Домашняя работа №1 по курсу «Методов Оптимизации»

Коврижных Дмитрий Б05-903

10.10.2021

К чему был вопрос про Disco Elysium?

1 Matrix

1 $\nabla f(x)$ **где** $f(x) = \frac{1}{2}\|Ax - b\|_2^2$
Я знаю, что $df(x) =< \nabla f(x), dx >$, поэтому $df(x) =< Ax - b, d(Ax - b) >$, так как $d(\|x\|_2^2) =< x, dx >$ и получится, что $df(x) =< Ax - b, A \cdot dx$, а затем используем перенос с транспонированием и хоба: $df(x) =< A^T(Ax - b), dx > => \nabla f(x) = A^T(Ax - b)$

2 $\nabla f(x)$ **где** $f(x) =< x, x >^{<x,x>}$**.**
$d(f(x)) = d(e^{<x,x>\cdot \ln <x,x>}) = [d(< x, x > \cdot \ln < x, x >) = (\frac{<x,x>}{<x,x>} \cdot < 2x, dx > + \ln < x, x > * < 2x, dx >] => \nabla f(x) = (2x) \cdot (1 + \ln < x, x >) < x, x >^{<x,x>}$

3 Calculate the Frobenious norm derivative: $\dfrac{\partial}{\partial X}\|X\|_F^2$
$d(\|X\|_F) = d(< X, X >^{\frac{1}{2}}) = \frac{1}{2} < X, X >^{\frac{-1}{2}} \cdot d(< X, X >)$ а как мы знаем из конспектов первого семинара, $d < X, X >=< 2X, dX > => f(x) =< X, X >^{(\frac{1}{2})\cdot(-1)} \cdot < X, dX >$ или же, хорошо бы написать ответ через норму Фробениуса: $df(x) = \|X\|_F^{-1} \cdot < X, dX >$
$d(\|X\|_F^2) = 2\|X\|_F^1 \cdot \|X\|_F^{-1} \cdot < X, dX >=< 2X, dX >$ Тогда: $\dfrac{\partial}{\partial X}\|X\|_F^2 = 2X$ - Ответ.

4 $f(t) = det(A - tI_n)$
Известно, что $d(det(X)) = det(X) < X^{-T}, dx >= det(X)X^{-T}$ поэтому $d(det(A - tI_n)) = det(A - tI_n) \cdot -I_n^T \cdot (A - tI_n)^{-T}$. То есть первая производная равна $det(A - tI_n) \cdot -I_n^T \cdot (A - tI_n)^{-T}$. Где I_n^T можно убрать, но я оставлю.
Тогда вторая производная равна: $d(det(A - tI_n)) \cdot -I_n^T \cdot (A - tI_n)^{-1} + det(A - tI_n) \cdot d(-I_n^T \cdot (A - tI_n)^{-T}) = det(A - tI_n) \cdot -I_n^T \cdot (A - tI_n)^{-T} \cdot -I_n^T \cdot (A - tI_n)^{-T} + H1 \cdot det(A - tI_n)$. Распишем $H1$ чтобы не таскать левую часть: $H1 = d(-I_n^T \cdot (A - tI_n)^{-T}) = -I_n^T \cdot < d((A - tI_n)^{-1}, d(tI_n) >= -I_n^T \cdot < d((A - tI_n)^{-1} \cdot -I_n^T, dt >$

2 Convex sets

1 Prove that the set of square symmetric positive definite matrices is convex.
Множество выпукло, если $\forall \theta \in [0, 1] : C = \theta A + (1 - \theta)B$; и оба множителя не равны нулю одновременно.

Для двух матриц A и B верно, что $\forall x \in \mathbb{R}\backslash 0 : x^T A x > 0$ аналогично для B, т.е они положительно определены.
Тогда $x^T C x = \theta x^T A x + (1-\theta) x^T B x > 0$, Данная сумма симметрична и положительно определена, поэтому это верно для C, и в силу произвольности C множество выпукло.

