Выпуклое программирование

Рассмотрим задачу математического программирования следующего вида:

$$(*)\ \begin{cases} \min \varphi(x) \\ x \in X, \text{ где допустимое множество } X: \\ X: \left\{ \begin{array}{l} x \in R^n : f_i(x) \le 0,\ i = 1,\ldots,m \\ \qquad\quad g_k(x) = 0, k = 1,\ldots,l \end{array} \right\} \end{cases}$$

Определение. Если в задаче (*) целевая функция $\varphi(x)$ – выпуклая и допустимое множество X – выпукло, то задача (*) называется задачей выпуклого программирования.

Рассмотрим теперь задачу математического программирования следующего вида:

$$(**)\ \begin{cases} \min \varphi(x) \\ x \in X, \text{ где допустимое множество } X: \\ X: \left\{ x \in R^n : f_1(x) \le 0, \ldots, f_m(x) \le 0 \right\}, \text{ где функции } f_i(x) \text{ – выпуклые на } R^n,\ i = 1,\ldots,m, \\ \text{и целевая функция } \varphi(x) \text{ – выпуклая на } R^n. \end{cases}$$

Покажем, что допустимое множество X задачи (**) – выпукло.

Действительно, пусть $x_1, x_2 \in X,\ \alpha \in (0,1)$

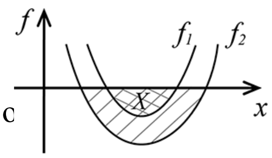
$\Rightarrow$ покажем, что $x = \alpha x_1 + (1-\alpha)x_2 \in X$, т.е. $\forall i \quad f_i(x) \le 0$.

Имеем, для $\forall i \quad f_i(x) = f_i(\alpha x_1 + (1-\alpha)x_2) \le \alpha f_i(x_1) + (1-\alpha) f_i(x_2) \le 0$

$\Rightarrow$ т.о. $\forall i\ f_i(x) \le 0 \Rightarrow$ точка $x \in X$, т.е. X – выпукло.

"Надграфик" выпуклой функции, т.е. множество

$$\Gamma \overset{def}{=} \left\{ \begin{pmatrix} a \\ x \end{pmatrix} : a \ge f(x),\ x \in D(f) \right\} \text{ – выпуклое множество}$$



Определение. Задача (**) называется основной задачей выпуклого программирования (ОЗВП).

Свойства выпуклых функций

1. Неравенство Йенсена

Пусть $f(x)$ – выпуклая функция на выпуклом множестве X. Тогда

$$f\left(\sum_{i=1}^{m} \lambda_i x^i\right) \le \sum_{i=1}^{m} \lambda_i f(x^i) \quad \text{при всех} \quad m = 1, 2, \ldots;\ x^i \in X;\ \lambda_i \ge 0;\ i = 1, \ldots, m;\ \sum_{i=1}^{m} \lambda_i = 1;$$

Доказательство. Индукция по m. Пусть $m = 1 \Rightarrow$ очевидно.

Пусть для $m = k$ утверждение верно.

Докажем для $m = k + 1$.

Пусть $x=\sum_{i=1}^{k+1}\lambda_i x^i,\ x^i\in X;\ \ \lambda_i\ge 0;\ i=1,\dots,k+1,\ \ \sum_{i=1}^{k+1}\lambda_i=1.$

Если $\lambda_{k+1}=1$, то $\lambda_1=\dots=\lambda_k=0\Rightarrow f(\lambda_{k+1}x^{k+1})=\lambda_{k+1}f(x^{k+1})$.

Если $\lambda_{k+1}>1\ \Rightarrow$ представим x в следующем виде:

$$x=(1-\lambda_{k+1})\overline{x}+\lambda_{k+1}x^{k+1},\ \text{где}\ \overline{x}=\sum_{i=1}^{k}\frac{\lambda_i}{1-\lambda_{k+1}}x^i\ .$$

Тогда:

$$f(x)\le(1-\lambda_{k+1})f(\overline{x})+\lambda_{k+1}f(x^{k+1})\le\left(1-\lambda_{k+1}\right)\sum_{i=1}^{k}\frac{\lambda_i}{1-\lambda_{k+1}}f(x^i)+\lambda_{k+1}f(x^{k+1})=\sum_{i=1}^{k+1}\lambda_i f(x^i)\text{, ч.т.д.}$$

2. Пусть $f(x)$ – выпуклая на выпуклом множестве $X\subset R^n$ функция. Тогда любой её локальный минимум на множестве X является одновременно и глобальным (доказательство было ранее в лекциях).

3. Пусть $f(x)$ – выпуклая функция на выпуклом множестве $X\subset R^n$, дифференцируемая в точке $x^*\in X$. Тогда $$f(x)-f(x^*)\ge(f'(x^*),x-x^*)\ \ \forall x\in X\ .$$

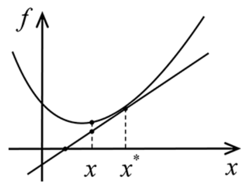


График f лежим не ниже касательной гиперплоскости к графику функции f в точке $(x^*, f(x^*))$.

