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 $x \in \text{int}U$, если существует линейный оператор $df(x) : U \to V$, т.ч.

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

$df(x)$ называется производной f в точке x.

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

Definition

Приращение дифференцируемой функции f в точке x с приращением h называется дифференциалом $df(x)[h] \in V$. Часто направление h обозначают как dx, а дифференциал как $df(x)[dx]$.

Производная по направлению

В одномерном случае $f : \mathbb{R} \to \mathbb{R}$, 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 $x \in \operatorname{int} U$ называется

$$\frac{\partial f}{\partial h}(x) := \lim_{t \to +0} \frac{f(x+th) - f(x)}{t}. \tag{1}$$

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.

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

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

$$df(x)[dx] = \langle \nabla f(x), dx \rangle, \quad \text{где вектор } \nabla f(x) \in \mathbb{R}^n \text{ зависит от } x.$$

Вектор $\nabla f(x)$ называется **градиентом** функции. Взяв $h = e_i = (0, \ldots, 0, \underset{i}{1}, 0, \ldots, 0) \in \mathbb{R}^n$, получим формулу градиента в стандартном базисе

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

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.

Градиент по матрице

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

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

где матрица $\nabla f(X) \in \mathbb{R}^{n\times m}$ зависит от X.
Матрица $\nabla f(X)$ также называется **градиентом** функции.
Аналогично взяв $h = e_{ij}$, получим формулу градиента в стандартном базисе

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

Матрица Якоби

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

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

Матрица $J_f(x)$ называется матрицей Якоби.
Аналогично взяв $h = e_i$, получим формулу матрицы Якоби в стандартном базисе

$$J_f(x) \equiv \frac{\partial f}{\partial x} := \left(\frac{\partial f_i}{\partial x_j}(x)\right)_{i,j} \in \mathbb{R}^{m\times n}. \tag{4}$$

Заметим, что $\nabla f(x) = J_x^\top$.

Таблица канонических видов

Вход \ Выход	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
Матрица	$df(X) = \langle \nabla f(X), dX \rangle$ $\nabla f(X)$ мат, dX мат	-

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

Подходы к вычислению производных

1. **Прямой подход**
   Идея: выразить функцию $f(x)$ через скалярную зависимость от каждой координаты x_i и напрямую искать частную производную $\frac{\partial f(x)}{\partial x_i}$.
2. **Дифференциальный подход**
   Идея: Используя правила вычисления дифференциалов, получить канонический вид из Таблицы (8) и выделить градиенты функций, гессиан или матрицу Якоби.

Дифференциальное исчисление: правила

Правила преобразования
$d(\alpha X) = \alpha dX$ $d(AXB) = AdXB$ $d(X+Y) = dX + dY$ $d(X^T) = (dX)^T$
$d(XY) = (dX)Y + X(dY)$ $d\langle X, Y\rangle = \langle dX, Y\rangle + \langle X, dY\rangle$
$d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$
$d(g(f(x))) = dg(f)[df(x)]$ $J_{g(f)} = J_g J_f \iff \frac{\partial g}{\partial x} = \frac{\partial g}{\partial f}\frac{\partial f}{\partial x}$
$df(x,y) = J_x dx + J_y dy$

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

Таблица стандартных производных
$dA = 0$ $d\langle A, X\rangle = \langle A, dX\rangle$ $d\langle Ax, x\rangle = \langle (A + A^\top)x, dx\rangle$ $d\,\mathrm{Tr}(X) = \mathrm{Tr}(dX)$
$d(\det(X)) = \det(X)\,\mathrm{Tr}(X^{-1}dX)$
$d(X^{-1}) = -X^{-1}(dX)X^{-1}$

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

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

$$
\begin{aligned}
&I = XX^{-1},\\
&dI = 0 = d(XX^{-1}) = (dX)X^{-1} + Xd(X^{-1}),\\
&d(X^{-1}) = -X^{-1}(dX)X^{-1}.
\end{aligned}
$$

Однако это не является доказательством **существования** дифференциала.

