

# Условия Каруша-Куна-Такера в задачах оптимизации

Мальковский Н. В.

Санкт-Петербургский Академический Университет



# Задачи оптимизации с ограничениями

Общая задача математической оптимизации – это

$$\begin{array}{ll} \text{минимизировать} & f(x), \\ \text{при условии} & x \in \mathcal{D}. \end{array} \quad (1)$$

Решение задачи – точка  $x^* \in \mathcal{D}$  такая, что  $\forall x \in \mathcal{D} : f(x) \geq f(x^*)$ .

# Задачи оптимизации с ограничениями

Общая задача математической оптимизации – это

$$\begin{array}{ll} \text{минимизировать} & f(x), \\ \text{при условии} & x \in \mathcal{D}. \end{array} \quad (1)$$

Решение задачи – точка  $x^* \in \mathcal{D}$  такая, что  $\forall x \in \mathcal{D} : f(x) \geq f(x^*)$ .

В подавляющем большинстве случаев множества в таких задачах представляются в виде набора уравнений и неравенств:

$$g(x) \leq 0, \quad g(x) < 0, \quad \text{или} \quad g(x) = 0.$$

# Задачи оптимизации с ограничениями

Общая задача математической оптимизации – это

$$\begin{array}{ll} \text{минимизировать} & f(x), \\ \text{при условии} & x \in \mathcal{D}. \end{array} \quad (1)$$

Решение задачи – точка  $x^* \in \mathcal{D}$  такая, что  $\forall x \in \mathcal{D} : f(x) \geq f(x^*)$ .

В подавляющем большинстве случаев множества в таких задачах представляются в виде набора уравнений и неравенств:

$$g(x) \leq 0, \quad g(x) < 0, \quad \text{или} \quad g(x) = 0.$$

Если  $g$  – дифференцируемая функция без нерегулярных точек, то  $g(x) < 0$  задает открытое множество,  $g(x) \leq 0$  задает замкнутое множество, а  $g(x) = 0$  задает гиперповерхность. Для первого случая работает условие стационарности, т.е. если  $x^* = \operatorname{argmin}_{g(x) < 0} f(x)$ , то  $\nabla f(x^*) = 0_n$ . Можно ли получить подобное для равенств и нестрогих неравенств?

# Условия стационарности для задач на выпуклых множествах

## Теорема

Пусть в задаче (1) множество  $\mathcal{D}$  выпукло, а функция  $f$  дифференцируема в точке  $x^*$ , тогда для точки минимума  $x^*$  выполняется условие

$$\nabla f(x^*)^T (x - x^*) \geq 0, \quad \forall x \in \mathcal{D}.$$

# Условия стационарности для задач на выпуклых множествах

## Теорема

Пусть в задаче (1) множество  $\mathcal{D}$  выпукло, а функция  $f$  дифференцируема в точке  $x^*$ , тогда для точки минимума  $x^*$  выполняется условие

$$\nabla f(x^*)^T(x - x^*) \geq 0, \quad \forall x \in \mathcal{D}.$$

**Док-во.** Предположим, что есть точка  $x \in \mathcal{D}$  такая, что  $\nabla f(x^*)^T(x - x^*) < 0$ , тогда в силу дифференцируемости  $f$  при достаточно малых  $\alpha > 0$  получаем

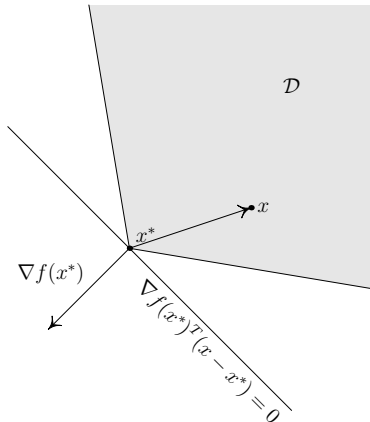
$$f(x^* + \alpha(x - x^*)) - f(x^*) = \alpha \nabla f(x^*)^T(x - x^*) + o(\alpha \|x - x^*\|) \leq$$

$$\alpha \nabla f(x^*)^T(x - x^*) + \alpha \|x - x^*\| \left( \frac{-\nabla f(x^*)^T(x - x^*)}{2\|x - x^*\|} \right) \leq \frac{\alpha}{2} \nabla f(x^*)^T(x - x^*) < 0.$$

Так как  $x, x^* \in \mathcal{D}$ ,  $\mathcal{D}$  – выпукло, то при  $0 < \alpha < 1$  точка  $x^* + \alpha(x - x^*)$  также принадлежит  $\mathcal{D}$ . ■

# Условия стационарности для задач на выпуклых множествах

**Замечание.**  $\nabla f(x^*) \neq 0_n$  допустимо только в случае, если  $x^*$  – граничная точка  $\mathcal{D}$ . Условие можно переформулировать в следующем виде: если  $\nabla f(x^*) \neq 0_n$ , то  $\nabla f(x^*)^T(x - x^*) = 0$  – опорная к  $\mathcal{D}$  гиперплоскость.