(*Напоминание: график линейной функции* $f(x)=f(x^*)+(f'(x^*),x-x^*)$ *называется касательной гиперплоскостью к графику функции* f *в точке* $(x^*, f(x^*))$).

Доказательство. По определению выпуклой функции для $\forall x, x^*\in X, \lambda\in[0,1]$ имеем:

$$f\left(\lambda x+(1-\lambda)x^*\right)\le\lambda f(x)+(1-\lambda)f(x^*)\ .$$

Преобразуя эту формулу, имеем:

$$f(x)-f(x^*)\ge\frac{f(x^*+\lambda(x-x^*))-f(x^*)}{\lambda}=\frac{(f'(x^*),\lambda(x-x^*))+o(\lambda)}{\lambda}=(f'(x^*),x-x^*)+\frac{o(\lambda)}{\lambda};$$

Переходя к пределу при $\lambda\to 0$, имеем искомое соотношение.

4. Пусть f – дважды непрерывно дифференцируемая функция на выпуклом множестве $X\subset R^n$. Тогда f выпукла на X $\Leftrightarrow$ матрица Гессе f'' неотрицательно определена, т.е. $\forall x^*\in X, \forall h^*\in R^n\ \ (f''(x^*)h,h)\ge 0$.

Был ранее без доказательства критерий сильной выпуклой функции f с параметром $\varnothing\ge 0$:
$(f''(x^*)h,h)\ge 2\varnothing\|h\|^2$

Доказательство.

Необходимость. Пусть f – выпукла на X.

а) Сначала предположим, что $x\in\operatorname{int}X$ (x^* – внутренняя точка множества X, т.е. существует ε-окрестность точки x^*, все точки которой принадлежат X). Тогда для $\forall h^*\in R^n$ имеем $x^*+\alpha h\in X$ при всех достаточно малых $\alpha>0$.Поскольку f – дважды дифференцируема в x^*, то можно записать:

$$f(x^*+\alpha h)=f(x^*)+(f'(x^*),\alpha h)+\frac{1}{2}(f''(x^*)\alpha h,\alpha h)+o(\alpha^2)\Rightarrow$$

$$\frac{\alpha^2}{2}(f''(x^*)h,h)+o(\alpha^2)=f(x^*+\alpha h)-f(x^*)-(f'(x^*),\alpha h)\ge 0 \Rightarrow$$

$$\frac{1}{2}(f''(x^*)h,h)+\frac{o(\alpha^2)}{\alpha^2}\ge 0.$$

Переходя к пределу при $\alpha \to 0$, имеем требуемое соотношение.

б) Пусть теперь $x^* \in X$ – произвольная точка $\Rightarrow$ существует последовательность точек $x^k \in \operatorname{int} X$ $(k=1,2,\ldots)$, сходящаяся к x^*. По доказанному выше, для $\forall h \in R^n$ имеем:

$$(f''(x^k)h,h)\ge 0 \ (k=1,2,\ldots),$$

причем последовательность матриц $f''(x^k)$ $(k=1,2,\ldots)$ сходится к $f''(x^*)$ в силу непрерывности $f''(x)$ в x^* (*непрерывность всех вторых частных производных*) $\Rightarrow$ имеем требуемое соотношение.

Достаточность: Пусть справедливо $(f''(x^*),h,h)\ge 0$. Тогда, рассмотрим произвольные точки $x, x^* \in X$ и положим $h=x-x^*$. Используя формулу Тейлора с остаточным членом в форме Лагранжа, имеем для некоторого $\alpha \in (0,1)$

$$f(x^*+h)-f(x^*)-(f'(x^*),h)=\frac{1}{2}(f''(x^*+\alpha h)h,h)\ge 0 \text{ (по предположению)}$$

Итак, имеем:

$$f(x^*+h)-f(x^*)-(f'(x^*),h)\ge 0 \text{ для } \forall x^* \in X, h \in R^n.$$

Надо показать, что для $\forall x', x'' \in X, \lambda \in (0,1)$ $f\left(\lambda x'+(1-\lambda)x''\right)\le \lambda f(x')+(1-\lambda)f(x'')$,

т.е. тем самым показать, что функция f – выпуклая.

Для этого зафиксируем произвольные $x', x'' \in X, \lambda \in (0,1)$ и рассмотрим точку x^*:

$x^*=\lambda x'+(1-\lambda)x'' \in X$ (*в силу выпуклости X*).

Тогда:

$f(x')-f(x^*)\ge\left(f'(x^*),x-x^*\right)$ *домножим на* λ

$f(x'')-f(x^*)\ge\left(f'(x^*),x''-x^*\right)$ *домножим на* $(1-\lambda)$

и сложим их: $\Rightarrow \lambda f(x')+(1-\lambda)f(x'')-f(x^*)\ge\left(f'(x^*),\ \underbrace{\lambda x'+(1-\lambda)x''}_{\overset{def}{=}x^*}-x^*\right)=0$

$$\Rightarrow f\left(\lambda x'+(1-\lambda)x''\right)\le \lambda f(x')+(1-\lambda)f(x''),$$

т.е. f – выпуклая, *ч.т.д.*

Пример. Пусть $f(x)=(Ax,x)+(b,x)$ – квадратичная функция, A – симметричная матрица. Тогда f – выпуклая $\Leftrightarrow$ A – неотрицательно определена.