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

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 $d^2 f(x)[h_1, h_2] : U \times U \to V$ имеет вид

$$d^2 f(x)[h_1, h_2] := d(df[h_1])(x)[h_2]. \tag{5}$$

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

Гессиан

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

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

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

$$\nabla^2 f(x) = (J_{\nabla f})^\top.$$

Квадратичная функция

Example

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

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

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

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

Example

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

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

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 $a \in \mathbb{R}^n$.

$$f(x) = \ln(1 + \exp\langle a, x\rangle) \qquad x, a \in \mathbb{R}^n$$

$$a \to 0,1$$

$$df(x) = d\ln(1 + z) = \qquad z := \exp\langle a, x\rangle$$

$$\frac{1}{1+z}\,dz = \frac{d(\exp\langle a,x\rangle)}{1+\exp\langle a,x\rangle} =$$

$$\frac{\exp\langle a,x\rangle\; d\langle a,x\rangle}{1+\exp\langle a,x\rangle} =$$

$$\Big\langle \underbrace{\frac{\exp\langle a,x\rangle}{1+\exp\langle a,x\rangle} \cdot a}_{\nabla f(x)}, dx \Big\rangle$$

$$\sigma(z) = \frac{1}{1+\exp(-z)} \qquad z \in \mathbb{R} \to [0,1]$$

$$\nabla f = \sigma(\langle a,x\rangle)\, a$$

$$\sigma'(z) = -\frac{1}{(1+\exp(-z))^2}\exp(-z)(-1)$$

$$= \frac{1}{(1+\exp(-z))}\,\frac{\exp(-z)}{(1+\exp(-z))} =$$

$$\sigma(z) \qquad \sigma(-z)$$

$$\sigma'(z) = \sigma(z)\,\sigma(-z)$$

$$d\nabla f = d\big(\sigma(\langle a, x\rangle)\, a\big) =$$

$$\sigma'(\langle a, x\rangle)\, a\, d\langle a, x\rangle =$$

$$\sigma'(\langle a, x\rangle)\, a\, \underbrace{\langle a, dx\rangle}_{a^T dx} = J\nabla f\, dx$$

$$\underbrace{\sigma(\langle a, x\rangle)\,\sigma(-\langle a, x\rangle)\, a a^T}_{J\nabla f \,=\, \nabla^2 f(x)}\, dx$$

$f: U \to V$

$$f(x+dx) - f(x) = \underbrace{f'(x)}_{U \to V} dx + o(\|dx\|)$$

$$df = f'(x)dx$$

$f: \mathbb{R}^n \to \mathbb{R}$

$$df = \langle \underbrace{\nabla f(x)}_{\begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}}, dx \rangle$$

$f: \mathbb{R}^{n \times m} \to \mathbb{R}$

$$df = \langle \underbrace{\nabla f(x)}_{\left(\frac{\partial f}{\partial x_{ij}}\right)}, dX \rangle$$

$f: \mathbb{R}^n \to \mathbb{R}^m$

$$df = J_f \, dx$$

$$\nabla f = (J_f)^T \quad J_f = \left(\frac{\partial f_i}{\partial x_j} \right)$$

$f: \mathbb{R}^n \to \mathbb{R}$

$$df(x)\,dx_1 \to g(x)$$

$$d(df(x)dx_1)\,dx_2 =$$

$$\langle \nabla^2 f(x)\,dx_1, dx_2 \rangle$$

$$J_{\nabla f} = (\nabla^2 f)^T \left| \frac{\partial^2 f}{\partial x_i \partial x_j} \right|$$

