Семинар № 8
«Разные конусы и сопряжённые функции»

Александр Катруца

6 ноября 2016 г.

1. Разбор промежуточной контрольной

Сначала несколько замечаний о прошедшей промежуточной контрольной.

1. Большинство написали все базовые определения и формулировки, и это уже неплохо. Те, кто не написал, пожалуйста, непременно выучите это к экзамену.

2. Почему-то почти никто не взялся сформулировать задачу проекции точки на вероятностный симплекс. Симплекс — это множество вида $\left\{\mathbf{x} \in \mathbb{R}^n_+ \mid \sum\limits_{i=1}^n x_i = 1\right\}$. В условии он был назван вероятностным, потому что $\mathbf{x}$ можно интерпретировать как дискретное вероятностное распределение. И чтобы навести вас на определение симплекса, если вы его забыли. Таким образом, задача нахождения проекции точки $\mathbf{y}$ на вероятностный симплекс формулируется следующим образом:

$$\begin{aligned} &\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x} - \mathbf{y}\|_2 \\ \text{s.t. } & x_i \geq 0 \\ & \sum_{i=1}^n x_i = 1 \end{aligned} \tag{1}$$

   и является выпуклой, так как целевая функция выпукла на выпуклой области определения — симплексе.

3. Далее, небольшое напоминание определения выпуклой задачи.

   Определение 1 Оптимизационная задача называется *выпуклой*, если целевая функция является выпуклой на выпуклом допустимом множестве.

4. В задаче про нормальный конус имелся в виду конус, с помощью которого опрелелялся условный субдифференциал. Но за ответ про конус, заданный некоторой нормой, я снижал совсем немного.

5. В задаче, где надо было проверить выпуклость множества, зависящего от параметра t, некоторые довольно мучительно, влоб, показали, что это множество пустое и поэтому выпуклое. Однако ожидалось, что вы заметите, что при фиксированном t — это

множество — шар, а значит выпукло. А дальше надо было привести утверждение про выпуклость пересечения любого числа выпукых множеств. Про этот приём я упоминал на семинаре.

6. В задаче про нахождение множества разделяющих плоскостей для заданной точки и множества, кажется, толкьо один человек правильно набросал путь решения. А именно надо было нарисовать картинку (специально задача дана в 2D) и понять, что для каждой точки, лежащей на оси ординат, между «углом» множества и данной точкой существует свой набор разделябщих гиперплоскостей. То есть первый параметр — это точка на оси ординат, второй параметр — это множество угловых коэффициентов, которое зависит от первого параметра. Далее нужно было аккуратно провести вычисления предельных положения разделяющих гиперплоскостей и записать ответ.

В целом, у меня осталось скорее положительное впечатление от вашей работы, однако я уже упоминал, что дальше контрольные будут оцениваться строже. По крайней мере по отношению к тем, кто не может написать базовые определения курса.

2. Конусы

2.1 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 2 Конусом возможных направлений для множества $G \subset \mathbb{R}^n$ в точке $\mathbf{x}_0 \in G$ будем называть такое множество $\Gamma(\mathbf{x}_0|G) = \{\mathbf{s} \in \mathbb{R}^n | \mathbf{x}_0 + \alpha\mathbf{s} \in G,\ 0 \le \alpha \le \overline{\alpha}(\mathbf{s})\}$, 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 3 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

$$G = \{\mathbf{x} \in \mathbb{R}^n | \varphi_i(\mathbf{x}) \le 0,\ i = \overline{0, n-1};\ \varphi_i(\mathbf{x}) = \mathbf{a}_i^\intercal \mathbf{x} - b_i = 0,\ i = \overline{n, m}\}, \tag{2}$$

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 (2) в точке $\mathbf{x}_0$ задаётся в виде:

$$\Gamma(\mathbf{x}_0|G) = \{\mathbf{s} \in \mathbb{R}^n | \nabla\varphi_i(\mathbf{x}_0)^\intercal \mathbf{s} \le 0, i \in I, \mathbf{a}_i^\intercal \mathbf{s} = 0, i = \overline{n, m}\}, \tag{3}$$

[1] То есть градиенты активных ограничений линейно незавиисмы.