3 Show that the hyperbolic set of $\{x \in \mathbb{R}^n_+ | \prod\limits_{i=1}^n x_i \geq 1\}$ is convex. Hint: For $0 \leq \theta \leq 1$ it is valid, that $a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b$ with non-negative a, b.

Пусть $_1, _2, x_i$ - элементы нашего множества L. Запишем исходное неравенство с учётом свойства множества L(1 и 2 - это индексы): $\prod\limits_{i=1}^n (x_i^1)^\theta (x_i^2)^{1-\theta} = (\prod\limits_{i=1}^n x_i^1)^\theta (\prod\limits_{i=1}^n x_i^2)^{1-\theta} \leq \prod\limits_{i=1}^n (\theta x_i^1 + (1-\theta) x_i^2) = RightPart.$

$$RightPart = \begin{cases} \prod\limits_{i=1}^n (x_i^1)^\theta \geq 1 & \theta = 1 \\ (\prod\limits_{i=1}^n x_i^2)^{1-\theta} \geq 1 & \theta = 0 \\ (\prod\limits_{i=1}^n x_i^1)^\theta (\prod\limits_{i=1}^n x_i^2)^{1-\theta} & \theta \in (0;1) \end{cases}$$

Тогда L выпукло по определению $\forall x^1, x^2 \in L, \forall \theta \in [0,1] : \theta x_i^1 + (1-\theta) x^2 \in L$

4 Условие:

В одну сторону просто: если S - выпукло, то $(\alpha + \beta) S = \alpha S + \beta S \Leftrightarrow x \in \alpha S + \beta S$. Т.е x принадлежит левой, то он принадлежит и правой и наоборот. По свойству выпуклых множеств. Обратно, утверждение верно для неотрицательных, коэффициентов, ок пусть они равны как в определении выпуклости $\alpha = \theta; \beta = 1 - \theta$, тогда по равенству из условия: $(\theta + 1 - \theta) S = \theta S + (1-\theta) S$. S выпукло по определению, если вместо S в правую часть подставить элементы из него. Чтд. (Пояснение в видео: https://www.youtube.com/watch?v=dQw4w9WgXcQ)

5 Здесь должно быть условие Интро: симплекс это выпуклое множество, также нам известно, что счётное пересечение не изменяет выпуклости, что позволяет нам пересекать симплекс и множество без потери выпуклости.
Разберём случаи:
1 $\alpha > a_i \Rightarrow \beta \geq 0$, так как вероятность нулевая и выполняется для вещественных иксов, иначе множество x - пустое. Оба варианта множеств выпуклы.
1(1)
Пусть некоторые $a_i > \alpha$, начиная с некоторого индекса в силу упорядоченности, тогда $\beta \geq \sum_{i=index}^n a_i p_i$. Используя выпуклость пересечения с симплексом, получим, что по определению $\sum_{i=j}^n a_i (_i^1 + (1-\theta) p_i^2) \leq \theta \beta + (1-\theta)\beta = \beta$. Доказано.
2
Так же, как и в первом возьмём два случайных вектора, записываем из них выпуклую комбинацию и как в неравенстве выше будет определение выпуклости.
3
Снова берём два произвольных вектора, так как ограничений на p нет и $\sum_i^n = j|a_i^2|(\theta p_i^1 + (1-\theta) p_i^2) = \theta \sum_{i=j}^n |a_i^2| p_i^1 + (1-\theta) \sum_{i=j}^n |a_i^2| p_i^2 \geq \theta\alpha + (1-\theta)\alpha = \alpha$ так как $beta \geq \sum_{i=index}^n a_i p_i$. Чтд.
4
$\mathbb{V}x = y^T p - p^T a a^T p$ где $y_i = a_i^2$ Распишем выпуклую комбинацию: $y^T(\theta p_i^1 + (1-\theta) p_i^2) - (\theta p_i^1 + (1-\theta) p_i^2)^T a a^T (\theta p_i^1 + (1-\theta) p_i^2) = y^T \theta p_i^1 + y^T (1-\theta) p_i^2) - \theta^2 (p_i^1)^T a a^T (p_i^1) - 2\theta(1-$

