Семинар № 8
«Разные конусы и сопряжённые функции»

Александр Катруца

10 ноября 2016 г.

1. Конусы

1.1 Конус возможных направлений

Надеюсь все помнят определение конуса, а также определение нормального конуса, данное в связи с понятим условного субдиференциала. Помимо нормального конуса важным понятием для сдачи курса является множество, называемое *конус возможных направлений*.

Определение 1 Конусом возможных направлений для множества $G \subset \mathbb{R}^n$ в точке $\mathbf{x}_0 \in G$ будем называть такое множество $\Gamma(\mathbf{x}_0|G) = \{\mathbf{s} \in \mathbb{R}^n | \mathbf{x}_0 + \alpha\mathbf{s} \in G,\ 0 \leq \alpha \leq \overline{\alpha}(\mathbf{s})\}$, где $\overline{\alpha}(\mathbf{s}) > 0$.

Определение достаточно интуитивно, а именно есть точка $\mathbf{x}_0 \in G$ и множество векторов с началом в этой точке, принадлежащих множеству G. Такое множество векторов и образует конус возможных направлений. Направления *возможные*, так как они не выводят за пределы множества. В частности, для выпуклого множества это определение можно переформулирвать в виде

Определение 2 Конусом возможных направлений для выпуклого множества $X \subset \mathbb{R}^n$ в точке $\mathbf{x}_0 \in X$ будем называть такое множество $\Gamma(\mathbf{x}_0|X) = \{\mathbf{s} \in \mathbb{R}^n | \mathbf{s} = \lambda(\mathbf{x} - \mathbf{x}_0),\ \lambda > 0, \forall \mathbf{x} \in X\}$.

Далее для вычисления конуса возможных направлений для множества G заданного в виде

$$G = \{\mathbf{x} \in \mathbb{R}^n | \varphi_i(\mathbf{x}) \leq 0,\ i = \overline{0, n-1};\ \varphi_i(\mathbf{x}) = \mathbf{a}_i^\intercal \mathbf{x} - b_i = 0,\ i = \overline{n, m}\}, \tag{1}$$

где $\varphi_i(\mathbf{x}),\ i = \overline{0, n-1}$ — выпуклы и G регулярно[1] воспользуемся следующим фактом, который легко устанавливается напрямую из определения. Конус возможных направлений для множества (1) в точке $\mathbf{x}_0$ задаётся в виде:

$$\Gamma(\mathbf{x}_0|G) = \{\mathbf{s} \in \mathbb{R}^n | \nabla\varphi_i(\mathbf{x}_0)^\intercal \mathbf{s} \leq 0, i \in I, \mathbf{a}_i^\intercal \mathbf{s} = 0, i = \overline{n, m}\}, \tag{2}$$

$I = \{i : \varphi_i(\mathbf{x}_0) = 0,\ i = \overline{0, n-1}\}$ и $\nabla\varphi_i(\mathbf{x}) \in \partial\varphi_i(\mathbf{x})$. Также получим выражение для сопряжённого конуса возможных направлений $\Gamma^*(\mathbf{x}_0|G)$:

$$\Gamma^*(\mathbf{x}_0|G) = \left\{ \mathbf{p} \in \mathbb{R}^n \middle| \mathbf{p} = \sum_{i=n}^{m} \lambda_i \mathbf{a}_i - \sum_{i \in I} \mu_i \nabla\varphi_i(\mathbf{x}_0) \right\}, \tag{3}$$

[1] То есть градиенты активных ограничений линейно незавиисмы.

где $\lambda_i \in \mathbb{R}$, $\mu_i \in \mathbb{R}_+$. Соотношение (3) может быть получено напрямую из определения сопряжённого конуса.

Упражнение. Покажите, что выполнено равенство $\mathcal{N}(\mathbf{x}_0|G) = -\Gamma^*(\mathbf{x}_0|G)$.