Вообще, можно привести *критерии строгой и сильной выпуклости функций* аналогично тем, которые сейчас были доказаны для выпуклой функции (доказательство – аналогичное).

- Итак, для дифференцируемой функции $f(x)$:
  - *строгая выпуклость* эквивалентна неравенству
$$f(x) > f(x^*) + \left(f'(x^*), x - x^* \right),$$
  - *сильная выпуклость* эквивалентна неравенству
$$f(x) \geq f(x^*) + \left(f'(x^*), x - x^* \right) + \theta \| x - x^* \|^2 .$$

  Графически:
  выпуклость – возможно касание касательной плоскости;
  строгая выпуклость – единственная точка касания с касательной плоскостью;
  сильная выпуклость – график расположен внутри некоторого параболоида вращения $(Z = f(x^*) + \theta \| x - x^* \|)$.
- Для дважды дифференцируемых функций $f(x)$:
  - достаточным условием *строгой выпуклости* $f(x)$ является *положительная определенность* при $\forall x \in X$ её матрицы Гессе $f''(x)$;
  - достаточным условием *сильной выпуклости* $f(x)$ является *положительная определенность* матрицы $f''(x) - lE$, где E – единичная матрица, а $l > 0$.

Эти критерии в сочетании с критерием Сильвестра дают удобный аппарат для проверки выпуклости функций небольшого числа переменных.

Функция Лагранжа

Рассмотрим основную задачу выпуклого программирования:
$$\begin{cases} \min \varphi(x) \\ x \in X \\ X = \left\{ x \in R^n : f_1(x) \leq 0, \ldots, f_m(x) \leq 0 \right\} \end{cases}$$
$\varphi(x)$, $f_i(x)$ – выпуклые функции.

В основной задаче выпуклого программирования имеем m условий, определяющих допустимое множество X.

Рассмотрим вектор $\lambda = \begin{pmatrix} \lambda_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \lambda_m \end{pmatrix} \in R^m, \quad f(x) = \begin{pmatrix} f_1(x) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ f_m(x) \end{pmatrix} \in R^m .$

Определение. Функция

$$L(x,\lambda)=\varphi(x)+(\lambda,\ f(x))=\varphi(x)+\sum_{i=1}^{m}\lambda_i f_i(x),$$

называется *функцией Лагранжа* для основной задачи выпуклого программирования, где $x\in R^n,\lambda\geq 0$.

Определение. Пара $(x^*,\ \lambda^*)$ называется *седловой точкой* функции Лагранжа на множестве $R^n\times\{\lambda\geq 0\}$, если $\forall x\in R^n$ и $\forall\lambda\geq 0$

$$L(x^*,\lambda)\leq L(x^*,\lambda^*)\leq L(x,\lambda^*),$$

т.е.

$$L(x^*,\lambda^*)=\min_{x\in R^n}\ \max_{\lambda\geq 0}L(x,\lambda)=\max_{\lambda\geq 0}\ \min_{x\in R^n}L(x,\lambda)$$

Наличие седловой точки означает, что операции минимизации и максимизации можно переставлять местами.

В задачах *классического* анализа об условном экстремуме (задачи, в которых допустимое множество задается *системой уравнений*) важную роль играет метод множителей Лагранжа: решение исходной задачи ищется среди стационарных точек функции $L(x,\ \lambda)$ – точек, удовлетворяющих системе уравнений

$$\frac{\partial L}{\partial x}=0,\quad \frac{\partial L}{\partial \lambda}=0.$$

В задачах выпуклого (и, в частности, линейного) программирования функции Лагранжа также отводится важное место: при весьма общих предположениях задача выпуклого программирования сводится к отысканию седловых точек функции Лагранжа.

Теорема *о седловой точке функции Лагранжа* (достаточные условия оптимальности). Если пара (x^*,λ^*) является седловой точкой функции Лагранжа $L(x,\lambda)=\varphi(x)+(\lambda,\ f(x))$ на множестве $x\in R^n$, $\lambda\geq 0$, то x^* – оптимальная точка основной ЗВП.

Доказательство. По определению седловой точки имеем:

$$\varphi(x^*)+\sum_{i=1}^{m}\lambda_i f_i(x^*)\leq\varphi(x^*)+\sum_{i=1}^{m}\lambda_i^* f_i(x^*)\leq\varphi(x)+\sum_{i=1}^{m}\lambda_i^* f_i(x) \qquad (*)$$

а) Из левого неравенства (*) убираем $\varphi(x^*)$, и получаем:

$$\sum_{i=1}^{m}\lambda_i f_i(x^*)\leq\sum_{i=1}^{m}\lambda_i^* f_i(x^*) \quad\Rightarrow f_i(x^*)\leq 0,\ \text{для}\ \forall i,\quad \text{т.е. } x^* \text{ – допустимая точка.}$$

Действительно, если бы существовал индекс $i: f_i(x^*)>0$, то слева имели бы неограниченную сумму (т.к. $\forall\lambda_i\geq 0$), а справа имеем ограничение $\Rightarrow$ для $\forall i\ \ f_i(x^*)\leq 0$.

