Условия Каруша-Куна-Такера в задачах оптимизации

Мальковский Н. В.

Санкт-Петербургский Академический Университет

Задачи оптимизации с ограничениями

Общая задача математической оптимизации – это

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

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

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

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:

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

Задачи оптимизации с ограничениями

Общая задача математической оптимизации – это

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

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:

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

Если g – дифференцируемая функция без нерегулярных точек, то $g(x) < 0$ задает открытой 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

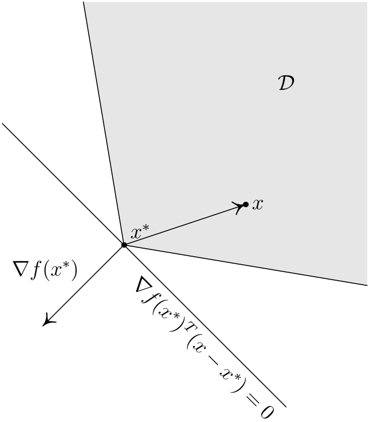
$$f(x^* + \alpha(x - x^*)) - f(x^*) = \alpha \nabla f(x^*)^T(x - x^*) + o(\alpha||x - x^*||) \leq$$

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

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, $f : \mathcal{D} \to \mathbb{R}$, $g : \mathcal{D} \to \mathbb{R}^m$ – непрерывно дифференцируемые функции. Рассмотрим задачу:

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

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

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

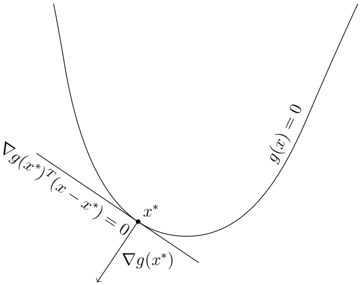
Коэффициенты λ_i принято называть *множителями Лагранжа*.

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

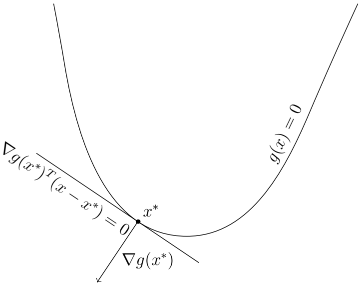
$$\nabla_x L(\lambda, x^*) = 0_n.$$

Геометрическая интерпретация множителей Лагранжа



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

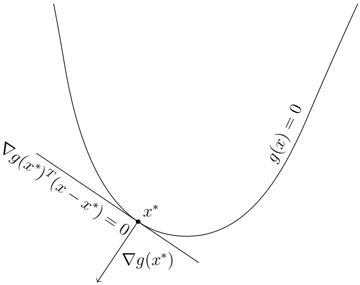
Геометрическая интерпретация множителей Лагранжа



- Вблизи x^* поверхность $g(x) = 0$ почти совпадает с гиперплоскостью $\nabla g_i(x^*)^T(x - x^*) = 0$.
- Если $g(x) = 0_m$. Существует единственное разложение градиента f в следующем виде: $\nabla f(x^*) = \lambda^T \nabla g(x^*) + v^T$, $\nabla g(x^*) v = 0_m$. Используя это разложение для допустимых точек x вблизи x^* получаем