Example

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

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

$$x \to \begin{pmatrix} \frac{\exp(x_1)}{\Sigma \exp} \\ \vdots \\ \frac{\exp(x_i)}{\Sigma \ldots} \end{pmatrix}$$

$$2k \quad \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \end{pmatrix}$$

$$J_S = \frac{\partial S_i}{\partial x_j}$$

$i \neq j$

$$S_i = \frac{\exp(x_i)}{\sum_{k=1}^{n} \exp(x_k)}$$

$$\frac{\partial S_i}{\partial x_j} = -\frac{\exp(x_i)}{\left(\sum_{k=1}^{n} \exp(x_k)\right)^2} \frac{\partial}{\partial x_j}\left(\sum_{k=1}^{n} \exp(x_k)\right)$$

$$= -\frac{\exp(x_i)\,\exp(x_j)}{\left(\sum_{k=1}^{n} \exp(x_k)\right)^2} = -S_i S_j$$

$i = j$

$$\frac{\partial S_j}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\frac{\exp(x_j)}{\sum_{k=1}^{n} \exp(x_k)}\right) =$$

$$\frac{\exp(x_j)\left(\sum_{k=1}^{n}\exp(x_k)\right) - \exp(x_j)\frac{\partial}{\partial x_j}\left(\sum\right)}{\left(\sum_{k=1}^{n}\exp(x_k)\right)^2}$$

$$= \frac{\exp(x_j)}{\sum_{k=1}^{n}\exp(x_k)} - \frac{\exp(x_j)\exp(x_j)}{\left(\sum_{k=1}^{n}\exp(x_k)\right)^2} =$$

$$\Downarrow$$

$$\Rightarrow \quad S_j' = S_j - S_j = \underbrace{S_j(1-S_j)}$$

$$(J_S)_{ij} = \begin{cases} -S_i S_j & , i \neq j \\ S_i(1-S_j) & , i = j \end{cases}$$

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

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.

Example

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

$$f(X) = \ln(\overbrace{\det(X)}^{>0})$$

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

$$df(X) = d\ln(\det(X)) = \frac{1}{\det(X)}\, d\det(X) =$$

$$d\det(X) = \det(X)\, Tr(X^{-1} dX) \Rightarrow$$

$$\Leftrightarrow \frac{\det(X)}{\det(X)} Tr(X^{-1} \underbrace{dX}) = \langle (X^{-1})^T, dX \rangle$$

$$\langle A, B \rangle = Tr(A^T B)$$

$$= \langle (X^{-1})^T, dX \rangle = \langle X^{-T}, dX \rangle \Leftrightarrow$$

$$(X^{-1})^T = (X^T)^{-1} = X^{-T}$$

$$\Leftrightarrow \langle \nabla f(X), dX \rangle \qquad \nabla f(X) = X^{-T}$$

$$d \langle \nabla f(X), dX_1 \rangle = d \langle X^{-T}, dX_1 \rangle =$$

$$= \langle -X^{-T} dX_2 X^{-T}, dX_1 \rangle$$

$$dX^{-1} = -X^{-1} dX X^{-1}$$

$$I = X^{-1} X \Rightarrow 0 = (dX^{-1}) X + X^{-1} dX$$

$$\underbrace{\langle -X^{-T} dX_2 X^{-T}, dX_1 \rangle}$$

$$\langle \nabla^2 f\, dX_1, dX_2 \rangle$$

$$\langle \mathbb{R}^{n\times n}, \mathbb{R}^{n\times n} \rangle$$

$$f : S^n_{++} \to \mathbb{R}$$

$$d^2 f(X)[\underline{H}, \underline{H}] \geq 0 \qquad \frac{(\nabla^2 f)_{ij}}{X \quad n}$$

$$B \quad \mathbb{R}^n \to \mathbb{R} \qquad \langle \nabla^2 f(X)\, \underline{dh}, \underline{dh} \rangle \geq 0$$

