \\ g_1(x_1, \ldots, x_n) = 0 \\ \dots \\ g_k(x_1, \ldots, x_n) = 0 \\ h_1(x_1, \ldots, x_n) \leqslant 0 \\ \dots \\ h_s(x_1, \ldots, x_n) \leqslant 0 \end{cases} \quad \text{или} \quad \begin{cases} f(x_1, \ldots, x_n) \to \max \\ g_1(x_1, \ldots, x_n) = 0 \\ \dots \\ g_k(x_1, \ldots, x_n) = 0 \\ h_1(x_1, \ldots, x_n) \leqslant 0 \\ \dots \\ h_s(x_1, \ldots, x_n) \leqslant 0 \end{cases}$$

В качестве решения мы хотим предъявить точку $x \in \mathbb{R}^n$ такую, что $f(x) \leqslant f(y)$ (для второй задачи $f(x) \geqslant f(y)$) для любой $y \in \mathbb{R}^n$ такой, что $g_1(y) = \ldots = g_k(y) = 0$ и $h_1(y) = \ldots = h_k(y) \leqslant 0$. Стандартный рецепт решения такой задачи следующий:

1. Надо составить так называемый Лагранжиан

   $$\mathcal{L} = \alpha f(x) - \lambda_1 g_1(x) - \ldots - \lambda_k g_k(x) - \mu_1 h_1(x) - \ldots - \mu_s h_s(x)$$

   где $\alpha, \lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_s \in \mathbb{R}$ – какие-то неизвестные константы. При этом для поиска максимума мы считаем, что $\alpha \geqslant 0$ и $\mu_i \geqslant 0$, а для поиска минимума мы считаем, что $\alpha \geqslant 0$ и $\mu_i \leqslant 0$.

2. Далее говорится, что если точка $x \in \mathbb{R}^n$ является точкой минимума или максимума f при условии, что все g_i равны нулю и все $h_i \leqslant 0$, то найдутся такие константы $\alpha, \lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_s \in \mathbb{R}$, что точка x будет удовлетворять системе (слева для минимума, а справа для максимума)

   $$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x}(x) = 0 \\ g_i(x) = 0 \\ \alpha \geqslant 0 \\ \mu_i h_i(x) = 0 \\ h_i(x) \leqslant 0 \\ \mu_i \leqslant 0 \end{cases} \quad \text{или} \quad \begin{cases} \dfrac{\partial \mathcal{L}}{\partial x}(x) = 0 \\ g_i(x) = 0 \\ \alpha \geqslant 0 \\ \mu_i h_i(x) = 0 \\ h_i(x) \leqslant 0 \\ \mu_i \geqslant 0 \end{cases}$$

В этой задаче еще менее очевидно, почему подобное барахло хоть как-то работает. Особенно получилась куча каких-то странных неравенств на константы, которых не было в предыдущем случае, когда были только уравнения и не было неравенств в системе. Как и раньше, я предлагаю разобрать механизм работы этого метода на конкретном примере малой размерности.

Наглядный пример с неравенствами

Давайте для простоты объяснения рассмотрим случай всего двух переменных и одной вспомогательной функции с неравенством, а именно

$$\begin{cases} f(x,y) \to \min \\ h(x,y) \leqslant 0 \end{cases} \quad \text{или} \quad \begin{cases} f(x,y) \to \max \\ h(x,y) \leqslant 0 \end{cases}$$

Более того, давайте рассмотрим конкретный пример

$$f(x,y) = (x-2)^2 + y^2 \quad \text{и} \quad h(x,y) = x^2 + y^2 - 1$$

В этом случае $h(x,y) \leqslant 0$ задает круг радиуса 1 с центром в нуле и теперь минимум и максимум функции мы ищем не только на окружности, но и внутри круга. По хорошему, это означает, что нам надо решить две разные задачи: 1) найти минимум внутри открытого круга, 2) найти минимум на окружности. После чего надо сравнить эти два минимума и выбрать меньший. Аналогично надо поступить в случае максимума.