$\theta)(p_i^2)^T aa^T(p_i^1) - (1-\theta)^2(p_i^2)^T aa^T(p_i^2)$ Выделяем полный квадрат и получаем выпуклость: $\alpha \leq \alpha + \theta(1-\theta)(a^Tp^1 - a^Tp^2)^2 \leq \mathbb{V}x$ Доказано.

3 Subgradient

1 Prove, that x_0 - is the minimum point of a convex function $f(x)$ if and only if $0 \in \partial f(x_0)$

Функция f - выпукла, а точка x_0 - её минимум.
В одну сторону $/ra$, у выпуклой функции $f(x) \geq f(x_0) = f(x_0) + \langle 0, x - x_0 \rangle$, так называемый "умный ноль". Т.е функция $g = 0$, получим, что $0 \in df(x_0)$.
В другую сторону, известно: $0 \in df(x_0)$, тогда возьмём $g = 0$, тогда $f(x_0) = f(x_0) + \langle 0, x - x_0 \rangle$, а так как $\langle 0, x - x_0 \rangle = 0$, получим утверждение в обратную сторону: $f(x) \geq f(x_0)$. Чтд.

2 Find $\partial f(x)$, if $f(x) = \textbf{ReLU}(x) = \max\{0, x\}$

Функции 0 и x выпуклы, а поточечный максимум определяется как $f(x) = \underset{i}{max} f_i(x)$ абъюзим (используем) теорему Дубовицкого - Милютина.

$$f_1(x) = 0 \Rightarrow \partial f_1(x) = 0;\ \ f_1(x) = x \Rightarrow \partial f_1(x) = 1;$$

$$\partial f(x) = \begin{cases} 1 & \text{x} > 0 \\ [0,1] & \text{x} = 0 \\ 0 & \text{x} < 0 \end{cases}$$

Ответ.

3 Find $\partial f(x)$, if $f(x) = \|x\|_p$ при $p = 1, 2, \infty$

1. **p = 1**

   Из функана я знаю, что манхэттенское расстояние (о, как загнул, ладно, первая норма) равна $\|x\|_1 = \sum_{i=0}^n |x_i|$, и, так как x - это вектор-столбец, то в каждой строчке у него одно значение x_i, то есть $\|x\|_1 = \sum_{i=0}^n |x_i| = \sum_{i=0}^n max(x_i, -x_i)$.

   Поэтому при дифференцировании x по одной из координат мы на её месте получим 1, а на остальных 0. Раз тут сумма, а осталась только одна теорема, то, так как все функции $x, -x$ под суммой выпуклые, а для максимума мы уже знаем из предыдущего пункта и $\partial(x_i) = 1;\ \partial(-x_i) = -1$, то:

$$\partial f(x) = \begin{cases} (0, \dots, 1, \dots 0) & x_i > 0 \\ (0, \dots, [-1,1], \dots 0) & x_i = 0 \\ (0, \dots, -1, \dots 0) & x_i < 0 \end{cases}$$

   Единица и минус единица стоят на i-ом месте. Уравнение похоже на функцию знака, кроме точки $x = 0$. В ней для каждого x_i все координаты кроме i равны 0, а она сама имеет значения из [-1, 1]. Так как в задаче спрашивают про $\partial f(x)$, то для случая, когда часть из координат равна нулю, а часть нет, то $A : i \in A,\ if\ x_i = 0$ и в этом случае A считаем непустым.

$$\partial f_i(x) = \begin{cases} \prod_{i=1}^{n} [-1,1] & x = 0 \\ (sign(x_1), \dots sign(x_n)) & x_i \neq 0; \forall i \text{ from 1 to n} \\ \prod_{i \notin A} (sign(x_i)) \times \prod_{i \in A} [-1,1] & \text{A} \neq \varnothing \end{cases}$$

2. **p = 2**

   Рассмотрим функцию $f(x) = \|x\|$, она дифференцируема, если $x \in \mathbb{R}^n \backslash \{0\}$, и тогда $\partial \| \cdot \| = x \cdot \|x\|^{-1}$.