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

Геометрическая интерпретация множителей Лагранжа



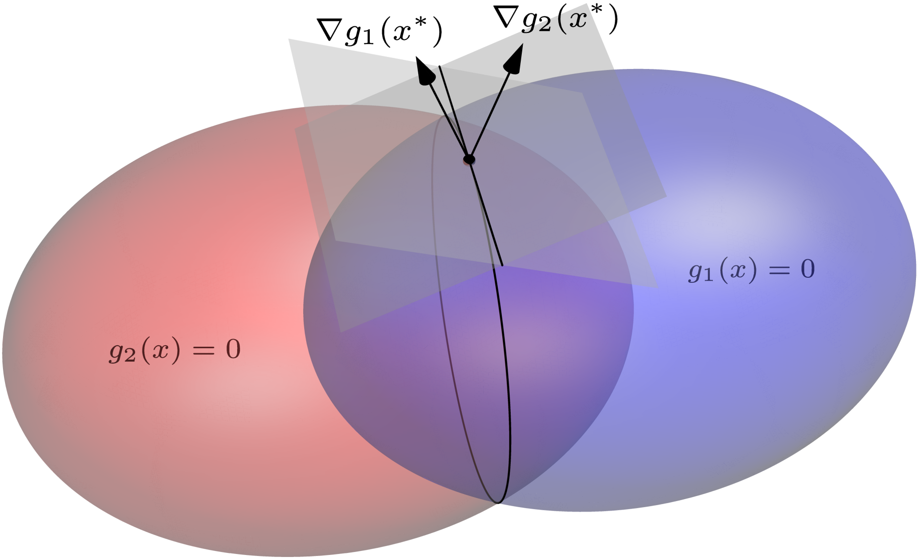
- Вблизи x^* поверхность $g(x) = 0$ почти совпадает с гиперплоскостью $\nabla g_i(x^*)^T(x - x^*) = 0$.
- Если $g(x) = 0_m$. Существует единственное разложение градиента f в следующем виде: $\nabla f(x^*) = \lambda^T \nabla g(x^*) + v^T$, $\nabla g(x^*)v = 0_m$. Используя это разложение для допустимых точек x вблизи x^* получаем

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

$$v^T(x - x^*) = v^T(x - x^*).$$

- 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 $-\lambda_1, \ldots, -\lambda_m$.

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 $\varphi : [-a, a] \to \mathbb{R}^n$, такая что $g(\varphi(t)) = 0$ при $t \in [-a, a]$ и при этом

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

Так как x^* – оптимальное значение вдоль кривой φ, 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 $\varphi : [-a, a] \to \mathbb{R}^n$, такая что $g(\varphi(t)) = 0$ при $t \in [-a, a]$ и при этом

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

Так как x^* – оптимальное значение вдоль кривой φ, 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 $\lambda_1, \ldots, \lambda_m$, такие что

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

Простой пример

Рассмотрим задачу

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

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

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

Лагранжиан:

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

Получаем условия стационарности

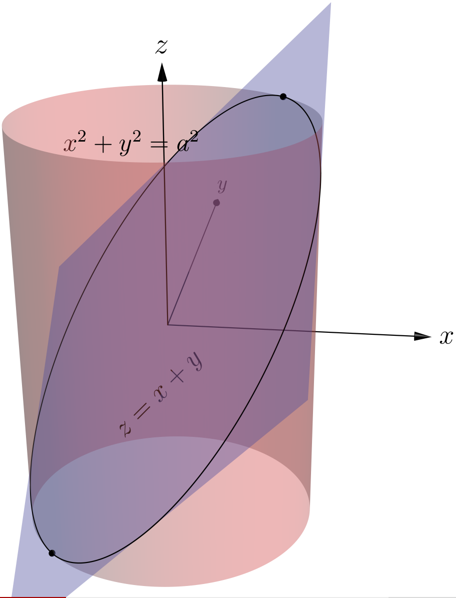
$$1 - 2\lambda x = 0,$$
$$1 - 2\lambda y = 0.$$

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 λ.

Простой пример



Простой пример

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

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

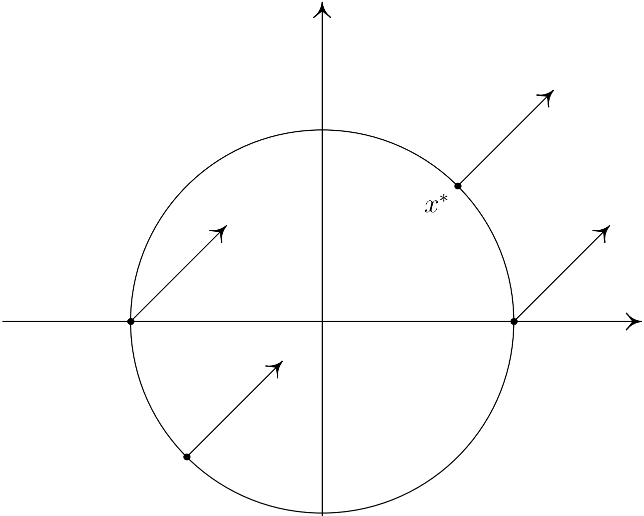
Применяя условия стационарности получаем

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

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

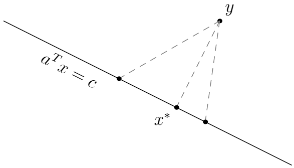
$$a^T x = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = c.$$

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

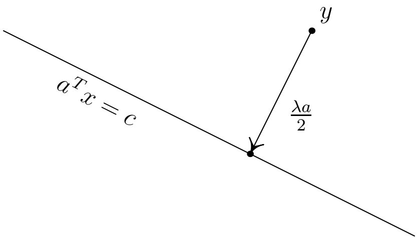
$$\frac{\lambda a^T a}{2} + a^T y = c, \quad \lambda = 2\frac{c - a^T y}{a^T a},$$

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

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.