$$\langle -X^{-T} H X^{-T}, H \rangle =$$

$$\forall A \in S^n \qquad A = S \Lambda S^T \qquad S^T S =$$

$$SS^T = I$$

$$\Lambda = \mathrm{diag}(\lambda(A))$$

$$\boxed{X^{-1} = S\Lambda S^T}$$

$$\Lambda = \mathrm{diag}(\lambda(X^{-1}))$$

$\lambda(X^{-1})$ – собств. числа $= \dfrac{1}{\lambda(X)}$

$$X = S\Lambda^{-1}S^T$$

$$X^{-1}X = S\Lambda \underbrace{S^T S}_{I} \Lambda^{-1} S^T = I$$

$$\underline{X^{-\frac{1}{2}}} := S\Lambda^{\frac{1}{2}}S^T \qquad X^{-\frac{1}{2}}X^{-\frac{1}{2}} = X^{-1}$$

$$\langle -X^{-1}HX^{-1}, H\rangle =$$

$$\langle -X^{-\frac{1}{2}}X^{-\frac{1}{2}}HX^{-\frac{1}{2}}X^{-\frac{1}{2}}, H\rangle$$

$$= -\mathrm{Tr}(X^{-\frac{1}{2}}X^{-\frac{1}{2}}HX^{-\frac{1}{2}}X^{-\frac{1}{2}}\ H)$$

$$=$$

$$= \mathrm{Tr}(\underbrace{X^{-\frac{1}{2}}HX^{-\frac{1}{2}}}\ \underbrace{X^{-\frac{1}{2}}HX^{-\frac{1}{2}}})$$

$$= \langle \underline{X^{-\frac{1}{2}}HX^{-\frac{1}{2}}}, \underline{X^{-\frac{1}{2}}HX^{-\frac{1}{2}}}\rangle =$$

$$-\| X^{-\frac{1}{2}} H \cancel{X} X^{-\frac{1}{2}} \|_F^2 \leq 0$$

$\forall X \in S^n_{++}$

$\forall H \in S^n$

Example

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

$$f(X) = \mathrm{Tr}(AX^\top BX^{-1}), \quad A, B, X \in \mathbb{R}^{n\times n}.$$

$\det(X) \neq 0$

$$f(X) = \mathrm{Tr}(AX^T B X^{-1})$$

$$df = \mathrm{Tr}(d(AX^T B X^{-1}))$$

$$= \mathrm{Tr}(A\,d(X^T) B X^{-1} + AX^T B\, d(X^{-1}))$$

$$= \mathrm{Tr}(A(dX)^T B X^{-1}) \to dX$$

$$+ \mathrm{Tr}(-AX^T B X^{-1} dX X^{-1}) \to dX$$

$$\mathrm{Tr}(A(dX)^T B X^{-1}) = \mathrm{Tr}((A(dX)^T B X^{-1})^T)$$

$$= \mathrm{Tr}(X^{-T} B^T dX A^T) =$$

$$\mathrm{Tr}(A^T X^{-T} B^T dX) =$$

$$= \langle (A^T X^{-T} B^T)^T, dX \rangle$$

$$= \langle B X^{-1} A, dX \rangle$$

$$\mathrm{Tr}(-AX^T B X^{-1} dX X^{-1}) =$$

$$\mathrm{Tr}(-X^{-1} A X^T B X^{-1} dX) =$$

$$\langle (-X^{-1} A X^T B X^{-1})^T, dX \rangle =$$

$$= \langle -X^{-T} B^T X A^T X^{-T}, dX \rangle$$

$$df = \langle \underbrace{B X^{-1} A - X^{-T} B^T X A^T X^{-T}}_{\nabla f(X)}, H \rangle$$

$$\operatorname{Tr}(A^T B) = \langle A, B \rangle$$

$$\operatorname{Tr}(A) = \operatorname{Tr}(A^T)$$

$$\operatorname{Tr}(AB) = \operatorname{Tr}(BA)$$

Example

Выразите первую $\phi'(\alpha)$ и вторую производную $\phi''(\alpha)$ функции $\phi : \mathbb{R} \to \mathbb{R}$:

$$\phi(\alpha) = f(x + \alpha p), \quad x, p \in \mathbb{R}^n, \text{ const}$$

через градиент ∇f и гессиан $\nabla^2 f$ функции $f : \mathbb{R}^n \to \mathbb{R}$.

