# Matrix-vector differentiation 2

## Mathematical Optimization

Georgiy Kormakov

CMC MSU

16 February 2024

# Differentiation definition

Let $U$ and $V$ be *finite-dimensional linear spaces with norms*.
**Examples:** $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^{n\times m}$, their Cartesian products.
Consider the function $f : X \to V, \quad X \subset U$.

# Differentiation definition

Let $U$ and $V$ be *finite-dimensional linear spaces with norms*.
**Examples:** $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^{n\times m}$, their Cartesian products.
Consider the function $f : X \to V, \quad X \subset U$.

## Differentiation

Let $x \in X$ be the inner point of $X$, and $L : U \to V$ be a linear operator. We will say that the function $f$ **is differentiable** at the point $x$ with the derivative $L$ if ...

# Differentiation definition

Let $U$ and $V$ be *finite-dimensional linear spaces with norms*.
**Examples:** $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^{n\times m}$, their Cartesian products.
Consider the function $f : X \to V, \quad X \subset U$.

## Differentiation

Let $x \in X$ be the inner point of $X$, and $L : U \to V$ be a linear operator. We will say that the function $f$ **is differentiable** at the point $x$ with the derivative $L$ if ... for all sufficiently small $h \in U$ it is true

$$\underline{f(x+h) = f(x)+L[h]+o(||h||)} \iff \lim_{h\to 0} \frac{||f(x+h) - f(x) - L[h]||}{||h||} = 0.$$

# Differential and directional derivative

## Differential

The **differential** $df(x)[h] \in V$ at the point $x \in X$ differentiability of the function $f$ and with an increment $h$ is called the vector $f'(x)[h]$.

Notation: $df(x)[h] \equiv Df(x)[h] \equiv f'(x)dx$. In practice, $h$ is removed, leaving $df(x)$, and $x$ is removed, leaving $df$ :)

# Differential and directional derivative

## Differential

The **differential** $df(x)[h] \in V$ at the point $x \in X$ differentiability of the function $f$ and with an increment $h$ is called the vector $f'(x)[h]$.

Notation: $df(x)[h] \equiv Df(x)[h] \equiv f'(x)dx$. In practice, $h$ is removed, leaving $df(x)$, and $x$ is removed, leaving $df$ :)

## Directional derivative

The derivative in the direction $h$ of the function $f$ at the point $x$ is called

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

- Partial derivative? Simply take the unit element of space

# Summary

| In \ Out | $\mathbb{R}$ | $\mathbb{R}^n$ | $\mathbb{R}^{n\times m}$ |
|---|---|---|---|
| $\mathbb{R}$ | $df(x) = f'(x)dx$<br>$f'(x)$ scalar, $dx$ scalar. | - | - |
| $\mathbb{R}^m$ | $df(x) = \langle \nabla f(x), dx \rangle$<br>$f(x)$ vector, $dx$ vector | $df(x) = J_x dx$<br>$J_x$ matrix, $dx$ vector | - |
| $\mathbb{R}^{n'\times m'}$ | $df(X) = \langle \nabla f(X), dX \rangle$<br>$\nabla f(X)$ matrix, $dX$ matrix | - | - |

# Second derivative

Let $f : U \to V$ be differentiable at each point $x \in U$. Consider the differential of the function $f$ with a fixed increment $h_1$ as a function of $x$:

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

## Second derivative

If at some point $x$ the function $g$ has a derivative, then it is called the second derivative, and the second differential has the form

$$D^2 f(x)[h_1, h_2] := D(Df[h_1])(x)[h_2].$$

What is the connection between Jacobian of $\nabla f$ and Hessian?

# Second derivative

Let $f : U \to V$ be differentiable at each point $x \in U$. Consider the differential of the function $f$ with a fixed increment $h_1$ as a function of $x$:

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

## Second derivative

If at some point $x$ the function $g$ has a derivative, then it is called the second derivative, and the second differential has the form

$$D^2 f(x)[h_1, h_2] := D(Df[h_1])(x)[h_2].$$

What is the connection between Jacobian of $\nabla f$ and Hessian?

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

# Tabular functions. The main table

| Transformation rules |
|---|
| $d(\alpha X) = \alpha dX$ |
| $d(AXB) = AdXB$ |
| $d(X+Y) = dX + dY$ |
| $d(X^T) = (dX)^T$ |
| $d(XY) = (dX)Y + X(dY)$ |
| $d\langle X, Y\rangle = \langle dX, Y\rangle + \langle X, dY\rangle$ |
| $d\left(\frac{X}{\phi}\right) = \frac{\phi dX - (d\phi)X}{\phi^2}$ |
| $d(g(f(x))) = g'(f)df(x)$ |
| $J_{g(f)} = J_g J_f \iff \frac{\partial g}{\partial x} = \frac{\partial g}{\partial f}\frac{\partial f}{\partial x}$ |

| Standard derivatives table |
|---|
| $dA = 0$ |
| $\langle A, X\rangle = \langle A, dX\rangle$ |
| $d\langle Ax, x\rangle = \langle (A + A^\top)x, dx\rangle$ |
| $d\mathrm{Tr}\,(X) = \mathrm{Tr}\,(dX)$ |
| $d(\det(X)) = \det(X)\mathrm{Tr}\left(X^{-1}dX\right)$ |
| $d(X^{-1}) = -X^{-1}(dX)X^{-1}$ |