# Условный экстремум и множители Лагранжа

Пусть  $\mathcal{D} \subset \mathbb{R}^n$ ,  $\mathcal{D}$  открыто,  $f : \mathcal{D} \rightarrow \mathbb{R}$ ,  $g : \mathcal{D} \rightarrow \mathbb{R}^m$  – непрерывно дифференцируемые функции. Рассмотрим задачу:

$$\begin{array}{ll} \text{минимизировать} & f(x), \\ \text{при условии} & g(x) = 0_m. \end{array} \quad (2)$$

функцией Лагранжа (или лагранжианом) соответствующей задаче (2) называется функция

$$L(\lambda, x) = f(x) + \lambda^T g(x), \quad \lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)^T.$$

Коэффициенты  $\lambda_i$  принято называть множителями Лагранжа.

## Теорема

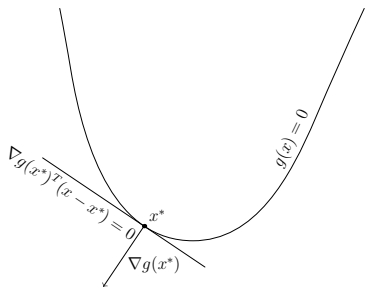
$x^*$  – точка минимума задачи (2) и векторы  $\nabla g_i(x^*)$  линейно независимы, тогда существует такой вектор  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)^T$ , что

$$\nabla_x L(\lambda, x^*) = 0_n.$$

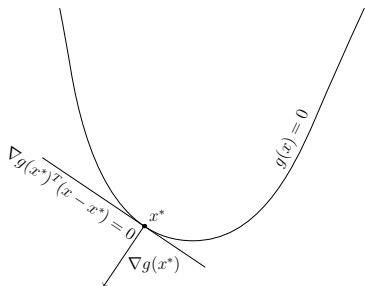


# Геометрическая интерпретация множителей Лагранжа

- Вблизи  $x^*$  поверхность  $g(x) = 0$  почти совпадает с гиперплоскостью  $\nabla g(x^*)(x - x^*) = 0$ .



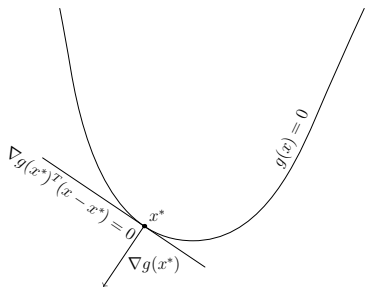
# Геометрическая интерпретация множителей Лагранжа



- Вблизи  $x^*$  поверхность  $g(x) = 0$  почти совпадает с гиперплоскостью  $\nabla g(x^*)(x - x^*) = 0$ .
- Если  $x^*$  – регулярная точка  $g$ , то существует единственное представление  $\nabla f(x^*) = \lambda^T \nabla g(x^*) + v^T$  при  $\nabla g(x^*)v = 0_m$ . Для точек  $x$  на гиперплоскости  $\nabla g(x^*)(x - x^*) = 0$  получаем

$$\nabla f(x^*)^T(x - x^*) = (-\lambda^T) \nabla g(x^*)(x - x^*) + v^T(x - x^*) = v^T(x - x^*).$$

