Методы оптимизации. Семинар 2. Матрично-векторное дифференцирование.

Корнилов Никита Максимович

мфти фивт

11 сентября 2025г

Дифференцируемость по Фреше

Работаем в множествах U,V линейных нормированных полных (банаховых) пространств.

Definition

Функция $f:U\to V$ дифференцируема по Фреше во внутренней точке $x\in {\rm int}\,U$, если существует линейный оператор $df(x):U\to V$, т.ч.

$$f(x+h) = f(x) + d[h] + o(||h||), ||h|| \to 0.$$

df(x) называется производной f в точке x.

Если точка x не является внутренней, то понятие дифференцируемости не определено.

Definition

Производная по направлению

В одномерном случае $f:\mathbb{R} o\mathbb{R}$, показателем скорости изменения f в точке x вдоль числовой прямой является производная:

$$f'(x) := \lim_{t \to 0} \frac{f(x+t) - f(x)}{t}.$$

В многомерном случае $f:U\to V$, направлений изменения не два, а бесконечно много. Производные по направлению отвечают за изменения функции вдоль одного направления $h\in U$:

Definition

Производной по направлению $h \in U$ функции $f: U \to V$ во внутреней точке $x \in \mathrm{int} U$ называется

$$\frac{\partial f}{\partial h}(x) := \lim_{t \to +0} \frac{f(x+y) - f(x)}{t} = \int f(x) dx$$

3 / 23

Если для любого $h\in U$ определена производная по направлению $\frac{\partial f}{\partial h}(x)$, то функция f дифференцируема по Гато в точке x.

Н. М. Корнилов 11 сентября 2025г

Связь определений

Lemma

Если функция f дифференцируема по Фреше в x, то производная по направлению $\frac{\partial f}{\partial h}(x)$ существует, линейна по h и равна дифференциалу df(x)[h].

В матанализе показывается, что из дифференцируемости по Фреше следует существование производных по всем направлением. Однако обратное неверно. Достаточным условием будет непрерывность всех частных производных.

Градиент по вектору

D P common f : Dn . D auch

ullet В случае $f:\mathbb{R}^n o\mathbb{R}$ дифференциал $df(x)[dx]\in\mathbb{R}$ можно представить в виде

$$df(x)[dx] = \langle \nabla f(x), dx \rangle$$
, где вектор $\nabla f(x) \in \mathbb{R}^n$ зависит от x . Вектор $\nabla f(x)$ называется градиентом функции. Взяв $h = e_i = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^n$, получим формулу градиента в стандартном базисе

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x)\right)^{\top} \in \mathbb{R}^n, \tag{2}$$

где $\frac{\partial f}{\partial x_i}(x):=\lim_{t\to 0}\frac{f(x+te_i)-f(x)}{t}$ — частные производные по i-ой координате.

H. М. Корнилов 11 сентября 2025г 5 / 23

Градиент по матрице

UV

ullet В случае $f:\mathbb{R}^{n imes m} o\mathbb{R}$ дифференциал $df(X)[dX]\in\mathbb{R}$ можно представить в виде

$$df(X)[dX] = \nabla f(X) dX \rangle,$$

где матрица $\nabla f(X) \in \mathbb{R}^{n imes m}$ зависит от X.

Матрица $\nabla f(X)$ также называется **градиентом** функции.

Аналогично взяв $h=e_{ij}$, получим ф $_{-}$ рмулу градиента в

стандартном базисе
$$\nabla f(X) = \left(\frac{\partial f}{\partial x_{ij}}(X)\right)_{i,j} \in \mathbb{R}^{n \times m}. \tag{3}$$

H. М. Корнилов 11 сентября 2025г 6 / 23

Матрица Якоби

ullet В случае $f:\mathbb{R}^n o\mathbb{R}^m$ дифференциал $df(x)[dx]\in\mathbb{R}^m$ можно представить в виде

$$df(x)[dx] = J_f(x)dx$$
, где матрица $J_f(x) \in \mathbb{R}^{m \times n}$ зависит от x .

Матрица $J_f(x)$ называется матрицей Якоби.

Аналогично взяв $h=e_i$, получим формулу матрицы Якоби в стандартном базисе

стандартном базисе
$$J_f(x) \equiv \frac{\partial f}{\partial x} := \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{i \not\mid j} \in \mathbb{R}^{m \times n}. \tag{4}$$
 Заметим, что $\nabla f(x) = J_x^\top$.

H. М. Корнилов 11 сентября 2025г 7 / 23

Таблица канонических видов

THE THE THE		
Выход Вход	Скаляр	Вектор
Скаляр	df(x) = f'(x)dx $f'(x)$ скаляр, dx скаляр.	-
Вектор	$df(x) = \langle \nabla f(x), dx \rangle$ $f(x)$ вектор, dx вектор	$df(x) = J_x dx$ J_x матрица, dx вектор

11 Ax> ~10fKl, Ax>

Стоит отметить, что данная таблица верна и для произвольных скалярных произведений, а не только для стандартного.