Задача Найдите $\Gamma(\mathbf{x}_0|G)$ и $\Gamma^*(\mathbf{x}_0|G)$ из геометрических соображений и испольщуя формулы (2), (3):

$$G = \{\mathbf{x} \in \mathbb{R}^2 | x_1^2 + 2x_2^2 \leq 3,\ x_1 + x_2 = 0\}.$$

1.2 Касательный (контингентный) конус

Ещё один конус, про который надо знать, — это *касательный конус*. Из названия не совсем очевидным образом следует формальное определение.

Определение 3 Касательным конусом к множеству G в точке $\mathbf{x}_0 \in \overline{G}$ называется следующее множество $T(\mathbf{x}_0|G) = \{\lambda \mathbf{z} | \lambda > 0,\ \exists \{\mathbf{x}_k\} \subset G,\ \mathbf{x}_k \to \mathbf{x}_0, \mathbf{x}_k \neq \mathbf{x}_0,\ \lim\limits_{k\to\infty} \frac{\mathbf{x}_k - \mathbf{x}_0}{\|\mathbf{x}_k - \mathbf{x}_0\|_2} = \mathbf{z}\}$

То есть это множество направлений, по которым можно сойтись по последовательностям из внутренности G к точке из границы G.

По аналогии с конусом возможных направлений рассмотрим как выглядит касательный конус для множества G:

$$G = \{\mathbf{x} \in \mathbb{R}^n | \varphi_i(\mathbf{x}) \leq 0, i = \overline{0, n-1}\ \varphi_i(\mathbf{x}) = 0, i = \overline{n, m}\}. \tag{4}$$

Из определения явно следует, что касательный конус для множества G (4) записывается как

$$T(\mathbf{x}_0|G) = \{\mathbf{z} \in \mathbb{R}^n | \nabla\varphi_i^\intercal(\mathbf{x}_0)\mathbf{z} \leq 0, i \in I,\ \nabla\varphi_i^\intercal(\mathbf{x}_0)\mathbf{z} = 0, i = \overline{n, m}\}$$

и соответствующий ему сопряжённый конус:

$$T^*(\mathbf{x}_0|G) = \left\{ \mathbf{p} \in \mathbb{R}^n \middle| \mathbf{p} = \sum_{i=n}^{m} \lambda_i \nabla\varphi_i(\mathbf{x}_0) - \sum_{i \in I} \mu_i \nabla\varphi_i(\mathbf{x}_0) \right\},$$

где $\mu_i \in \mathbb{R}_+$, $\lambda_i \in \mathbb{R}$, $I = \{i | \varphi_i(\mathbf{x}_0) = 0, i = \overline{0, n-1}\}$ Заметим, что для выпуклых множеств касательный конус совпадает с конусом возможных направлений.

Пример: найти $T(\mathbf{x}_0|G)$ и $T^*(\mathbf{x}_0|G)$ для множества $G = \{\mathbf{x} \in \mathbb{R}^2 | x_1 + x_2 \leq 1,\ x_1^2 + 2x_2^2 = 1\}$

1.3 Острый экстремум

Введём ещё один тип минимумов (экстремумов) функции.

Определение 4 Точка $\mathbf{x}^* \in G$ является точкой острого минимума функции f на множестве G, месли существует такое число $\gamma > 0$, что $f(\mathbf{x}) - f(\mathbf{x}^*) \geq \gamma \|\mathbf{x} - \mathbf{x}^*\|_2$ для всех $\mathbf{x} \in G$.

Определение неудобно для проверки точек минимума на остроту, поэтому сформулируем следующий факт.

Факт об остром минимуме (максимуме): пусть f дифференцируемая функция на $G \subset \mathbb{R}^n$. Точка $\mathbf{x}^* \in G$ — точка острого минимума (максимума) функции f на множестве G тогда и только тогда, когда существует $\alpha > 0$ такое что $\nabla f^\intercal(\mathbf{x}^*)h \geq \alpha > 0$ ($-\nabla f^\intercal(\mathbf{x}^*)h \geq \alpha > 0$) для всех $h \in T(\mathbf{x}^*)$ и $\|h\|_2 = 1$.