б) В частности, левое неравенство (*) верно и для $\lambda = 0$, тогда имеем $(\lambda^*, f(x^*)) \geq 0$, но

$$\lambda^* \geq 0,\ f(x^*) \leq 0 \Rightarrow \boxed{(\lambda^*, f(x^*)) = 0} \qquad (**)$$

в) Подставим (**) в правое неравенство (*):

$$\varphi(x^*) \leq \varphi(x) + (\lambda^*, f(x)) \leq \varphi(x)$$

Поскольку, для $\forall x \in X \quad f(x) \leq 0 \qquad \Rightarrow (\lambda^*, f(x)) \leq 0$

Итак, получили, что для $\forall x \in X \quad \varphi(x^*) \leq \varphi(x)$, т.е. x^* – оптимальная точка, *ч.т.д.*

Отметим, что при доказательстве теоремы нигде не использовались ни свойства выпуклости функций $\varphi(x)$, $f_i(x)$, ни свойства выпуклости множества R^n, ни какие-либо свойства гладкости.

Т.о., наличие седловой точки (x^*, λ^*) функции Лагранжа определяет оптимальность точки x^* для общей задачи математического программирования. Обратное утверждение, что из оптимальности точки x^* следует существование седловой точки (x^*, λ^*) функции Лагранжа, справедливо лишь для задачи выпуклого программирования при выполнении определенных ограничений относительно допустимого множества X.

Сформулируем эти ограничения и саму теорему, известную как *теорема Куна-Таккера*.

Определение 1. Рассмотрим допустимое множество

$$X = \left\{ x \in R^n : f_i(x) \leq 0,\ i = 1,...,m \right\}.$$

Если для всех $i \in 1,\ldots,m$ существует такая точка $x_i \in X$, что $f_i(x_i) < 0$, то говорят, что допустимое множество X удовлетворяет *условию регулярности*.

Определение 2. Пусть существует такая точка $x \in X$, что для всех $i \in 1,\ldots,m$ выполняется $f_i(x) < 0$. Тогда говорят, что допустимое множество удовлетворяет *условию регулярности Слейтера*.

Определения (1) и (2) – эквивалентны.

Действительно, из (2) $\Rightarrow$ (1) – очевидно ($x_i \equiv x$).

Пусть теперь выполнено (1).

Выберем $\quad x = \sum_{i=1}^{m} \alpha_i x_i, \quad \sum_{i=1}^{m} \alpha_i = 1, \quad \alpha_i \geq 0, \quad i = 1,...,m,$

Неравенство Йенсена для выпуклых функций

тогда для $\forall i \in 1,\ldots,m$ имеем: $\quad f_i(x) = f_i\left(\sum_{i=1}^{m} \alpha_i x_i \right) \leq \sum \alpha_i f_i(x_i) < 0$, *ч.т.д.*

Условие (2) означает, что существует точка внутри допустимого множества.

Теорема (Куна-Таккера (необходимые и достаточные условия оптимальности)). Пусть в основной задаче выпуклого программирования допустимое множество X обладает свойством регулярности. Тогда необходимым и достаточным условием оптимальности точки x^* является существование такого $\lambda^* \geq 0$, чтобы пара (x^*, λ^*) была седловой точкой для функции Лагранжа на множестве $x \in R^n$, $\lambda \geq 0$.

Доказательство.

Достаточность доказана в теореме о седловой точке функции Лагранжа.

Необходимость.

Пусть x^* – оптимальная точка. Рассмотрим два множества в пространстве R^{m+1}:

– множество $P \overset{def}{=} \left\{ \begin{pmatrix} z_0 \\ z_1 \\ . \\ . \\ . \\ . \\ z_m \end{pmatrix} : \begin{matrix} z_0 \leq \varphi(x^*) \\ z_i \leq 0 \\ i = 1, \dots, m \end{matrix} \right\}$;

– и множество $S \overset{def}{=} \bigcup\limits_{x \in R^n} S_x$, где $\forall x \;\; S_x \overset{def}{=} \left\{ \begin{pmatrix} \omega_0 \\ \omega_1 \\ \vdots \\ \omega_m \end{pmatrix} : \begin{matrix} \omega_0 \geq \varphi(x) \\ \omega_i \geq f_i(x) \\ i = 1, \dots, m \end{matrix} \right\}$.

Множество P – выпукло.

Действительно, пусть $z', z'' \in P \Rightarrow$ рассмотрим $z = \alpha z' + (1 - \alpha) z''$ $\forall \alpha \in [0,1]$ и покажем что $z \in P$.

Положим $z' \overset{def}{=} \begin{pmatrix} z_0' \\ z_i' \end{pmatrix}, z'' \overset{def}{=} \begin{pmatrix} z_0'' \\ z_i'' \end{pmatrix}, z \overset{def}{=} \begin{pmatrix} z_0 \\ z_i \end{pmatrix}$

$$\left. \begin{matrix} z_0 = \alpha z_0' + (1-\alpha) z_0'' \leq \alpha \varphi(x^*) + (1-\alpha)\varphi(x^*) \leq \varphi(x^*) \\ z_i = \alpha z_i' + (1-\alpha) z_i'' \leq 0 \end{matrix} \right\} z \in P$$

Множество S – выпукло.