$$d\,\varphi(\alpha) = d f(x+\alpha p) = \langle \nabla f(x+\alpha p), d(x+\alpha p) \rangle$$

$$= \langle \nabla f(x+\alpha p), p\, d\alpha \rangle =$$

$$\underbrace{\langle \nabla f(x+\alpha p), p \rangle}_{\varphi'(\alpha)} \boxed{d\alpha} = \varphi'(\alpha)\, d\alpha$$

$$d\varphi'(\alpha) = d\langle \nabla f(x+\alpha p), p \rangle =$$

$$d\nabla f(z) = J_{\nabla f}\, dz = (\nabla^2 f)^T dz$$

$$d\varphi'(\alpha) = \langle (\nabla^2 f(x+\alpha p))^T d(x+\alpha p), p \rangle$$

$$= \langle (\nabla^2 f(x+\alpha p))^T p\, d\alpha, p \rangle =$$

$$\underbrace{\langle (\nabla^2 f(x+\alpha p))^T p, p \rangle}_{\varphi''(\alpha)\, d\varphi} d\alpha =$$

Поэлементные функции

Example

Найдите градмент $\nabla f(x)$ и гессиан $\nabla^2 f(x)$ функции $f : \mathbb{R}^n \to \mathbb{R}$:

$$f = h(g(x)),$$

где $g : \mathbb{R}^n \to \mathbb{R}^n$ действует поэлементно

$$g(x) = \sin(x), \quad \cos, \exp, \log$$

а функция $h : \mathbb{R}^n \to \mathbb{R}$ суммирует вектор $:= \begin{pmatrix} \sin(x_1) \\ \vdots \\ \sin(x_i) \\ \vdots \\ \sin(x_n) \end{pmatrix}$

$$h(u) = \sum_{i=1}^{n} u_i.$$

$$\| \quad \langle u, \vec{1} \rangle$$

$$f = h(g(x))$$

$$df = \underbrace{J_h}_{= (\nabla h)^T} \underbrace{J_g}\, d\mathbf{x} = \underbrace{J_f}_{= (\nabla f)^T}\, d\mathbf{x}$$

$$(\nabla h)^T J_g = (\nabla f)^T \Rightarrow$$

$$\nabla f = J_g^T (\nabla h)$$

$$h(v) = \langle \vec{1}, v \rangle \qquad dh = \langle \vec{1}, dv \rangle$$

$$\nabla h = \vec{1}$$

$$J_g = \qquad g: x \to \sin(x) \qquad \begin{pmatrix} \sin(x_1) \\ \vdots \\ \sin(x_i) \\ \vdots \\ \sin(x_n) \end{pmatrix}$$

$$\boxed{J_g = \mathrm{diag}(g'(x)^T)}$$

$$\frac{\partial g_i}{\partial x_j} = \begin{cases} 0, & i \neq j \\ \cos(x_j), & i = j \end{cases}$$

$$= \underbrace{\begin{pmatrix} \cos(x_1) & & 0 \\ & \ddots & \\ 0 & & \cos(x_n) \end{pmatrix}} = \underbrace{\mathrm{diag}}([\cos(x)]^T)$$

$$\nabla f = (J_g)^T \underbrace{\nabla h}_{\vec{1}} = \underbrace{\cos(x)}$$

$$d \nabla f = d\, \underbrace{\cos(x)} = J_{\cos(x)}\, dx =$$

$$\underbrace{\mathrm{diag}([-\sin(x)]^T)}_{J \nabla f = \nabla^2 f(x)}\, d\mathbf{x} \qquad \nabla^2 f = \begin{pmatrix} -\sin(x_1) & & \\ & \ddots & \\ & & -\sin(x_n) \end{pmatrix}$$