Пример.

$$
\begin{aligned}
&\min x_1^2 + x_2^2 \\
\text{s.t. } &x_1^2 + 2x_2^2 = 2 \\
&x_1 + x_2 \leq 1
\end{aligned}
\tag{5}
$$

Будем решать задачу графически с помощью линий уровня. Из картинки следует, что $\mathbf{x}_1^* = (0, 1)$ и $\mathbf{x}_2^* = (0, -1)$ — точки минимума и $f(\mathbf{x}_{1,2}^*) = 1$, где f — целевая функция. Найдём касательный конус для множества G в этих точках. На графике показано, что $T(\mathbf{x}_1^*|G) = \left\{ \lambda \begin{pmatrix} -1 \\ 0 \end{pmatrix} \middle| \lambda \in \mathbb{R}_+ \right\}$ и $T(\mathbf{x}_2^*|G) = \left\{ \lambda \begin{pmatrix} -1 \\ 0 \end{pmatrix} \middle| \lambda \in \mathbb{R} \right\}$. Также прямое вычисление даёт $\nabla f(\mathbf{x}_{1,2}^*) = \begin{pmatrix} 0 \\ \pm 2 \end{pmatrix}$. Из выражений для градиента целевой функции и векторов из касательных конусов в точках минимума, очевидно, что их скалярное произведение для точек $\mathbf{x}_{1,2}^*$ равно нулю. Таким образом, эти точки не являются точками острого экстремума.

Далее рассмотрим точки локального максимума $\mathbf{x}_3^* = (-\sqrt{2}, 0)$ и $\mathbf{x}_4^* = (4/3, -1/3)$.

- Точка $\mathbf{x}_3^*$. В этой точке градиент целевой функции $\nabla f(\mathbf{x}_3^*) = \begin{pmatrix} -2\sqrt{2} \\ 0 \end{pmatrix}$ и касательный конус $T(\mathbf{x}_3^*|G) = \left\{ \mathbf{p} \middle| \mathbf{p} = \lambda \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \lambda \in \mathbb{R} \right\}$ Скалярное произведение $\nabla f^\mathsf{T}(\mathbf{x}_3^*)\mathbf{p}$ может быть как положительным, так и отрицательным. Следовательно, точка $\mathbf{x}_3^*$ не является точкой острого экстремума.

- Точка $\mathbf{x}_4^*$. В этой точке $\nabla f(\mathbf{x}_4^*) = \dfrac{2}{3} \begin{pmatrix} 4 \\ -1 \end{pmatrix}$. Обозначим первое ограничение $\varphi_1(\mathbf{x}) = 0$, а второе $\varphi_2(\mathbf{x}) \leq 0$. Тогда $\nabla \varphi_1(\mathbf{x}_4^*) = \dfrac{4}{3} \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ и $\nabla \varphi_2(\mathbf{x}_4^*) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Тогда $T(\mathbf{x}_4^*|G) = \{\mathbf{p} \in \mathbb{R}^2 \mid p_1 + p_2 \leq 0,\ 2p_1 - p_2 = 0\} = \left\{ \mathbf{p} \in \mathbb{R}^2 \,\middle|\, \mathbf{p} = \lambda \begin{pmatrix} -1 \\ -2 \end{pmatrix},\ \lambda \in \mathbb{R}_+ \right\}$. Определим для какой λ выполнено $\|\mathbf{p}\|_2 = 1$. Элементарные вычисления дают $\lambda^* = \dfrac{1}{\sqrt{5}}$. Тогда, используя факт об остром экстремуме для максимума, посчитаем скалярное произведение $\dfrac{2}{3} \begin{pmatrix} -4 \\ 1 \end{pmatrix}^\mathsf{T} \dfrac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ -2 \end{pmatrix} \geq 10^{-10} > 0$ Таким образом, точка $\mathbf{x}_4^*$ является точкой острого максимума.