 $df(X) = \langle \nabla f(X), dX \rangle$ $\nabla f(X)$ mat, dX mat

Матрица

Подходы к вычислению производных

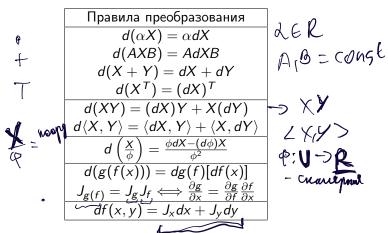
£(x/= £(x1,..., xn/

1 Прямой подход Идея: выразить функцию f(x) через скалярную зависимость от каждой координаты x_i и напрямую искать частную производную

Дифференциальный подход

Идея: Используя правила вычисления дифференциалов, получить канонический вид из Таблицы (8) и выделить градиенты функций, гессиан или матрицу Якоби.

Дифференциальное исчисление: правила



Стоит отметить, что данная таблица верна и для произвольных скалярных произведений и гельдоровых ЛНП.

Н. М. Корнилов 11 сентября 2025г 10 / 23

Дифференциальное исчисление: табличные производные

Стоит отметить, что данная таблица верна и для произвольных скалярных произведений и гельдоровых ЛНП.

Hint. Для запоминания формулы $d(X^{-1})$

Однако это не является доказательством существования дифференциала.