   Сначала скажем, что Евклидова норма самосопряженная и $X* = X$, а сопряженная норма имеет вид $\|n\|_* = sup_{\|m\|<1}\langle n, m \rangle$

   Предположим, что в нуле $\partial \| \cdot \|(0) = SHAR_1(0)$, то есть шар радиуса 1. В обратную сторону очевидно, используя неравенство Коши-Буняковского-Шварца, $|\langle y, x\rangle| \leq \|y\| \cdot \|x\| \Leftrightarrow 1 = f(x)$. Поэтому $y \in \partial f(0)$.

   Прямо: предположим, что $y \in \partial f(0)$ и норма y лежит вне единичного шара, но тогда есть точка, где скалярное произведение больше 1, так как рассмотрим сопряженную норму $sup_{\|x\|,1}\langle y, x\rangle = \|y\| > 1$. Тогда для некой точки x_a, для которой это верно, имеем, что $\|x_a\| \leq 1 < \langle y, x_a\rangle$. Но возникает противоречие для нашей точки.

   $$\partial f(x) = \begin{cases} SHAR_1(0) & \text{x} = 0 \\ \mathrm{e}^{\|x\|} \cdot x \cdot \|x\|^{-1} & \text{x} \neq 0 \end{cases}$$

3. **p = ∞**

   Вернёмся к p = 1, и снова применим теорему Дубовицкого-Милютина о поточечном максимуме. Рассмотрим наш вектор x. Когда вектор равен нулю, то берём выпуклую оболочку, согласно теореме, а именно $\prod_{i=1}^{n}[-1, 1]$.

   Когда он не равен нулю, то есть 2 варианта, то пусть максимум достигается не в одной точке, а на нескольких $f_i(x)$, ответ будет выпуклой оболочкой всех

   $$\partial f_i(x) = \begin{cases} (0, ..., 1, ..., 0) & x_i > 0 \\ (0, ..., -1, ..., 0) & x_i < 0 \end{cases}$$

   Если максимум один, т.е на одной $f_i(x)$, то жизнь проще, берем значение

   $$\partial f_i(x) = \begin{cases} (0, ..., 1, ..., 0) & x_i > 0 \\ (0, ..., -1, ..., 0) & x_i < 0 \end{cases}$$

   Ответом будет объединение трёх вариантов выше.

4 Find $\partial f(x)$, if $f(x) = \|Ax - b\|_1$

$g(x) = \|x\|_1$, тогда $\partial g(x) = \partial \|x\|_1(x)$ *and* $\phi = x$ и 0, если $x < 0$, тогда $\partial f(x) = \partial(\|Ax - b\|_1) = A^T \partial \| \cdot \|_1(Ax - b)$, так как g выпуклая и я использовал chain rule.
Ответ: $A^T \partial \| \cdot \|_1(Ax - b)$, а $\partial \| \cdot \|_1$ был найден в задаче 3.

5 Find $\partial f(x)$, if $f(x) = e^{\|x\|}$

Chain rule, я выбираю тебя! Мне кажется очевидным, что $\| \cdot \|$ выпукла, по крайней мере так сказали на функане. Пусть $\phi = exp$ тоже выпукла, тогда найдём по правилу:

$$\partial f(x) = e^{\|x\|} \cdot \partial \| \cdot \|(x)$$

Ответ: $\partial f(x) = e^{\|x\|} \cdot \partial \| \cdot \|(x)$, а $\partial \| \cdot \|$ был найден в задаче 3.