Упражнение. Вывести формулы расстояний до гиперплоскости, заданными несколькими линейными уравнениями.

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

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

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

- Метод Лагранжа позволяет позволяет эффективно решать задачи с ограничением в виде равенств.

Ограничения в виде неравенств

- Метод Лагранжа позволяет позволяет эффективно решать задачи с ограничением в виде равенств.
- Во многих задачах оптимизации возникают ограничения в виде неравенств, свести которые к равенством с сохранением “хороших” свойств (дифференцируемость, выпуклость и т.д.) не представляется возможным.

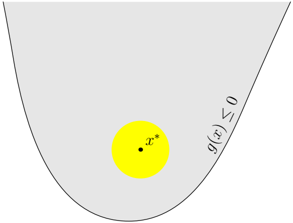
Ограничения в виде неравенств

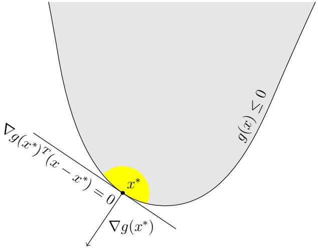
- Метод Лагранжа позволяет позволяет эффективно решать задачи с ограничением в виде равенств.
- Во многих задачах оптимизации возникают ограничения в виде неравенств, свести которые к равенством с сохранением “хороших” свойств (дифференцируемость, выпуклость и т.д.) не представляется возможным.
- Как оказалось, метод множителей Лагранжа можно обобщить и на случай с неравенствами. Такое обобщение было независимо предложено Вильямом Карушем (1939) и Харольдом Куном, Альбертом Такером (1951), и на сегодняшний день принято называть условиями Каруша-Куна-Такера.

Ограничения в виде неравенств

Рассмотрим функцию $g : \mathbb{R}^n \to \mathbb{R}^m$ и область, заданную неравенством $g(x) \le 0_m$ (здесь и далее $x \le y$ при $x, y \in \mathbb{R}^m$ означает $x_i \le y_i$, $1 \le i \le m$). 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

$$\begin{array}{ll} \text{минимизировать} & f(x), \\ \text{при условии} & g(x) \le c, \\ & h(x) = d, \end{array}$$

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

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

Градиенты $\nabla g_i(x^*(c,d))$, $\nabla h_j(x^*(c,d))$ *при* $i : g_i(x^*(c,d)) = c_i$ *линейно независимы. Тогда*

$$\nabla_c f(x^*(c,d)) = -\lambda(c,d),$$

$$\nabla_d f(x^*(c,d)) = -\mu(c,d).$$

Лемма о чувствительности

Док-во. Очевидным образом, если $g_i(x^*(c,d)) < c_i$, то

$$\frac{\partial}{\partial c_i} f(x^*(c,d)) = 0.$$

Далее считаем для простоты, что $g_i(x^*(c,d)) = c_i$ для всех $i \in \{1, \dots, m\}$. Применяя условия стационарности получаем систему

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

Лемма о чувствительности

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

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

С другой стороны, из условий задачи следует, что

$$
\begin{aligned}
\nabla_c g(x^*(c, d)) &= I_m,\\
\nabla_d g(x^*(c, d)) &= 0,\\
\nabla_c h(x^*(c, d)) &= 0,\\
\nabla_d h(x^*(c, d)) &= I_k.
\end{aligned}
$$