Действительно, пусть $\begin{pmatrix} \omega_0' \\ \omega' \end{pmatrix}, \begin{pmatrix} \omega_0'' \\ \omega'' \end{pmatrix} \in S \Rightarrow$

Рассмотрим

$$\begin{pmatrix} \omega_0 \\ \omega \end{pmatrix} = \alpha \begin{pmatrix} \omega_0' \\ \omega' \end{pmatrix} + (1-\alpha) \begin{pmatrix} \omega_0'' \\ \omega'' \end{pmatrix}, \; \forall \alpha \in [0,1]$$

По определению множества S:

$$\left.\begin{array}{l}\exists\, x' : \begin{pmatrix}\omega_0' \\ \omega'\end{pmatrix} \in S_x' \\[2ex] \exists\, x'' : \begin{pmatrix}\omega_0'' \\ \omega''\end{pmatrix} \in S_x''\end{array}\right\} \Rightarrow \text{рассмотрим } x = \alpha x' + (1-\alpha)x'' \text{ и покажем, что } \begin{pmatrix}\omega_0 \\ \omega\end{pmatrix} \in S_x .$$

Т.к. φ – выпуклая функция, то $\varphi(x) \le \alpha\varphi(x') + (1-\alpha)\varphi(x'') \le \alpha\omega_0' + (1-\alpha)\omega_0'' = \omega_0$

Т.к. $f_i(x)$ – выпуклая функция, то $f_i(x) \le \alpha f_i(x') + (1-\alpha) f_i(x'') \le \alpha\omega_i' + (1-\alpha)\omega_i'' = \omega_i \;\; \forall i$

$$\Rightarrow \begin{pmatrix}\omega_0 \\ \omega\end{pmatrix} \in S_x \subset S \;\Rightarrow S \text{ – выпукло.}$$

Рассмотрим $P_0 \overset{def}{=} \left\{ \begin{pmatrix} z_0 \\ z_i \end{pmatrix} : \begin{matrix} z_0 < \varphi(x^*) \\ z_i < 0 \end{matrix} \right\}$ – множество внутренних точек P и покажем, что пересечение $P_0 \cap S = \varnothing$.

- Для $\forall x \in X$ $\underbrace{\omega_0 \ge \varphi(x) \ge \varphi(x^*)}_{(\cdot)\text{ из множества } S}$ (оптимальность), но $\underbrace{z_0 < \varphi(x^*)}_{(\cdot)\text{ из мн-ва } P_0}$.
- Для $\forall x \notin X$ $\exists i :$ $\underbrace{\omega_i \ge f_i(x) > 0}_{(\cdot)\text{ из мн-ва} S}$, но $\underbrace{z_i < 0}_{(\cdot)\text{ из мн-ва } P_0}$.

$\Rightarrow$ общих точек в множествах P_0 и S – нет.

Применим к множествам P и S теорему о разделяющей гиперплоскости.
Существует разделяющая гиперплоскость, т.е. существует

$$\begin{pmatrix}u_0 \\ u\end{pmatrix} \in R^{m+1},\; \begin{pmatrix}u_0 \\ u\end{pmatrix} \ne 0 :\; u_0\omega_0 + (u, \omega) \ge u_0 z_0 + (u, z) \quad \text{для } \forall \begin{pmatrix}z_0 \\ z\end{pmatrix} \in P \text{ и } \forall \begin{pmatrix}\omega_0 \\ \omega\end{pmatrix} \in S .$$

При этом вектор $\begin{pmatrix}u_0 \\ u\end{pmatrix} \ge 0$, т.к. компоненты векторов из P неограниченны снизу.

Выберем $\begin{pmatrix}z_0 \\ z\end{pmatrix}$ на границе множества P: $\begin{matrix} z_0 = \varphi(x^*) \\ z = 0 \end{matrix}$ и $\begin{pmatrix}\omega_0 \\ \omega\end{pmatrix} \in S$: $\begin{matrix} \omega_0 = \varphi(x) \\ \omega = f(x) \end{matrix}$.

Тогда получим:

$$\forall x \in R^n \;\; u_0\varphi(x) + (u, f(x)) \ge u_0\varphi(x^*) \qquad (*)$$

Покажем что $u_0 \ne 0$ (тем самым покажем, что $u_0 > 0$, т.к. по условию $u_0 \ge 0$).
Допустим, что $u_0 = 0$, тогда $(u, f(x)) \ge 0 \;\; \forall x \in R^n$.

При этом, поскольку $\begin{pmatrix}u_0 \\ u\end{pmatrix} \ne 0$, то существует индекс i: $u_i \ne 0$, т.е. $u_i > 0$.

С другой стороны, $\forall x \in X$ $f(x) \le 0 \Rightarrow$ для $u_i > 0$ для $\forall x \in X$ $f_i(x) = 0$, что противоречит свойству регулярности.

Итак, $u_0 > 0$, и определим $\lambda^* \overset{def}{=} \dfrac{1}{u_0} u \ge 0$.