# Геометрическая интерпретация множителей Лагранжа

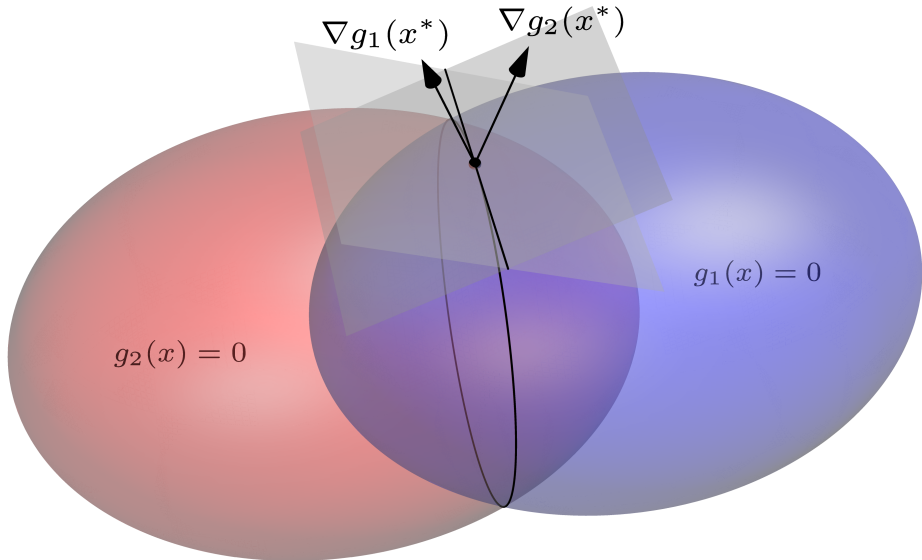


- Вблизи  $x^*$  поверхность  $g(x) = 0$  почти совпадает с гиперплоскостью  $\nabla g(x^*)(x - x^*) = 0$ .
- Если  $x^*$  – регулярная точка  $g$ , то существует единственное представление  $\nabla f(x^*) = \lambda^T \nabla g(x^*) + v^T$  при  $\nabla g(x^*)v = 0_m$ . Для точек  $x$  на гиперплоскости  $\nabla g(x^*)(x - x^*) = 0$  получаем

$$\nabla f(x^*)^T(x - x^*) = (-\lambda^T \nabla g(x^*)(x - x^*) + v^T(x - x^*)) = v^T(x - x^*).$$

- Если  $v^T(x - x^*) \neq 0$ , то  $v \neq 0_n$  и  $x^*$  не является точкой минимума. Если же  $v = 0_n$ , то выполняется условие Лагранжа с множителями  $-\lambda_1, \dots, -\lambda_m$ .

# Геометрическая интерпретация множителей Лагранжа



## Доказательство метода Лагранжа

**Док-во:** Обозначим за  $M$  гиперплоскость

$$M = \{y \in \mathbb{R}^n \mid \nabla g(x^*)^T y = 0_m\}.$$

Пусть  $y \in M$ ,  $y \neq 0_n$ , тогда для некоторого  $a > 0$  существует кривая  $\varphi : [-a, a] \rightarrow \mathbb{R}^n$ , такая что  $g(\varphi(t)) = 0$  при  $t \in [-a, a]$  и при этом

$$\varphi(0) = x^*, \quad \varphi'(0) = y.$$

Так как  $x^*$  – оптимальное значение вдоль кривой  $\varphi$ , то в силу условий стационарности

$$0 = \frac{d}{dt} f(\varphi(t))|_{t=0} = \nabla f(\varphi(t))^T \varphi'(t)|_{t=0} = \nabla f(x^*)^T y.$$

## Доказательство метода Лагранжа

**Док-во:** Обозначим за  $M$  гиперплоскость

$$M = \{y \in \mathbb{R}^n \mid \nabla g(x^*)^T y = 0_m\}.$$

Пусть  $y \in M$ ,  $y \neq 0_n$ , тогда для некоторого  $a > 0$  существует кривая  $\varphi : [-a, a] \rightarrow \mathbb{R}^n$ , такая что  $g(\varphi(t)) = 0$  при  $t \in [-a, a]$  и при этом

$$\varphi(0) = x^*, \quad \varphi'(0) = y.$$

Так как  $x^*$  – оптимальное значение вдоль кривой  $\varphi$ , то в силу условий стационарности

$$0 = \frac{d}{dt} f(\varphi(t))|_{t=0} = \nabla f(\varphi(t))^T \varphi'(t)|_{t=0} = \nabla f(x^*)^T y.$$

Таким образом, получаем, что вектор  $\nabla f(x^*)$  ортогонален  $M$ , с другой стороны из определения  $M$  следует, что векторы  $\nabla g_i(x^*)$  образуют базис ортогонального дополнения  $M$ . Следовательно,  $\nabla f(x^*)$  является линейной комбинацией  $\nabla g_i(x^*)$ , т.е. существуют коэффициенты  $\lambda_1, \dots, \lambda_m$ , такие что

$$\nabla f(x^*) = \sum_{i=1}^m \lambda_i \nabla g_i(x^*) \Leftrightarrow \nabla f(x^*) + (-\lambda^T) \nabla g(x^*) = 0_n. \quad \blacksquare$$

# Простой пример

Рассмотрим задачу

$$\begin{array}{l} \text{максимизировать} \\ \text{при условии} \end{array} \quad \begin{array}{l} x + y, \\ x^2 + y^2 = a^2. \end{array}$$

## Простой пример

Рассмотрим задачу

$$\begin{array}{ll} \text{максимизировать} & x + y, \\ \text{при условии} & x^2 + y^2 = a^2. \end{array}$$