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

$$\Gamma^*(\mathbf{x}_0|G) = \left\{\mathbf{p} \in \mathbb{R}^n \middle| \mathbf{p} = \sum_{i=n}^{m} \lambda_i \mathbf{a}_i - \sum_{i\in I} \mu_i \nabla\varphi_i(\mathbf{x}_0)\right\}, \tag{4}$$

где $\lambda_i \in \mathbb{R}$, $\mu_i \in \mathbb{R}_+$. Соотношение (4) 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 (3), (4):

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

2.2 Касательный (контингентный) конус

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

Определение 4 Касательным конусом к множеству G в точке $\mathbf{x}_0 \in \overline{G}$ называется следующее множество $T(\mathbf{x}_0|G) = \{\lambda\mathbf{z} | \lambda > 0,\ \exists\{\mathbf{x}_k\} \subset G,\ \mathbf{x}_k \to \mathbf{x}_0, \mathbf{x}_k \ne \mathbf{x}_0,\ \lim\limits_{k\to\infty} \frac{\mathbf{x}_k - \mathbf{x}_0}{\|\mathbf{x}_k - \mathbf{x}_0\|_2} = \mathbf{z}\}$

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:

$$G = \{\mathbf{x} \in \mathbb{R}^n | \varphi_i(\mathbf{x}) \le 0, i = \overline{0, n-1}\ \varphi_i(\mathbf{x}) = 0, i = \overline{n, m}\}. \tag{5}$$

Из определения явно следует, что касательный конус для множества G (5) записывается как

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

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 $G = \{\mathbf{x} \in \mathbb{R}^2 | x_1 + x_2 \le 1,\ x_1^2 + 2x_2^2 = 1\}$

2.3 Острый экстремум

Введём ещё один тип минимумов (экстремумов) функции.

Определение 5 Точка $\mathbf{x}^* \in G$ является точкой острого минимума функции f на множестве G, месли существует такое число $\gamma > 0$, что $f(\mathbf{x}) - f(\mathbf{x}^*) \ge \gamma\|\mathbf{x} - \mathbf{x}^*\|_2$ для всех $\mathbf{x} \in G$.

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

Факт об остром минимуме: 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 $\nabla f^\mathsf{T}(\mathbf{x}^*)h \geq \alpha > 0$ $(-\nabla f^\mathsf{T}(\mathbf{x}^*)h \geq \alpha > 0)$ для всех $h \in T(\mathbf{x}^*)$ и $\|h\|_2 = 1$.

Пример.

$$\begin{aligned} &\min x_1^2 + x_2^2 \\ \text{s.t. } &x_1^2 + 2x_2^2 = 2 \\ &x_1 + x_2 \leq 1 \end{aligned} \tag{6}$$

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 $\nabla f(\mathbf{x}_4^*) = \frac{2}{3}\begin{pmatrix} 4 \\ -1 \end{pmatrix}$. Обозначим первое ограничение $\varphi_1(\mathbf{x}) = 0$, а второе $\varphi_2(\mathbf{x}) \leq 0$. Тогда $\nabla \varphi_1(\mathbf{x}_4^*) = \frac{4}{3}\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ 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 $\lambda^* = \frac{1}{\sqrt{5}}$. Тогда, используя факт об остром экстремуме для максимума, посчитаем скалярное произведение $\frac{2}{3}\begin{pmatrix} -4 \\ 1 \end{pmatrix}^\mathsf{T} \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ -2 \end{pmatrix} \geq 10^{-10} > 0$ Таким образом, точка $\mathbf{x}_4^*$ является точкой острого максимума.

3. Сопряжённые функции