Для этого вектора соотношение (*) примет вид:

$\forall x \in R^n \;\; \varphi(x^*) \le \varphi(x) + (\lambda^*, f(x))$ (**)

$\Rightarrow$ при $x = x^*$ $(\lambda^*, f(x^*)) \ge 0$.

Но т.к. $\lambda^* \ge 0$, а $f(x^*) \le 0$ (*поскольку* $x^* \in X$) $\Rightarrow (\lambda^*, f(x^*)) = 0$.

Далее, для $\forall \lambda \ge 0 \quad (\lambda, f(x^*)) \le 0$.

Собирая все вместе, получим:

$$\varphi(x^*) + \underbrace{(\lambda, f(x^*))}_{\le 0} \;\le\; \varphi(x^*) + \underbrace{(\lambda^*, f(x^*))}_{=0} \;\le\; \varphi(x) + (\lambda^*, f(x)),$$

или $\quad L(x^*, \lambda) \le L(x^*, \lambda^*) \le L(x, \lambda^*) \;\; \forall \lambda \ge 0, \forall x \in R^n,$

т.е. (x^*, λ^*) – седловая точка функции Лагранжа, *ч.т.д.*

Замечание.

Теорема Куна-Таккера лежит в *основе теории двойственности* математического программирования. Она также находит *применение в численных методах* решения задач математического программирования. Она позволяет исходную задачу заменить задачей отыскания седловой точки функции Лагранжа, т.е. задачей вида:

$$\min_{x \in R^n} \max_{\lambda \ge 0} L(x, \lambda)$$

"Простые" ограничения этой задачи позволяют применять для её решения методы, во многом схожие с численными методами безусловной оптимизации, достаточно хорошо изученные и апробированные.

Двойственность.

В формулировке теоремы Куна-Таккера прямые и двойственные переменные (x и λ) входят симметричным образом, поэтому можно ожидать, что аналогичная симметрия существует и для задач оптимизации (относительно прямых и двойственных переменных). Действительно, рассмотрим функцию

$$g(x) \overset{def}{=} \sup_{\lambda \ge 0} L(x, \lambda), \; где \;\; L(x, \lambda) = \varphi(x) + \big(\lambda, f(x)\big),$$

Очевидно, что

$$g(x) = \begin{cases} \varphi(x), \; если \;\; f_i(x) \le 0, \; i = 1, \dots, m \\ \infty, \text{ в противном случае} \end{cases}$$

$\Rightarrow$ исходная задача $\quad \min_{x \in X} \varphi(x)$ (1)

может быть представлена в виде $\min_{x \in X} g(x)$ (1-а)

Совершенно аналогично, введем функцию $\psi(\lambda) \overset{def}{=} \inf_{x \in R^n} L(x, \lambda)$ и рассмотрим задачу

$$\max_{\lambda \ge 0} \psi(\lambda) \qquad (2)$$

Задача (2) называется двойственной, а задача (1) или (1-а) – прямой.

Теорема (двойственность). Справедливы следующие соотношения:

1) Для всех допустимых x и λ (т.е. $x\in X$, $\lambda \geq 0$)

$$\varphi(x) = \psi(\lambda) \qquad (3)$$

2) Если прямая задача регулярна, x^* – её решение, λ^* – множители Лагранжа, то λ^* – решение задачи (2) и справедливо

$$\varphi(x^*) = \psi(\lambda^*) \qquad (4)$$

3) Если для допустимых x^* и λ^* имеет место (4), то x^* – решение прямой задачи, а λ^* – решение двойственной задачи.

Доказательство.

1) Если $x\in X$, $\lambda \geq 0$, то имеем:

$$\varphi(x) \geq \varphi(x) + (\lambda, f(x)) = L(x,\lambda) \geq \inf_{x' \in R^n} L(x',\lambda) = \psi(\lambda), \text{ ч.т.д.}$$

2) Пусть x^* – решение задачи (1), λ^* – множители Лагранжа, тогда

$$\psi(\lambda^*) = \inf_{x \in R^n} L(x,\lambda^*) = L(x^*,\lambda^*) \geq L(x^*,\lambda) \geq \inf_{x \in R^n} L(x,\lambda) = \psi(\lambda)$$

для $\forall \lambda \geq 0$, т.е. λ^* – решение (2), при этом, поскольку $L(x^*,\lambda^*) = \varphi(x^*)$, то

$$\psi(\lambda^*) = \varphi(x^*), \text{ ч.т.д.}$$

3) Пусть $x^*\in X$, $\lambda^* \geq 0$ и выполняется соотношение $\varphi(x^*) = \psi(\lambda^*)$, тогда рассмотрим произвольные допустимые x, $\lambda \Rightarrow$ в силу (3) имеем

$$\varphi(x) \geq \psi(\lambda^*) = \varphi(x^*) \geq \psi(\lambda),$$

т.е. x^* – решение прямой задачи, λ^* – решение двойственной задачи, ч.т.д.

Замечания.

1. Можно свести задачу к другой с размерностью m, которая может оказаться при $m \ll n$ значительно проще.
2. Неравенство (3) позволяет получить оценку снизу для min в задаче (1) $\Rightarrow$ можно оценить точность приближенного решения.