Лемма о чувствительности

Последовательно применяя эти равенства получаем

$$\nabla_c f(x^*(c,d)) = \nabla_x f(x^*(c,d))\nabla_c x^*(c,d) =$$

$$-(\lambda(c,d)^T\nabla_x g(x^*(c,d)) + \mu(c,d)^T\nabla_x h(x^*(c,d)))\nabla_c x^*(c,d) =$$

$$-\lambda(c,d)^T \underbrace{\nabla_x g(x^*(c,d))\nabla_c x^*(c,d)}_{=\nabla_c g(x^*(c,d))=I_m} - \mu(c,d)^T \underbrace{\nabla_x h(x^*(c,d))\nabla_c x^*(c,d)}_{=\nabla_c h(x^*(c,d))=0} =$$

$$-\lambda(c,d)^T$$

и

$$\nabla_d f(x^*(c,d)) = \nabla_x f(x^*(c,d))\nabla_d x^*(c,d) =$$

$$-(\lambda(c,d)^T\nabla_x g(x^*(c,d)) + \mu(c,d)^T\nabla_x h(x^*(c,d)))\nabla_d x^*(c,d) =$$

$$-\lambda(c,d)^T \underbrace{\nabla_x g(x^*(c,d))\nabla_d x^*(c,d)}_{=\nabla_d g(x^*(c,d))=0} - \mu(c,d)^T \underbrace{\nabla_x h(x^*(c,d))\nabla_d x^*(c,d)}_{=\nabla_d h(x^*(c,d))=I_k} =$$

$$-\mu(c,d)^T. \quad \blacksquare$$

Условия Каруша-Куна-Такера

Пусть $\mathcal{D} \subset \mathbb{R}^n$, $\mathcal{D}$ открыто, $f : \mathcal{D} \to \mathbb{R}$, $g : \mathcal{D} \to \mathbb{R}^m$, $h : \mathcal{D} \to \mathbb{R}^k$ – 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: $\nabla f(x^*) + \lambda^T \nabla g(x^*) + \mu \nabla h(x^*) = 0_n$.
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

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

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 λ, μ такие, что

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

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

Из леммы о чувствительности имеем

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

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

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 λ можно избавиться следующим образом:

$$-\frac{1}{\alpha_i + x_i} - \lambda_i + \mu = 0, \;\; \lambda \geq 0 \Leftrightarrow -\frac{1}{\alpha_i + x_i} + \mu \geq 0.$$

Дополняющая нежесткость можно переписать в виде

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

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

$$\mu \geq \frac{1}{\alpha_i + x_i} = \frac{1}{\alpha_i},$$

а значит из условий дополняющей нежесткости получаем, что $\mu = 1/(\alpha_i + x_i)$.
Если $\mu \geq \frac{1}{\alpha_i}$, то $x_i = 0$, так как при $x_i > 0$

$$\mu \geq \frac{1}{\alpha_i} > \frac{1}{\alpha_i + x_i},$$

что нарушает условие дополняющей нежесткости.

Пример: water-filling

Итого имеем

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

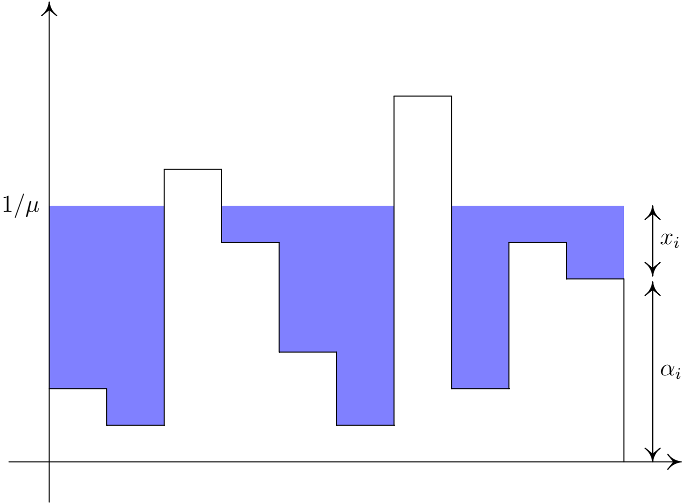
В более простой форме, $x_i = \max\{0, 1/\mu - \alpha_i\}$. Подставляя в исходное условие получаем

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

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 α_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 λ и μ для любого x, а значит, вогнута по λ и μ. 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 $\lambda \geq 0_m$, μ 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^*, λ^*, μ^* 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^*, λ^*, μ^*, что

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

то x^* – оптимальное решение прямой задачи, а λ^*, μ^* – решение двойственной задачи. Свойство, когда такая ситуация возможна, принято называть *сильной двойственностью*. К сожалению, сильная двойственность не всегда имеет место. Наиболее распространенное условие выполнение 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.