Давайте начнем со случая поиска минимума (или максимума) внутри круга. Тогда мы решаем задачу

$$\begin{cases} f(x,y) \to \min \\ h(x,y) < 0 \end{cases} \quad \text{или} \quad \begin{cases} f(x,y) \to \max \\ h(x,y) < 0 \end{cases}$$

Так как условие $h(x,y) < 0$ задает открытое подмножество, то необходимое условие для минимума или максимума – градиент f равен нулю, то есть

$$\begin{cases} \dfrac{\partial f}{\partial x}(x,y) = 0 \\ \dfrac{\partial f}{\partial y}(x,y) = 0 \\ \quad h(x,y) < 0 \end{cases}$$

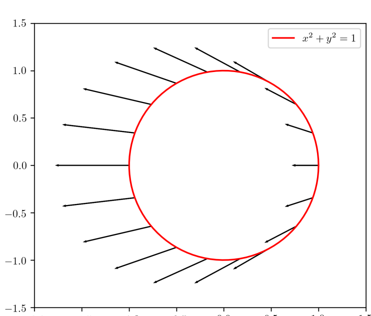
Теперь, если мы ищем минимум или максимум на границе, то задача превращается в предыдущую и мы решаем

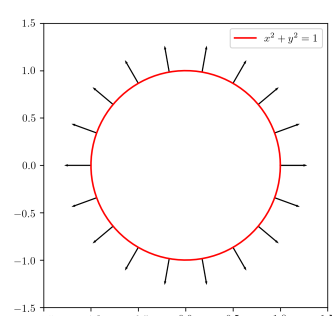
$$\begin{cases} \dfrac{\partial f}{\partial x}(x,y) = \mu \dfrac{\partial h}{\partial x}(x,y) \\ \dfrac{\partial f}{\partial y}(x,y) = \mu \dfrac{\partial h}{\partial y}(x,y) \\ \quad h(x,y) = 0 \end{cases}$$

Я здесь не пишу α, потому что мы уже знаем, что градиент на окружности не зануляется.

Теперь давайте обсудим странные неравенства и равенства, которые возникли в общем случае, на этом примере. Первое неравенство было $\alpha \geqslant 0$. Это условие нужно, чтобы критические точки функций f и αf не сменили своего характера. Точнее, при $\alpha > 0$ точка является минимум для одной из них тогда и только тогда, когда она минимум для другой. А при $\alpha = 0$ у нас от функции в лагранжиане ничего не зависит и этот случай нужен лишь для случая плохой параметризации.

Причина, по которой мы хотим иметь одинаковый характер поведения в критических точках для f и αf очень простой – мы можем распознать минимумы и максимумы в этом случае. Давайте посмотрим на градиенты f и h





Где будет максимум у функции f на окружности? Там где жидкость пытается утечь наружу. А минимум? Там где жидкость пытается затечь внутрь окружности. То есть в точке максимума на границе $\operatorname{grad} f$ и $\operatorname{grad} h$ смотрят в одну сторону, а в точке минимума $\operatorname{grad} f$ и $\operatorname{grad} h$ смотрят в разные стороны. Потому для минимума мы требуем $\mu \leqslant 0$, а для максимума $\mu \geqslant 0$. Случай равенства нулю нужен если эта точка оказалась стационарной для f (например точка была точкой максимума или минимума без ограничения $h = 0$).

Осталось понять зачем нужно условие $\mu h(x, y) = 0$. Но в этом случае все просто. В общем случае, когда у нас много неравенств, нам надо разобрать для каждого уравнения два случая $h_i < 0$ или $h_i = 0$. А как записать эти два условия одновременно?[5] Условие $\mu h(x, y) = 0$ гарантирует, что либо $\mu = 0$ и значит условие минимизации превращается в $\operatorname{grad} f$ (то есть мы ищем точку минимума без ограничений). А если $\mu \neq 0$, то условие $\mu h(x, y) = 0$ означает, $h(x, y) = 0$ и значит мы вышли на границу и условие минимизации означает $\operatorname{grad} f = \mu \operatorname{grad} h$. И для того, чтобы в первом случае $\mu = 0$ мы учли неравенство $h(x, y) < 0$ в систему все равно добавляем уравнение $h(x, y) \leqslant 0$. Лишнее равенство нулю тут ничему не мешает формально, мы лишь учтем минимум на границе дважды, но ничего не потеряем.

[5] Не то чтобы это поможет при решении. Так как решая составленное уравнение, вы все равно будете перебирать случаи. Это вопрос лишь краткой записи и так понятных условий.