   Все зависит от того, насколько просто можно вычислить $\psi(\lambda^*)$

Двойственные задачи линейного программирования

Рассмотрим задачу линейного программирования в основной форме:

$$\varphi(x) = (c,x) \to \min$$

Допустимое множество $X = \{x \in R^n : (A_i, x) \geq b_i;\ i = 1,\dots,m;\ x \geq 0\}$.

По теореме Куна-Таккера для задачи выпуклого программирования (а ЗЛП есть ЗВП) наличие оптимальной точки x^* эквивалентно наличию седловой точки (x^*,λ^*) функции Лагранжа:

$$L(x^*,\lambda^*)=\max_{\lambda\ge 0}\min_{x\in R^n}L(x,\lambda)=\min_{x\in R^n}\max_{\lambda\ge 0}L(x,\lambda)\text{, где}$$

$$L(x,\lambda)=\varphi(x)+(\lambda,\,f(x))=\varphi(x)+\sum_{i=1}^{m}\lambda_i f_i(x)$$

(*предполагаем, что условия регулярности выполняются*).

Если обозначить $\psi(\lambda)=\min\limits_{x\in R^n}L(x,\lambda)$, то получаем *двойственную задачу*:

$$\max_{\lambda\ge 0}\psi(\lambda)=\psi(\lambda^*)=\varphi(x^*)=\min_x\varphi(x).$$

Построим двойственную задачу к исходной задаче линейного программирования, рассматривая её, *как задачу выпуклого программирования* (напомним, что там допустимое множество задается неравенством вида $f_i(x)\le 0$).

Имеем $(m+n)$-ограничений:

$$\left.\begin{array}{l|l} f_i(x)=b_i-(A_i,x)\le 0,\ i\in 1\ldots m & \lambda_i \\ f_{m+j}(x)=-x_j\le 0,\ j\in 1\ldots n & \lambda_j{}' \end{array}\right\}\ \lambda>0$$

Каждому ограничению сопоставим элементы λ_i, $\lambda_j{}'$ - компоненты вектора $\lambda>0$.

Тогда функция Лагранжа:

$$L(x,\lambda,\lambda')=(c,x)+\sum_{i=1}^{m}\lambda_i[b_i-(A_i,x)]+\sum_{j=1}^{n}\lambda_j{}'(-x_j)$$

Т.к. $\lambda=\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_m\end{pmatrix}$, $\lambda'=\begin{pmatrix}\lambda_1{}'\\ \vdots\\ \lambda_n{}'\end{pmatrix}$, $A=\begin{pmatrix}A_1\\ \vdots\\ A_m\end{pmatrix}$, $b=\begin{pmatrix}b_1\\ \vdots\\ b_m\end{pmatrix}$, то

$$L(x,\lambda,\lambda')=(c,x)+(\lambda,b)-(\lambda,Ax)-(\lambda',x)=(b,\lambda)+\left(x,c-A^T\lambda-\lambda'\right)$$

$$(\lambda,Ax)=\left(A^T\lambda,x\right)$$

Введем функцию ψ

$$\psi(\lambda,\lambda')\overset{def}{=}\min_{x\in R^n}L(x,\lambda,\lambda')=\begin{cases}-\infty,\ \text{если } c-A^T\lambda-\lambda'\ne 0\\ (b,\lambda),\ \text{если } c-A^T\lambda-\lambda'=0\end{cases}$$

Итак,

$$\max_{\substack{\lambda\ge 0,\\ \lambda'\ge 0}}\psi(\lambda,\lambda')\quad=\quad\max_{\substack{\lambda\ge 0,\ \lambda'\ge 0,\\ \underbrace{c-A^T\lambda}_{\ge 0}-\lambda'=0}}(b,\lambda)\quad=\quad\max_{\substack{\lambda\ge 0,\\ A^T\lambda\le c}}(b,\lambda)$$

по условию $\lambda'\ge 0\Leftrightarrow c-A^T\lambda\ge 0\Leftrightarrow A^T\lambda\le c$

Получаем *двойственную задачу линейного программирования*:

Целевая функция $\psi(\lambda)=(b,\lambda)\to\max$

Допустимое множество: $\Lambda=\left\{\lambda\in R^m : A^T\lambda\le c,\ \lambda\ge 0\right\}$

При этом:

- размерность исходной ЗЛП (n) совпадает с числом ограничений в двойственной: $A^T\lambda \le c$, и наоборот, число ограничений (m) в исходной ЗЛП совпадает с размерностью двойственной;
- min меняется на max, знаки неравенств меняются на противоположные.

Справедливы следующие утверждения:

1) *Двойственность взаимна*, т.е. задача, двойственная к двойственной – исходная.

   Действительно,

   рассмотрим задачу, эквивалентную двойственной:

$$\min_{\Lambda}\left(-b,\lambda\right), \quad \Lambda = \left\{\lambda \in R^m : -A^T\lambda \ge c,\ \lambda \ge 0\right\}$$

   $\Rightarrow$ получили ЗЛП в основной форме. Построим к ней двойственную:

$$\max_{X}\left(-c,x\right), \quad X = \left\{x \in R^n : -Ax \le -b,\ x \ge 0\right\}$$

   $\Rightarrow$ эта задача эквивалентна исходной:

$$\min_{X}\left(c,x\right), \quad X = \left\{x \in R^n : Ax \ge b,\ x \ge 0\right\}.$$

2) Если решение исходной задачи линейного программирования существует, то существует и решение двойственной ЗЛП, причем *экстремумы целевых функций совпадают* (было доказано в теореме о двойственности).