Лагранжиан:

$$F(x, y, \lambda) = x + y - \lambda(x^2 + y^2 - a^2).$$

Получаем условия стационарности

$$1 - 2\lambda x = 0,$$

$$1 - 2\lambda y = 0.$$



## Простой пример

Рассмотрим задачу

$$\begin{array}{l} \text{максимизировать} \\ \text{при условии} \end{array} \quad \begin{array}{l} x + y, \\ x^2 + y^2 = a^2. \end{array}$$

Лагранжиан:

$$F(x, y, \lambda) = x + y - \lambda(x^2 + y^2 - a^2).$$

Получаем условия стационарности

$$1 - 2\lambda x = 0,$$

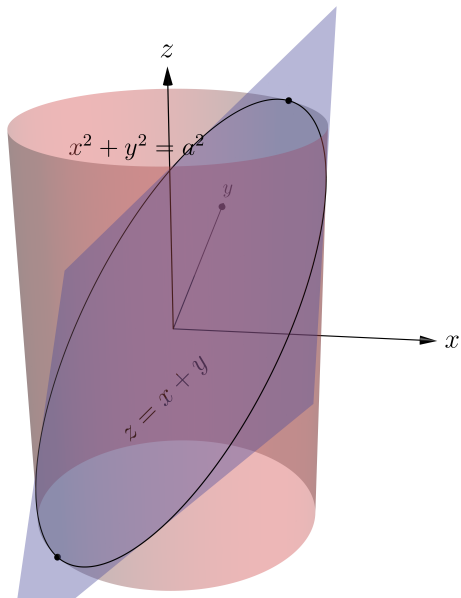
$$1 - 2\lambda y = 0.$$

Из этих уравнений следует, что  $\lambda \neq 0$ . Вычитая первое из второго получаем

$$2\lambda(x - y) = 0 \Rightarrow x = y.$$

Подставляя в исходное уравнение получаем  $x = \pm \frac{a}{\sqrt{2}}$ . При этом, если  $a = 0$ , то  $x = 0$  и условия стационарности не выполняются ни при каком  $\lambda$ .

# Простой пример



## Простой пример

Решение параметризацией: окружность  $x^2 + y^2 = a^2$  легко параметризуется:

$$x^2 + y^2 = a^2 \Leftrightarrow \exists t \in \mathbb{R} : (x, y) = (a \cos(t), a \sin(t)).$$

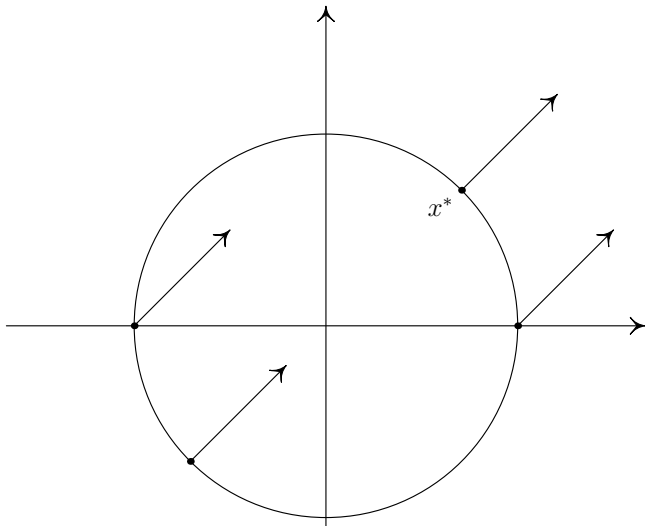
Применяя условия стационарности получаем

$$0 = \frac{d}{dt}[a \cos(t) + a \sin(t)] = a(\cos(t) - \sin(t)) \Rightarrow \cos(t) = \sin(t) = \pm \frac{1}{\sqrt{2}}.$$

Отсюда получаем, что  $x^* = y^* = \frac{a}{\sqrt{2}}$ .

## Простой пример

Решение методом множителей Лагранжа: найти такую точку на окружности, что касательная в этой точке перпендикулярна вектору  $\nabla f(x, y) \equiv (1, 1)$ .



## Пример: расстояние от точки до гиперплоскости

Пусть в  $\mathbb{R}^n$  задана гиперплоскость

$$a^T x = a_1 x_1 + a_2 x_2 + \dots + a_n x_n = c.$$

Требуется найти расстояние от этой гиперплоскости до некоторой точки  $y$ , что можно представить в виде следующей задачи оптимизации:

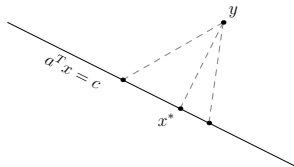
$$\begin{array}{ll} \text{минимизировать} & \|x - y\|^2 = \sum_{k=1}^n (x_k - y_k)^2, \\ \text{при условии} & a^T x = c. \end{array}$$