H. М. Корнилов 11 сентября 2025г 11 / 23

Квадратичная функция

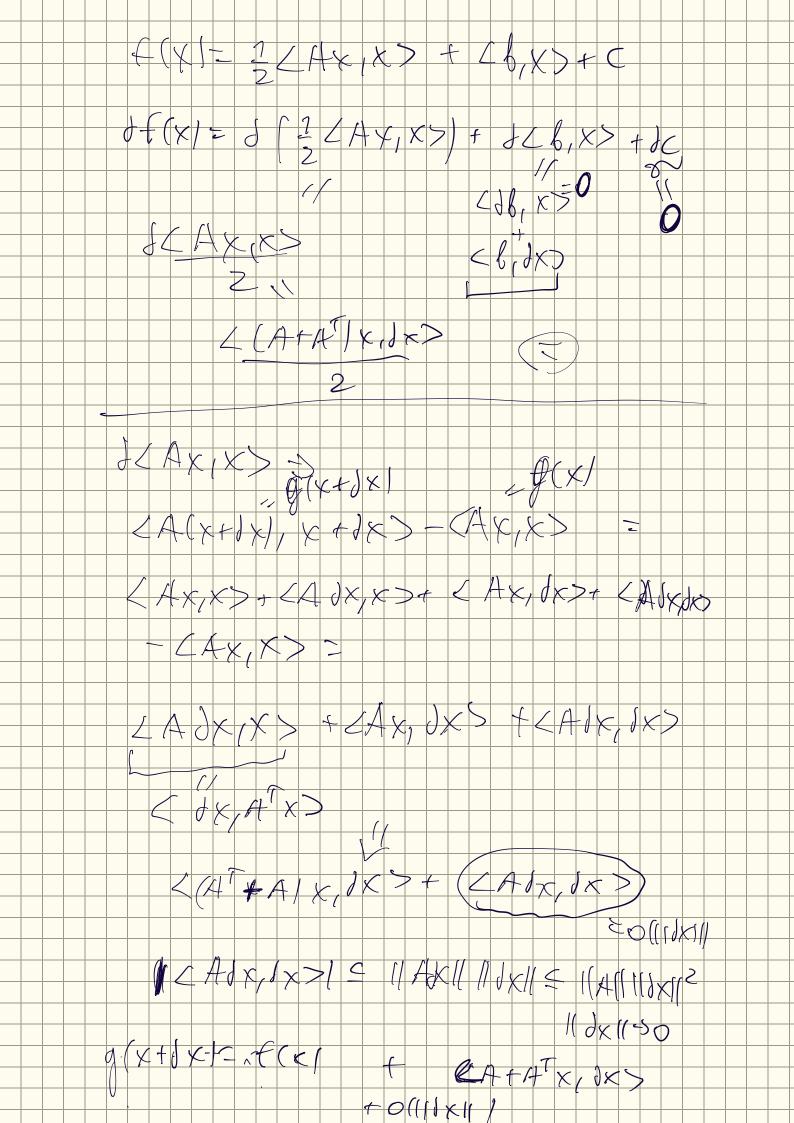
Example

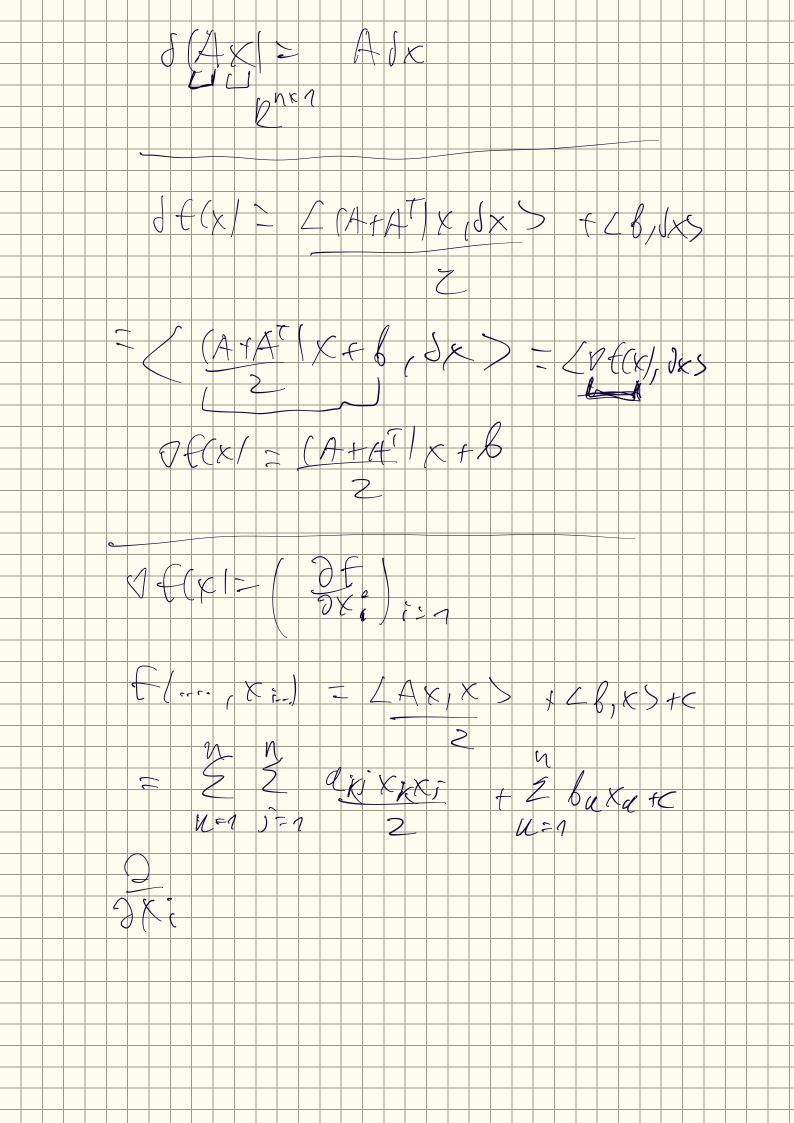
Найдите первый дифференциал и градиент функции $\nabla f(x)$ для

$$f(x) = \frac{1}{2}\langle Ax, x \rangle + \langle b, x \rangle + c,$$

где $A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n, c \in \mathbb{R}$.

Н. М. Корнилов





Вторая производная

Пусть $f:U\to V$ дифференцируема в каждой точке $x\in U$. Рассмотрим дифференциал функции $df(x)[h_1]$ при фиксированном приращении $h_1\in U$ как функцию от x:

$$g(x) = df(x)[h_1].$$

Definition (Вторая производная)

Если в некоторой точке $x\in U$ функция $g:U\to V$ дифференцируема, то второй дифференциал $d^2f(x)[h_1,h_2]:U\times U\to V$ имеет вид

$$d^{2}f(x)[h_{1},h_{2}] := d(df[h_{1}])(\mathbf{x})[h_{2}].$$
 (5)

Можно показать, что $d^2f(x)[h_1,h_2]$ билинейная функция по h_1,h_2 . По аналогии определяется третий дифференциал $d^3f(x)[h_1,h_2,h_3]$, четвёртый и так далее.

H. M. Корнилов 11 сентября 2025г 13 / 23

Гессиан

В случае $f:\mathbb{R}^n \to \mathbb{R}$ второй дифференциал, как и любую билинейную функцию, можно представить с помощью матрицы

$$d^2f(x)[dx_1,dx_2] = \langle \nabla^2 f(x)dx_1,dx_2 \rangle.$$

Матрица $\nabla^2 f(x)$ называется **гессианом** функции. В стандартном базисе гессиан имеет вид

$$\nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{ij} \in \mathbb{R}^{n \times n}.$$

Напомним, что для дважды непрерывно дифференцируемой функции гессиан - симметричная матрица. В общем случае, удобно считать гессиан как