3) Экстремальная точка λ^* двойственной задачи является *векторным коэффициентом чувствительности* исходной задачи по вектору b.

   Рассмотрим видоизмененную задачу с вектором правых частей $b+\Delta b$:

$$b + \Delta b = b + \begin{pmatrix} 0 \\ \vdots \\ \varepsilon_i \\ \vdots \\ 0 \end{pmatrix}$$

Для пассивных ограничений $(A_i, x^*) > b_i$ небольшое изменение b_i не нарушит строгого неравенства.

При этом заметим, что из условия $\lambda_i^* f_i(x^*) = \lambda_i^*\left[b_i - (A_i, x^*)\right] = 0$, которое называется условием дополняющей нежесткости, следует, что $\lambda_i^* = 0$ для пассивных ограничений. Для активных ограничений $(A_i, x^*) = b_i$ изменение b_i может привести к большому изменению экстремума.

λ_i^* *характеризует скорость изменения экстремума, т.е.* $\dfrac{\partial \varphi}{\partial \varepsilon_i} = \lambda_i^*$.

Действительно, для линейной задачи функция Лагранжа L имеет вид

$$L(x,\lambda)=(c,x)+(\lambda,b-Ax),\quad \varphi(x^*)=L(x^*,\lambda^*)$$

$$\Delta\varphi=\Delta L=(c,x)+(\lambda^*,b+\varepsilon-Ax)-(c,x)-(\lambda^*,b-Ax)=(\lambda^*,\varepsilon)\Rightarrow\Delta\varphi=(\lambda^*,\varepsilon)$$

$$\Rightarrow\lim_{\varepsilon_i\to 0}\frac{\Delta\varphi}{\varepsilon_i}=\lambda_i^*,\quad ч.т.д.$$

Пример.

Провести анализ чувствительности в следующей задаче оптимизации. Для изготовления изделий четырех видов $A_1,...,A_4$ используют ресурсы трех типов, причем запасы ресурсов ограничены. Исходные данные задачи представлены в таблице.

Тип ресурсов	Расход ресурсов на изготовление одного изделия A_j при его стоимости c_j				Запасы ресурсов
	A_1	A_2	A_3	A_4	
	$c=27$	$c=10$	$c=15$	$c=28$	
I	3	2	1	2	20
II	3	1	3	4	50
III	2	1	1	2	60

Цель: составить план выпуска изделий A_j, обеспечивающий max стоимость произведенной продукции.

Взяв в качестве управляемых переменных $x_j,\ j=1,...,4$ – количество выпускаемых изделий A_j получим следующую математическую модель:

$$\varphi(x)=-27x_1-10x_2-15x_3-28x_4\to\min$$

$$\begin{cases}-3x_1-2x_2-x_3-2x_4\ge-20\\-3x_1-x_2-3x_3-4x_4\ge-50\\-2x_1-x_2-x_3-2x_4\ge-60\\x_j\ge 0,\ j=1,...,4\end{cases}$$

Решив задачу симплекс-методом, найдем: $x^*=(0,0,10,5)$, т.е. max стоимость произведенной продукции $\varphi^*=290$ будет получена, если изделия A_1 и A_2 не выпускать, а изготовить 10 изделий A_3 и 5 изделий A_4.

Двойственная задача имеет вид:

$$\psi(y) = -20y_1 - 50y_2 - 60y_3 \to \max$$

$$\begin{cases} -3y_1 - 3y_2 - 2y_3 \le -27 \\ -2y_1 - y_2 - y_3 \le -10 \\ -y_1 - 3y_2 - y_3 \le -15 \\ -2y_1 - 4y_2 - 2y_3 \le -28 \\ y_i \ge 0, i = 1,2,3 \end{cases}$$

Решив её, получим $y^* = (12,1,0)$, $\psi(y^*) = 290$.

Из этого решения видно, что при небольших приращениях Δb_1 запасов ресурса I максимально достижимая стоимость изготовленной продукции φ^* изменится на величину $12\Delta b_1$

Например, если этот ресурс представляет собой сырье, то увеличение его запасов на 1 кг при оптимальном планировании, вызовет возрастание стоимости изготовленной продукции на 12 руб. То же приращение ресурсов II обеспечит увеличение объема продукции только на 1 руб. И, наконец, изменение в небольших пределах запасов ресурса III вовсе не повлияет на стоимость произведенной продукции.

Это означает, что запасы ресурса III при оптимальном плане расходуются не полностью и являются избыточными.

$\Rightarrow$ Наиболее дефицитным в рассматриваемой задаче является ресурс I, его запасы следует по возможности, увеличивать в первую очередь. Второй ресурс менее дефицитен, а запасы ресурса III превосходят потребности, соответствующие оптимальному плану выпуска изделий.