Функция Лагранжа имеет вид

$$L(\lambda, x) = \|x - y\|^2 - \lambda(a^T x - c).$$

Получаем необходимые условия

$$2(x - y) - \lambda a = 0.$$



## Пример: расстояние от точки до гиперплоскости

Разрешим относительно  $x$

$$x = \frac{\lambda a}{2} + y.$$

Подставляем в уравнение гиперплоскости

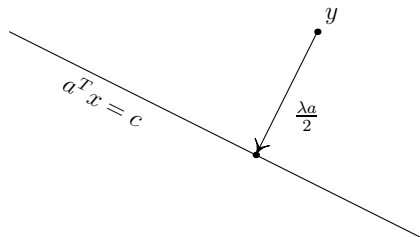
$$\frac{\lambda a^T a}{2} + a^T y = c, \quad \lambda = 2 \frac{c - a^T y}{a^T a},$$

$$x = y + a \frac{c - a^T y}{a^T a}.$$

И наконец, квадрат расстояния равен

$$\|x - y\|^2 = (a^T a) \frac{\lambda^2}{4} = (a^T a) \frac{(c - a^T y)^2}{(a^T a)^2} = \frac{(c - a^T y)^2}{(a^T a)}$$

## Пример: расстояние от точки до гиперплоскости



Значение  $\frac{\lambda}{2}$  можно интерпретировать как расстояние от  $y$  до  $a^T x = c$ , измеренная в векторе  $a$ .

## Пример: электрический ток

Оказывается, из первого закона Кирхгофа и одного из вариационных принципов можно вывести остальные законы Кирхгофа. Рассмотрим электрическую цепь из  $n$  узлов,  $r_{ij}$  – сопротивление участка цепи между узлами  $i$  и  $j$ ,  $B$  – матрица инцидентности соответствующего графа с произвольной ориентацией дуг. Положим, что мы пустили единичный заряд из  $s$  в  $t$ . Обозначим

$$\chi_i^{st} = \begin{cases} -1 & i = s, \\ 1 & i = t, \\ 0 & i \neq s, t \end{cases}$$

и пусть  $I_{ij}$  – сила тока на участке цепи  $ij$ ,  $I$  – соответствующий вектор, согласованный с нумерацией и ориентацией  $B$ , тогда первый закон Кирхгофа для этой цепи будет иметь вид

$$BI = \chi^{st}.$$



## Пример: электрический ток

Далее полагаем, что ток протекает минимизируя энергию

$$\frac{1}{2} \sum_{ij} r_{ij} I_{ij}^2$$

Посмотрим на это как на задачу оптимизации относительно  $I_{ij}$ . Лагранжиан имеет вид

$$F(I, \lambda) = \frac{1}{2} \sum_{ij} r_{ij} I_{ij}^2 - \lambda^T B I.$$

Дифференцируя по  $I_{ij}$  получаем

$$\frac{\partial}{\partial I_{ij}} F(I, \lambda) = r_{ij} I_{ij} - \lambda_j + \lambda_i = 0.$$

## Пример: электрический ток

Отсюда вытекает, что

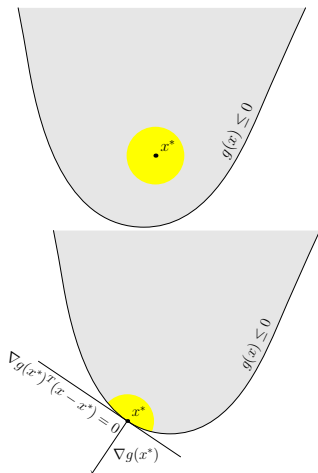
$$r_{ij}I_{ij} = \lambda_i - \lambda_j.$$

Другими словами, в этом случае множители Лагранжа являются величиной, которую в физике принято называть *потенциалом*, а последнее неравенство является ничем иным как закон Ома.

## Ограничения в виде неравенств

Рассмотрим функцию  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  и область, заданную неравенством  $g(x) \leq 0_m$  (здесь и далее  $x \leq y$  при  $x, y \in \mathbb{R}^m$  означает  $x_i \leq y_i, 1 \leq i \leq m$ ). Для обобщения множителей Лагранжа достаточно подметить два важных момента:

- Если  $g_i(x) < 0$  и  $g_i(x)$  непрерывна в точке  $x^*$ , то  $g_i(x) < 0$  в некоторой окрестности  $x^*$ , а значит это ограничение не влияет на локальные свойства задачи для точки  $x^*$ .
- Если  $g_i(x) = 0$ , то как и в случае ограничений в виде равенств,  $\nabla g_i(x^*)$  задает направление, вдоль которого сдвигаться нельзя. Отличие неравенств состоит в том, что нельзя сдвигаться только против градиента, но можно сдвигаться вдоль.