Н. М. Корнилов 11 сентября 2025г

Квадратичная функция

$$\int \partial t = \left(\frac{\partial \nabla f_i}{\partial x_i}\right) = \left(\frac{\partial^2 f}{\partial x_i} \partial x_i\right) \quad \text{Shungh reup.}$$

$$\left(\int \partial f_i - \nabla^2 f dx_i\right) = \left(\int \partial f_i - \nabla^2 f dx_i\right) \quad \text{Shungh reup.}$$

Example

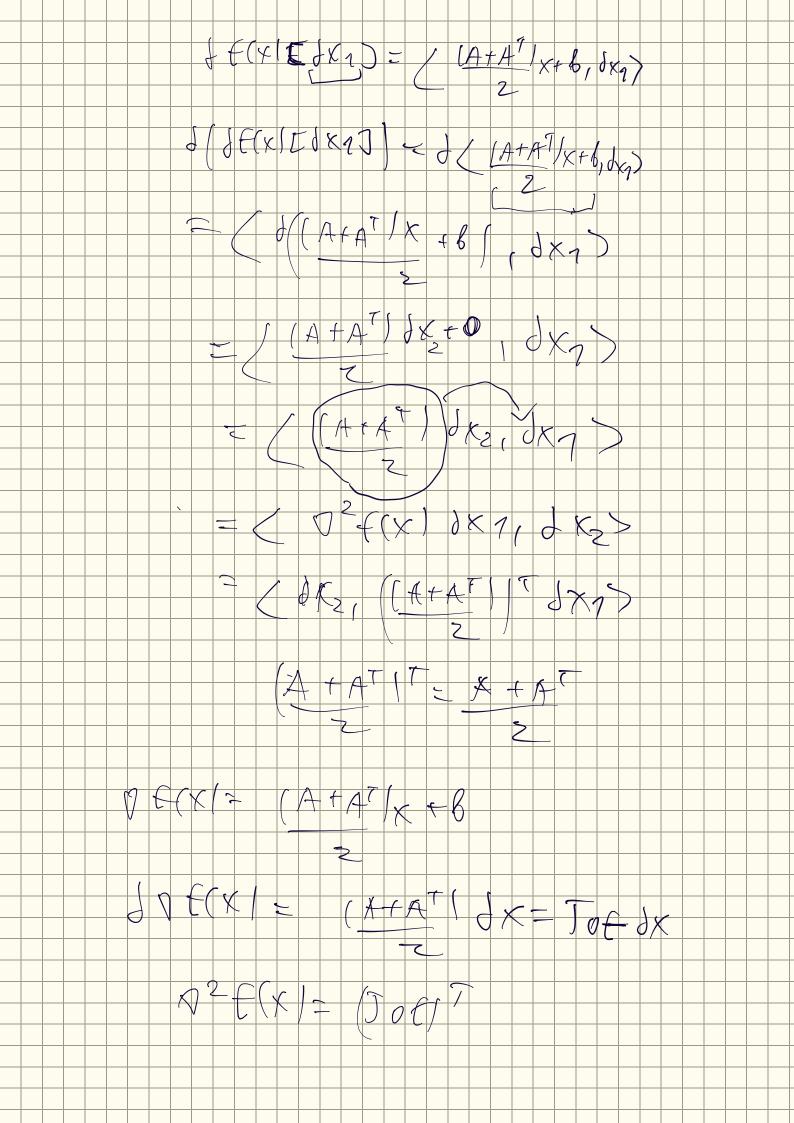
Найдите второй дифференциал и гессиан функции $abla^2 f(x)$ для

$$f(x) = \frac{1}{2}\langle Ax, x \rangle + \langle b, x \rangle + c,$$

где $A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n, c \in \mathbb{R}$.