2. Сопряжённые функции

Важным общематематическим понятием является понятие *сопряжённой функции*.

Определение 5 Пусть $f : \mathbb{R}^n \to \mathbb{R}$. Функция $f^* : \mathbb{R}^n \to \mathbb{R}$ называется сопряжённой функцией к функции f и определена как

$$
f^*(\mathbf{y}) = \sup_{\mathbf{x} \in dom\, f} (\mathbf{y}^\mathsf{T}\mathbf{x} - f(\mathbf{x})).
$$

Область определения f^* — это множество таких $\mathbf{y}$, что супремум конечен.

Для нахождения сопряжённой функции необходимо задать её область определения, а затем оценить выражение под супремумом сверху.

Далее рассмотрим некоторые примеры.

Пример 1.

Найти сопряжённую функцию к функции $f(x) = ax + b$.

Решение.

По определедению $f^*(y) = \sup\limits_{x \in \mathbb{R}}(yx - ax - b) = \sup\limits_{x \in \mathbb{R}}((y - a)x - b)$. Под супремумом стоит линейная функцию по x, которая ограничена только при $y = a$. Следовательно, областью определения сопряжённой функции f^* является одна точка $\{a\}$, и значение сопряжённой функции в этой точке равно $f^*(a) = -b$.

Пример 2.

Найти сопряжённую функцию к функции $f(x) = x \log x$ при $x > 0$.

Решение.

Аналогично предыдущему примеру: $f^*(y) = \sup\limits_{x>0}(xy - xlogx)$. Функция под супремумом ограничена сверху на области определения при любом y (проверьте!). Поэтому областью определения f^* является $\mathbb{R}$. Найдём максимум из условия первого порядка: $g'(x) = y - \log x - 1 = 0$. Откуда $x^* = e^{y-1}$ и $f^*(y) = ye^{y-1} - e^{y-1}(y - 1) = e^{y-1}$.

Пример 3.

Найти сопряжённую функцию к функции $f(\mathbf{x}) = \|\mathbf{x}\|$.

Решение.

По традиции рассмотрим определение $f^*(\mathbf{y}) = \sup\limits_{\mathbf{x} \in \mathbb{R}^n} (\mathbf{y}^\intercal \mathbf{x} - \|\mathbf{x}\|)$. Вспомним (или узнаем), что двойственная норма $\|\cdot\|_*$ определяется как $\|\mathbf{z}\|_* = \sup\limits_{\|\mathbf{x}\| \leq 1} (\mathbf{z}^\intercal \mathbf{x})$. Напрямую из определения следует, что для всех $\mathbf{x}$ и $\mathbf{z}$ выполнено неравенство $\mathbf{z}^\intercal \mathbf{x} \leq \|\mathbf{x}\| \|\mathbf{z}\|_*$. Если $\|\mathbf{y}\|_* > 1$, тогда существует вектор $\mathbf{z}$ такой что $\mathbf{z}^\intercal \mathbf{y} > 1$. Возьмём $\mathbf{x} = t\mathbf{z}$ при $t \to \infty$ и получим неограниченность функции под супремумом:

$$\mathbf{y}^\intercal \mathbf{x} - \|\mathbf{x}\| = t(\mathbf{y}^\intercal \mathbf{z} - \|\mathbf{z}\|) \to \infty$$

Наоборот, если $\|\mathbf{y}\|_* \leq 1$, тогда $\mathbf{y}^\intercal \mathbf{x} \leq \|\mathbf{x}\| \|\mathbf{y}\|_*$ и $\mathbf{y}^\intercal \mathbf{x} \leq \|\mathbf{x}\|$. Значит функция под супремумом ограничена сверху 0 и достигается на нулевом векторе. Таким образом,

$$f^*(\mathbf{y}) = \begin{cases} 0 & \|\mathbf{y}\|_* \leq 1 \\ \infty, & \text{otherwise.} \end{cases}$$