# Лемма о чувствительности

## Лемма

Рассмотрим задачу

$$\begin{aligned} &\text{минимизировать } f(x) \\ &h(x) = d, \end{aligned}$$

где  $d \in \mathbb{R}^k$ . Пусть  $x^*(d), \lambda(d)$  – решение задачи и соответствующие ему множители Лагранжа, удовлетворяющие

$$\nabla f(x^*(d)) + \lambda(d)^T \nabla h(x^*(d)) = 0_n,$$

при этом система

$$\begin{aligned} \nabla f(x) + \lambda^T \nabla h(x) &= 0_n \\ h(x) &= d \end{aligned}$$

регулярна относительно  $x$  в точке  $x^*(d)$ , тогда

$$\nabla_d f(x^*(d)) = -\lambda(d).$$

# Лемма о чувствительности

**Док-во.** Применяя условия стационарности получаем систему

$$\begin{aligned}\nabla f(x^*(d)) + \lambda(d)^T \nabla h(x^*(d)) &= 0_n \\ h(x^*(d)) &= d.\end{aligned}$$

## Лемма о чувствительности

**Док-во.** Применяя условия стационарности получаем систему

$$\begin{aligned}\nabla f(x^*(d)) + \lambda(d)^T \nabla h(x^*(d)) &= 0_n \\ h(x^*(d)) &= d.\end{aligned}$$

По теореме о неявном отображении  $x^*(d)$  – дифференцируема (вместе с  $\lambda(d)$ ), при этом выполняются следующие равенства

$$\begin{aligned}\nabla_d f(x^*(d)) &= \nabla_x f(x^*(d)) \nabla_d x^*(d), \\ \nabla_d h(x^*(d)) &= \nabla_x h(x^*(d)) \nabla_d x^*(d).\end{aligned}$$

## Лемма о чувствительности

**Док-во.** Применяя условия стационарности получаем систему

$$\begin{aligned}\nabla f(x^*(d)) + \lambda(d)^T \nabla h(x^*(d)) &= 0_n \\ h(x^*(d)) &= d.\end{aligned}$$

По теореме о неявном отображении  $x^*(d)$  – дифференцируема (вместе с  $\lambda(d)$ ), при этом выполняются следующие равенства

$$\begin{aligned}\nabla_d f(x^*(d)) &= \nabla_x f(x^*(d)) \nabla_d x^*(d), \\ \nabla_d h(x^*(d)) &= \nabla_x h(x^*(d)) \nabla_d x^*(d).\end{aligned}$$

С другой стороны, из условий задачи следует, что

$$\nabla_d h(x^*(d)) = I_k.$$

# Лемма о чувствительности

Последовательно применяя эти равенства получаем

$$\begin{aligned}\nabla_d f(x^*(d)) &= \nabla_x f(x^*(d)) \nabla_d x^*(d) \\ &= -(\lambda(d)^T \nabla_x h(x^*(d))) \nabla_d x^*(d) \\ &= -\lambda(d)^T \underbrace{\nabla_x h(x^*(d)) \nabla_d x^*(d)}_{=\nabla_d h(x^*(d))=l_k} \\ &= -\lambda(d)^T. \blacksquare\end{aligned}$$



# Условия Каруша-Куна-Такера

Пусть  $\mathcal{D} \subset \mathbb{R}^n$ ,  $\mathcal{D}$  открыто,  $f : \mathcal{D} \rightarrow \mathbb{R}$ ,  $g : \mathcal{D} \rightarrow \mathbb{R}^m$ ,  $h : \mathcal{D} \rightarrow \mathbb{R}^k$  – непрерывно дифференцируемые функции. Рассмотрим задачу

$$\begin{array}{ll} \text{минимизировать} & f(x), \\ \text{при условии} & g(x) \leq 0_m, \\ & h(x) = 0_k. \end{array}$$

При выполнении некоторых условий регулярности (например,  $\nabla g_i(x^*)$ ,  $\nabla h_i(x^*)$  линейно независимы), если  $x^*$  является решением задачи, тогда существует вектора  $\lambda \in \mathbb{R}^m$  и  $\mu \in \mathbb{R}^k$  такие что выполняются следующие условия:

1. Стационарность:  $\nabla f(x^*) + \lambda^T \nabla g(x^*) + \mu^T \nabla h(x^*) = 0_n$ .
2. Прямая выполнимость:  $g(x^*) \leq 0_m$ ,  $h(x^*) = 0_k$ .
3. Двойственная выполнимость:  $\lambda \geq 0_m$  ( $\leq 0_m$  для задачи максимизации).
4. Дополняющая нежесткость:  $\lambda_i g_i(x^*) = 0$ ,  $1 \leq i \leq m$ .