15 / 23

Н. М. Корнилов 11 сентября 2025г



Практика

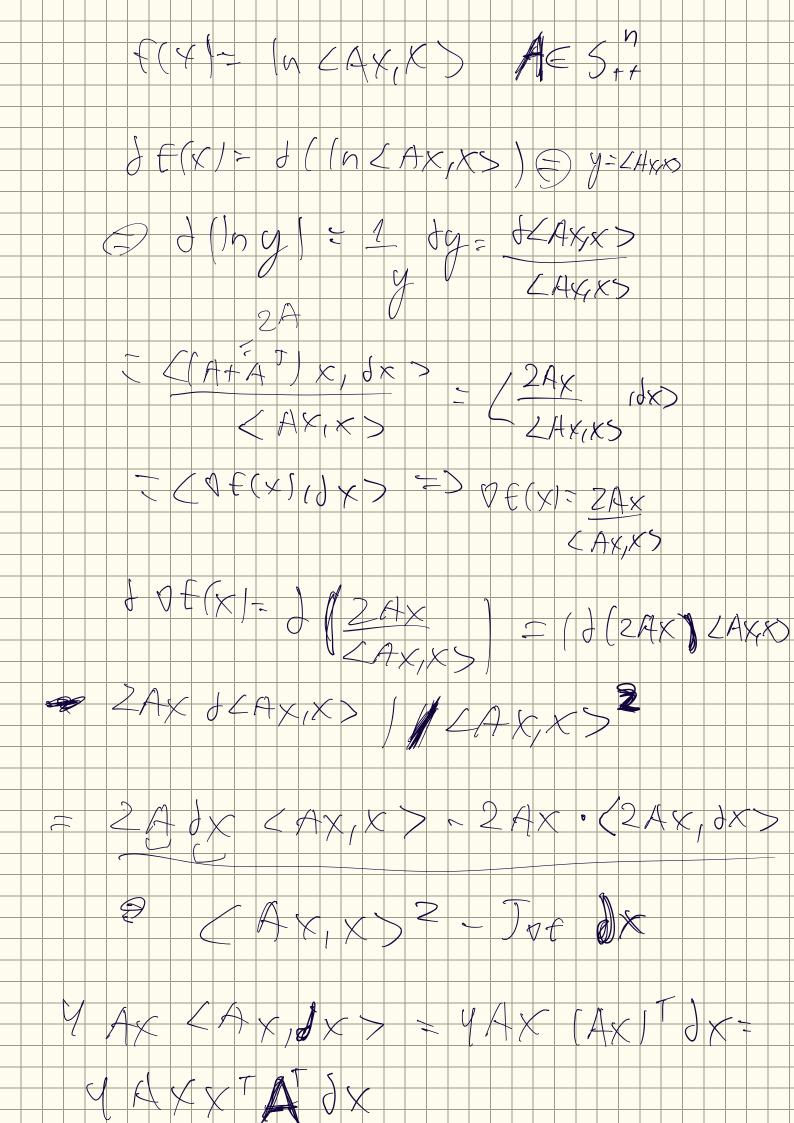
Example

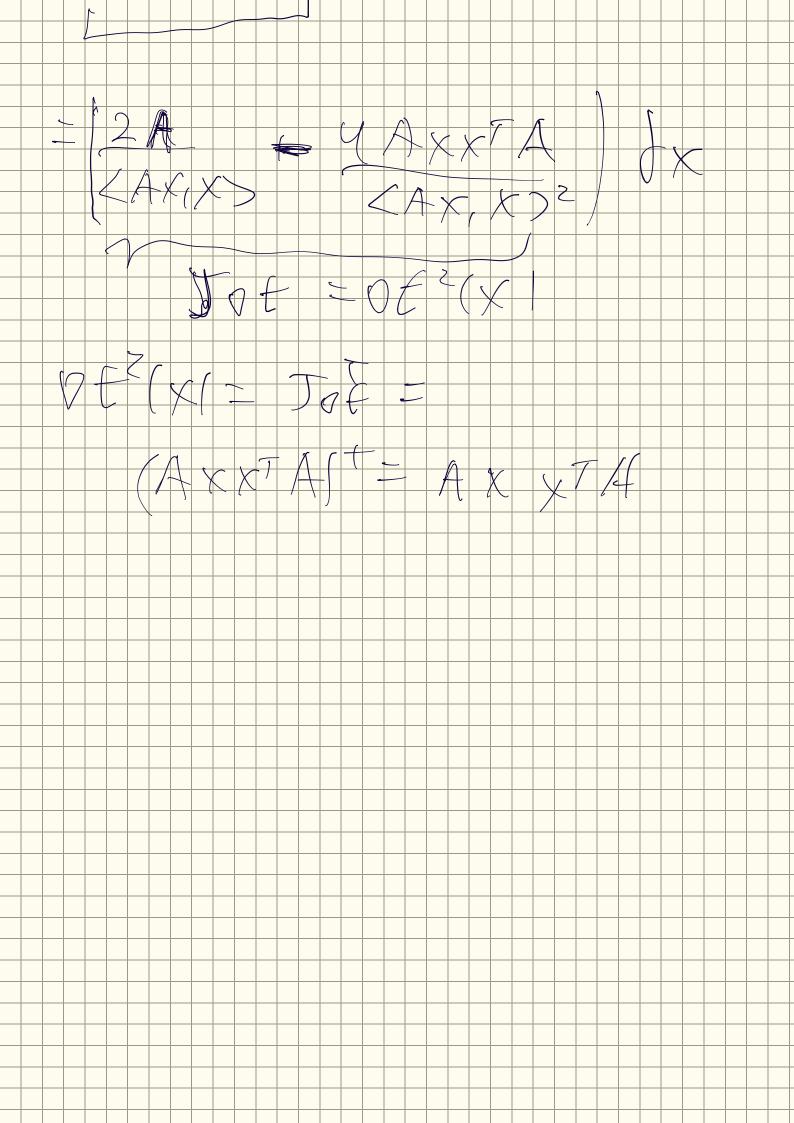
Найдите первый и второй дифференциалы, а также градиент $\nabla f(x)$ и гессиан $\nabla^2 f(x)$ функции

$$f(x) = \ln \langle Ax, x \rangle$$

где $x \in \mathbb{R}^n, A \in \mathbb{S}^n_{++}$.







Евклидова норма

Example

Найдите первый и второй дифференциалы, а также градиент $\nabla f(x)$ и гессиан $\nabla^2 f(x)$ функции

$$f(x) = \frac{1}{3} ||x||_2^3, \quad x \in \mathbb{R}^n \setminus \{0\}.$$

Логистическая регрессия

Example

Найдите первый и второй дифференциалы, а также градиент $\nabla f(x)$ и гессиан $\nabla^2 f(x)$ функции

$$f(x) = \ln(1 + \exp(\langle a, x \rangle)),$$

где $a \in \mathbb{R}^n$.

Softmax

Example

Найдите матрицу Якоби функции $s(x) = \operatorname{softmax}(x)$

$$\operatorname{softmax}(x) := \left(\frac{\exp(x_1)}{\sum_{i=1}^n \exp(x_i)}, \dots, \frac{\exp(x_n)}{\sum_{i=1}^n \exp(x_i)}\right)^\top.$$



Н. М. Корнилов

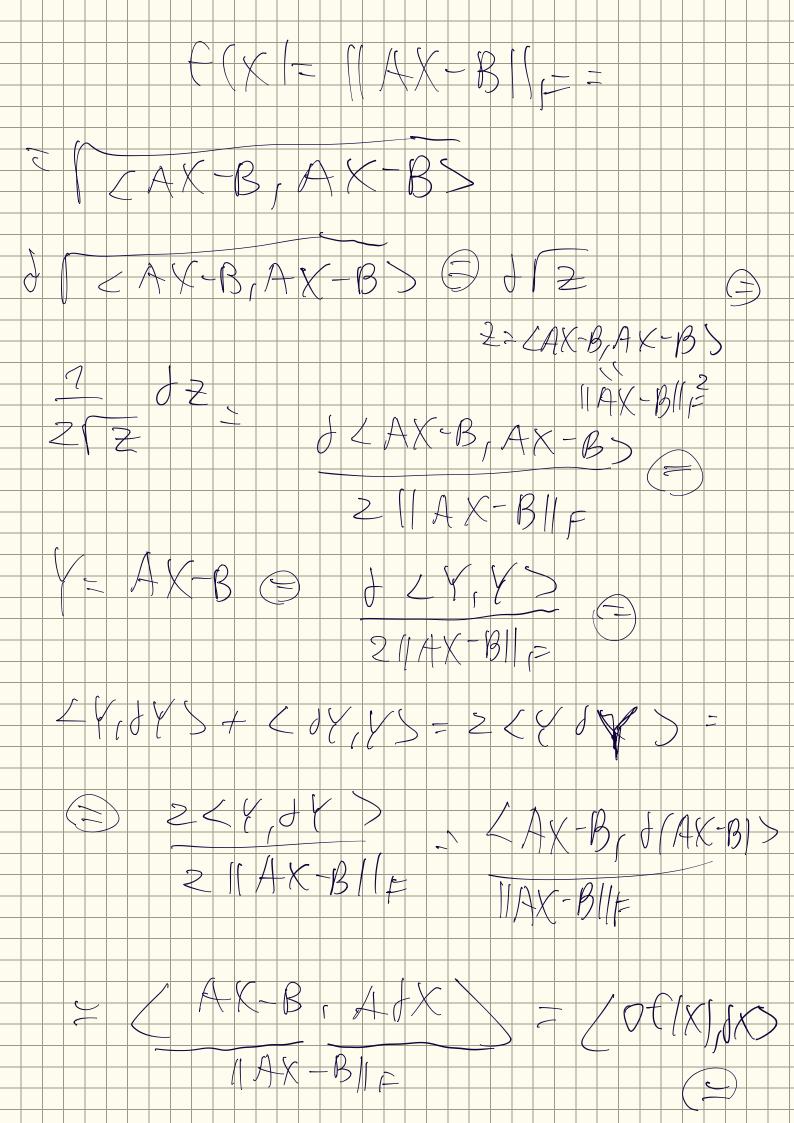
Фробениусова норма

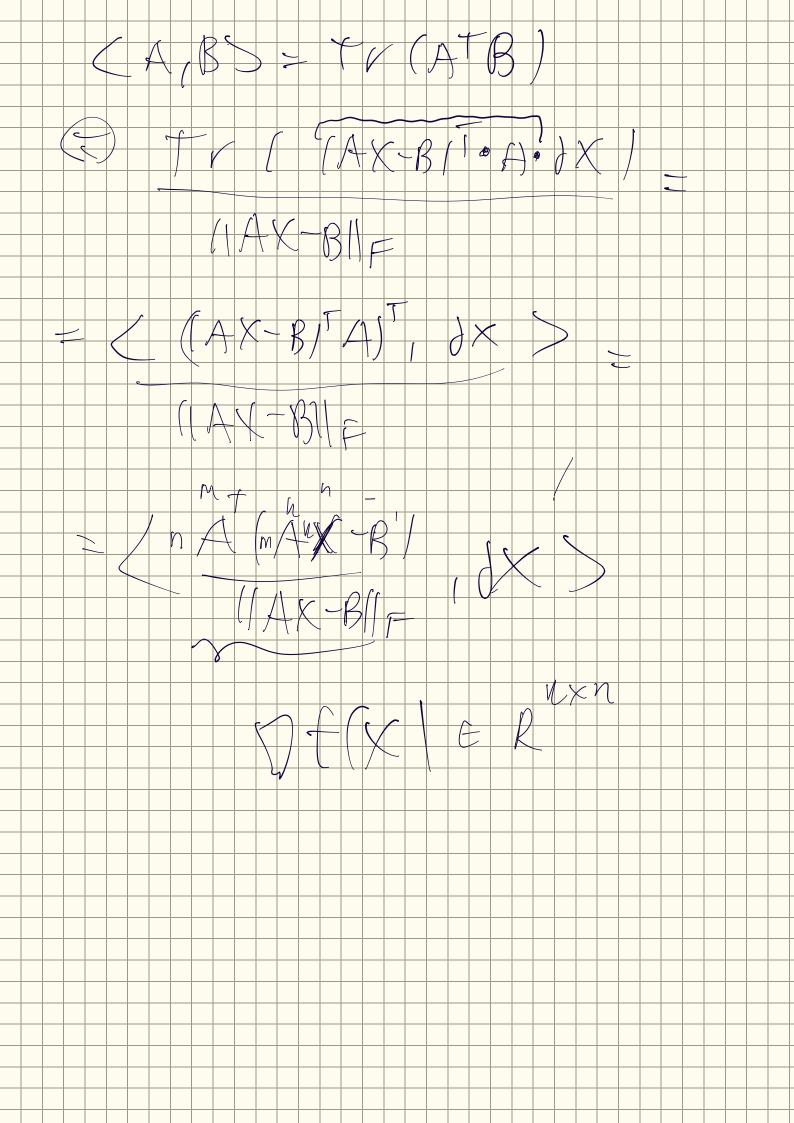
Example

Найти градиент $\nabla f(X)$ и дифференциал функции f(X)

$$f(X) = ||AX - B||_F, \quad X \in \mathbb{R}^{k \times n},$$

где $A \in \mathbb{R}^{m \times k}$, $B \in \mathbb{R}^{m \times n}$.