## Доказательство условий ККТ

Для начала введем дополнительные переменные  $y_1, \dots, y_m$ ,  $z = (y_1^2, \dots, y_m^2)^T$  и заметим, что исходная задача эквивалентна следующей

$$\begin{aligned} & \text{минимизировать} && f(x), \\ & \text{при условии} && g(x) + z = 0_m, \\ & && h(x) = 0_k. \end{aligned}$$

Функция Лагранжа имеет вид

$$F(x, y, \lambda, \mu) = f(x) + \lambda^T (g(x) + z) + \mu^T h(x).$$

Если  $x^*$  – решение задачи, то из теоремы Лагранжа получаем, что существуют векторы  $\lambda, \mu$  такие, что

$$\begin{aligned} \nabla_x F(x^*, y^*, \lambda, \mu) &= \nabla f(x^*) + \lambda^T \nabla g(x^*) + \mu^T \nabla h(x^*) = 0_n, \\ \frac{\partial}{\partial y_i} F(x^*, y^*, \lambda, \mu) &= 2\lambda_i y_i^* = 0, \end{aligned}$$

Заметим, что если  $\lambda_i = 0$ , то  $0 = \lambda_i g(x^*)_i$ . Если же  $\lambda_i \neq 0$ , то  $y_i^* = 0$ , а значит  $g_i(x) = -(y_i^*)^2 = 0$ . Другими словами, из второго равенства следует

$$\lambda_i g_i(x^*) = 0.$$

## Доказательство условий ККТ

Предположим, что вместо условия  $g(x) = 0_m$  имеет место условие  $g(x) = d$ .  
Из леммы о чувствительности имеем

$$-\lambda^T = \nabla_d f(x^*(d))|_{d=0}.$$

Очевидным образом, при увеличении любой компоненты вектора  $d$  в неравенстве  $g(x) \leq d$  ведет за собой увеличение множества, на котором это неравенство выполняется  $\Rightarrow \nabla_d f(x^*(d)) \leq 0 \Rightarrow \lambda \geq 0_m$  ■.

## Пример: water-filling

Рассмотрим задачу

$$\begin{array}{ll} \text{минимизировать} & -\sum_{i=1}^n \log(\alpha_i + x_i), \\ \text{при условии} & x \geq 0, \\ & \mathbf{1}_n^T x = 1, \end{array}$$

где  $\alpha_i > 0$ . Введем двойственные переменные  $\lambda \in \mathbb{R}^n$  для неравенства  $x \geq 0$  и  $\mu \in \mathbb{R}$  для равенства  $\mathbf{1}_n^T x = 1$ . Получаем условия стационарности

$$-\frac{1}{\alpha_i + x_i} - \lambda_i + \mu = 0.$$

Двойственная выполнимость

$$\lambda \geq 0.$$

Дополняющая нежесткость

$$\lambda_i x_i = 0.$$

## Пример: water-filling

От  $\lambda$  можно избавиться следующим образом:

$$-\frac{1}{\alpha_j + x_j} - \lambda_j + \mu = 0, \lambda \geq 0 \Leftrightarrow -\frac{1}{\alpha_j + x_j} + \mu \geq 0.$$

Дополняющая нежесткость можно переписать в виде

$$\left( \mu - \frac{1}{\alpha_j + x_j} \right) x_j = 0.$$

Если  $\mu < \frac{1}{\alpha_j}$ , то  $x_j > 0$ , так как при  $x_j = 0$

$$\mu \geq \frac{1}{\alpha_j + x_j} = \frac{1}{\alpha_j},$$

а значит из условий дополняющей нежесткости получаем, что  $\mu = 1/(\alpha_j + x_j)$ .

Если  $\mu \geq \frac{1}{\alpha_j}$ , то  $x_j = 0$ , так как при  $x_j > 0$

$$\mu \geq \frac{1}{\alpha_j} > \frac{1}{\alpha_j + x_j},$$

что нарушает условие дополняющей нежесткости.