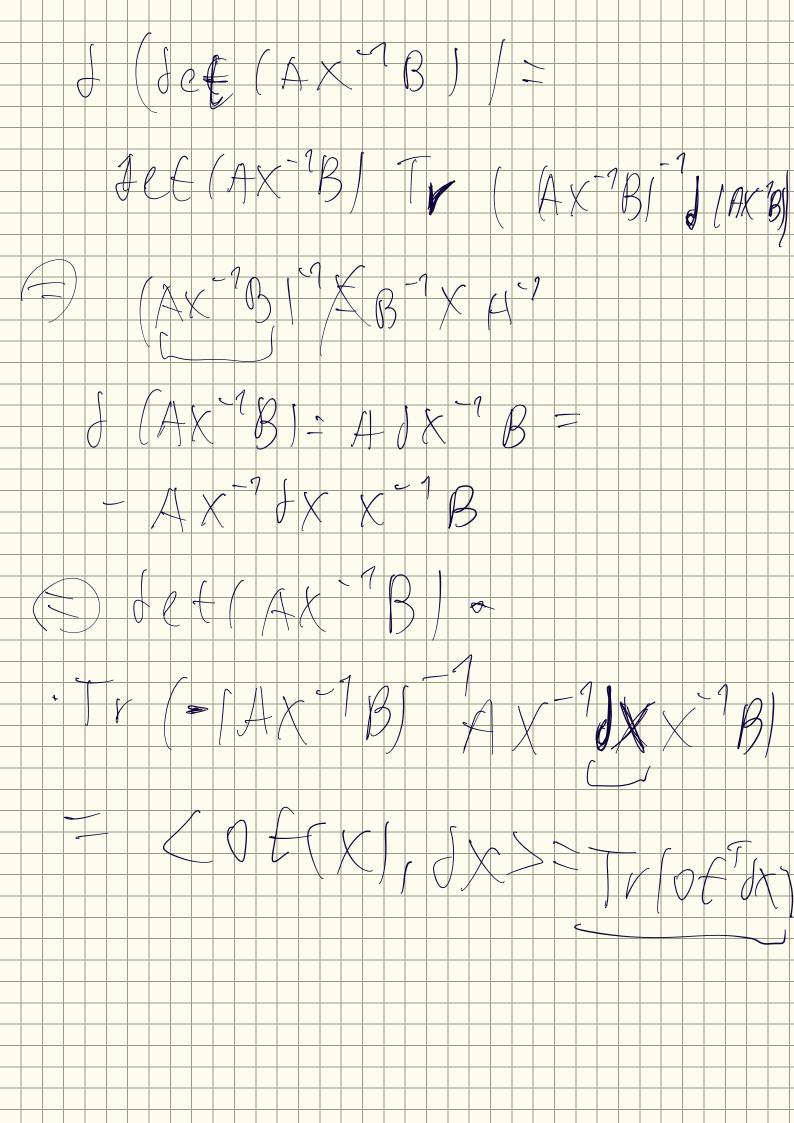
Практика

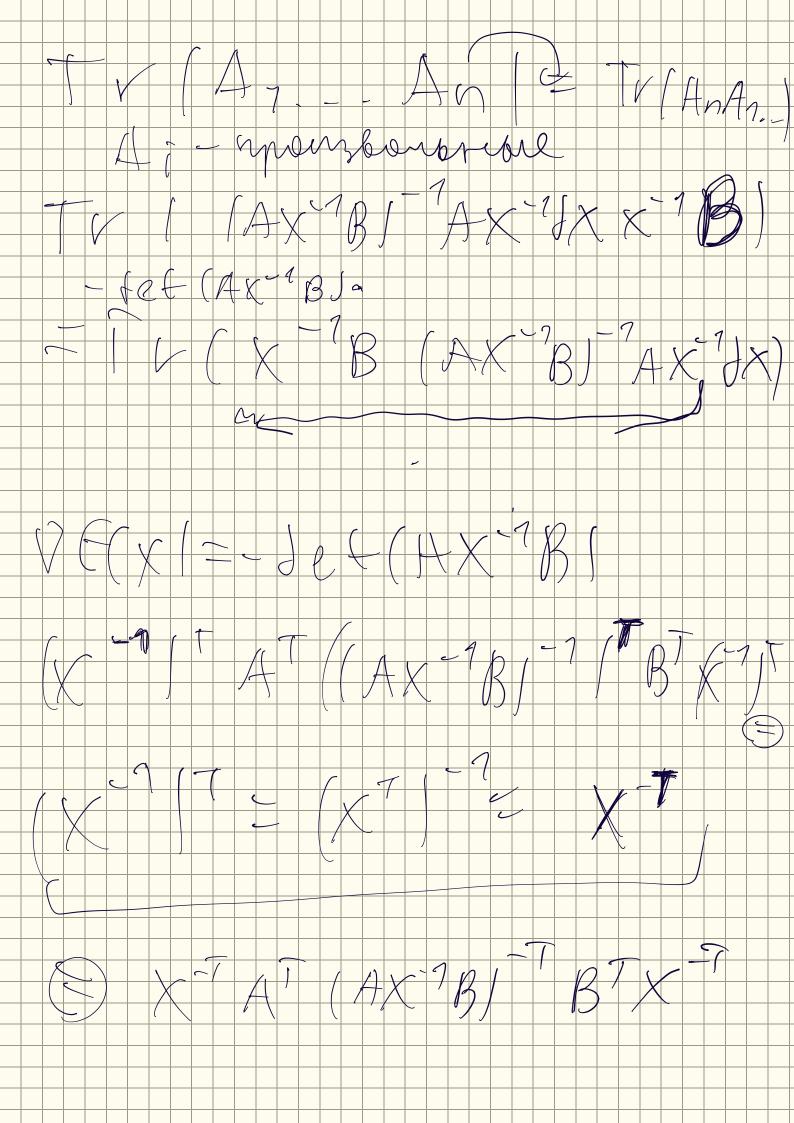
Example

Найдите первый дифференциал и градиент $\nabla f(X)$ функции f(X)

$$f(X) = \det(AX^{-1}B),$$

где A, X, B – такие матрицы с нужными размерностями, что $AX^{-1}B$ обратима.





Логарифм определителя

Example

Найдите первый и второй дифференциалы, а также градиент $\nabla f(X)$ функции f(X)

$$f(X) = \ln(\det(X))$$

заданной на множестве $X \in \mathbb{S}^n_{++}$ в пространстве \mathbb{S}^n .

Практика

Example

Найти градиент $\nabla f(X)$ и дифференциал функции f(X)

$$f(X) = \operatorname{Tr}(AX^{\top}X).$$