## Пример: water-filling

Итого имеем

$$x_i = \begin{cases} 1/\mu - \alpha_i & \mu < 1/\alpha_i, \\ 0, & \mu \geq 1/\alpha_i. \end{cases}$$

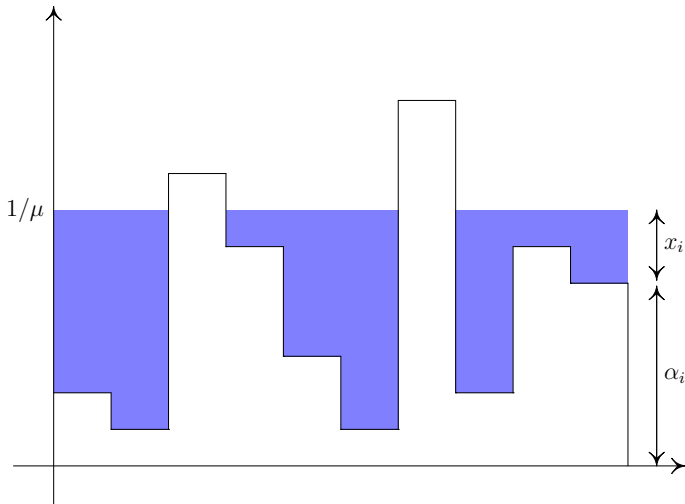
В более простой форме,  $x_i = \max\{0, 1/\mu - \alpha_i\}$ . Подставляя в исходное условие получаем

$$\mathcal{L}(1/\mu) = \sum_{i=1}^n \max\{0, 1/\mu - \alpha_i\} = 1.$$

Заметим, что  $\mathcal{L}$  – кусочно-линейная функция, при этом  $\mathcal{L}(-\infty) = 0$ ,  $\mathcal{L}(+\infty) = +\infty$ ,  $\mathcal{L}(t)$  строго возрастает при  $t > \min_i \alpha_i$  и  $\mathcal{L}(\min_i \alpha_i) = 0$ , а значит уравнение  $\mathcal{L}(t) = 1$  имеет единственное решение.

## Пример: water-filling

Нахождение решения  $\mathcal{L}(t) = 1$  зачастую называют *water-filling algorithm* ( $\mathcal{O}(n \log n)$  или  $\mathcal{O}(n)$ , если известен отсортированный порядок  $\alpha_i$ ) из-за того, что распределение  $x_i$  напоминают распределение воды в неоднородной среде



# Двойственная задача

Лагранжиан:

$$L(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x).$$

Двойственной функцией Лагранжа называется функция

$$q(\lambda, \mu) := \inf_{x \in \mathcal{D}} L(x, \lambda, \mu).$$

Двойственная задача:

$$\begin{array}{ll} \text{максимизировать} & q(\lambda, \mu), \\ \text{при условии} & \lambda \geq 0_m, \end{array}$$

Часто удобно непосредственно добавлять к этой задаче ограничение  $q(\lambda, \mu) > -\infty$ .



## Свойства двойственной задачи

- Функция  $q$  – вогнута (выпукла вверх):  $L(x, \lambda, \mu)$  линейна по  $\lambda$  и  $\mu$  для любого  $x$ , а значит, вогнута по  $\lambda$  и  $\mu$ . Легко проверить, что если  $F(x, y)$  вогнута по  $y$  для любого  $x$ , то  $\inf_x F(x, y)$  также вогнута. Как следствие, двойственная задача всегда имеет решение.
- Если условия  $g(x) = 0_m$ ,  $h(x) = 0_k$  выполнимы, то для любых допустимых  $x$  и  $\lambda \geq 0_m$ ,  $\mu$  имеет место неравенство

$$q(\lambda, \mu) \leq f(x).$$

Пусть  $x$  – допустимая точка, тогда

$$f(x) \geq f(x) + \underbrace{\lambda^T g(x)}_{\leq 0} + \underbrace{\mu^T h(x)}_{=0} \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \mu) = q(\lambda, \mu).$$

В частности, если обе задачи разрешимы, то для оптимальных значений  $x^*$ ,  $\lambda^*$ ,  $\mu^*$  выполняется неравенство

$$q(\lambda^*, \mu^*) \leq f(x^*).$$

Это свойство принято называть *слабой двойственностью*.

# Сильная двойственность

Очевидным образом, если для прямой и двойственных задач найдены такие значения  $x^*$ ,  $\lambda^*$ ,  $\mu^*$ , что

$$f(x^*) = q(\lambda^*, \mu^*),$$

то  $x^*$  – оптимальное решение прямой задачи, а  $\lambda^*$ ,  $\mu^*$  – решение двойственной задачи. Свойство, когда такая ситуация возможна, принято называть *сильной двойственностью*. К сожалению, сильная двойственность не всегда имеет место. Наиболее распространенное условие выполнения сильной двойственности – так называемое *условие Слейтера*: Если в задаче

$$\begin{array}{ll} \text{минимизировать} & f(x), \\ \text{при условии} & g(x) \leq 0_m, \\ & Ax = b \end{array}$$

с выпуклыми функциями  $f$ ,  $g$  существует такая точка  $\tilde{x}$ , что  $g(\tilde{x}) < 0_m$ ,  $A\tilde{x} = b$ , то для этой задачи и двойственной к ней выполняется сильная двойственность.