# Tabular functions (some matrix example)

3. Let $S := \{X \in \mathbb{R}^{n\times n} : \det(X) \neq 0\}$ and the function $f : S \to S$ reverses the matrix $f(X) = X^{-1}$. For an arbitrary small increment of $H$, we calculate

$$f(X+H) - f(X) = (X+H)^{-1} - X^{-1} = (X(I_n + X^{-1}H))^{-1} - X^{-1} =$$
$$= ((I_n + X^{-1}H)^{-1} - I_n)X^{-1}.$$

## Neumann series.

Let $A \in \mathbb{R}^{n\times n}$ be a matrix such that $\|A\| < 1$, then the matrix $(I_n - A)$ is invertible and

$$(I_n - A)^{-1} = \sum_{k=0}^{\infty} A^k.$$

# Tabular functions (still some matrix example)

In our case, we can apply the Neumann series due to the smallness of $H$

$$(I_n + X^{-1}H)^{-1} = I_n - X^{-1}H + \sum_{k=2}^{\infty}(-X^{-1}H)^k.$$

# Tabular functions (still some matrix example)

In our case, we can apply the Neumann series due to the smallness of $H$

$$(I_n + X^{-1}H)^{-1} = I_n - X^{-1}H + \sum_{k=2}^{\infty}(-X^{-1}H)^k.$$

Let's estimate the norm of the last term

$$\left\|\sum_{k=2}^{\infty}(-X^{-1}H)^k\right\| \leq \sum_{k=2}^{\infty}\|(-X^{-1}H)^k\| \leq \sum_{k=2}^{\infty}\|-X^{-1}\|^k\|H\|^k =$$

$$= \frac{\|X^{-1}\|^2\|H\|^2}{1-\|X^{-1}\|\|H\|} = o(\|H\|),$$

As a result, we get the difference

$$f(X+H) - f(X) = -X^{-1}HX^{-1} + o(\|H\|),$$

in this case, the mapping $H \to -X^{-1}HX^{-1}$ is linear. That is, by definition

$$Df(X)[H] = -X^{-1}HX^{-1}.$$

# Determinant by definition

Let $S := \{X \in \mathbb{R}^{n\times n} : \det(X) \neq 0\}$ and the function $f : S \to \mathbb{R}$ equals $f(X) = \det(X)$. Evaluate the differential.

$$f(X+H) - f(X) = \det(X + H) - \det(X) = \det(X(I_n + X^{-1}H)) - \det(X)$$
$$= \det(X)(\det(I_n + X^{-1}H) - 1).$$

# Determinant by definition

Let $S := \{X \in \mathbb{R}^{n \times n} : \det(X) \neq 0\}$ and the function $f : S \to \mathbb{R}$ equals $f(X) = \det(X)$. Evaluate the differential.

$$f(X+H) - f(X) = \det(X+H) - \det(X) = \det(X(I_n + X^{-1}H)) - \det(X)$$
$$= \det(X)(\det(I_n + X^{-1}H) - 1).$$

$$\det(I_n + X^{-1}H) = \prod_{i=1}^{n} [1 + \lambda_i(X^{-1}H)] =$$

$$= 1 + \sum_{i=1}^{n} \lambda_i(X^{-1}H) + \underbrace{\left( \sum_{1 \le i \le j \le n} \lambda_i(X^{-1}H)\lambda_j(X^{-1}H) + \dots \right)}_{o(\|H\|)} =$$

$$|X| \cdot \operatorname{tr}(X^{-1}H) = \langle X^{-T}, H \rangle$$

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

# Determinant by definition

Let $S := \{X \in \mathbb{R}^{n \times n} : \det(X) \neq 0\}$ and the function $f : S \to \mathbb{R}$ equals $f(X) = \det(X)$. Evaluate the differential.

$$\begin{aligned} f(X+H) - f(X) = \det(X+H) - \det(X) &= \det(X(I_n + X^{-1}H)) - \det(X) \\ &= \det(X)(\det(I_n + X^{-1}H) - 1). \end{aligned}$$

$$\det(I_n + X^{-1}H) = \prod_{i=1}^{n} [1 + \lambda_i(X^{-1}H)] =$$

$$= 1 + \sum_{i=1}^{n} \lambda_i(X^{-1}H) + \left( \sum_{1 \leq i \leq j \leq n} \lambda_i(X^{-1}H)\lambda_j(X^{-1}H) + \dots \right) =$$

$$= 1 + \operatorname{Tr}\left(X^{-1}H\right) + o(\|H\|).$$

Then, final formula for differential...

# Determinant by definition

Let $S := \{X \in \mathbb{R}^{n\times n} : \det(X) \neq 0\}$ and the function $f : S \to \mathbb{R}$ equals $f(X) = \det(X)$. Evaluate the differential.

$$f(X+H) - f(X) = \det(X+H) - \det(X) = \det(X(I_n + X^{-1}H)) - \det(X)$$
$$= \det(X)(\det(I_n + X^{-1}H) - 1).$$

$$\det(I_n + X^{-1}H) = \prod_{i=1}^{n}[1 + \lambda_i(X^{-1}H)] =$$
$$= 1 + \textstyle\sum_{i=1}^{n}\lambda_i(X^{-1}H) + \left(\displaystyle\sum_{1\le i\le j\le n}\lambda_i(X^{-1}H)\lambda_j(X^{-1}H) + \dots\right) =$$
$$= 1 + \operatorname{Tr}\left(X^{-1}H\right) + o(\|H\|).$$

Then, final formula for differential...$d(\det(X)) = \det(X)\langle X^{-\top}, dX\rangle$

# Determinant by definition

Let $S := \{X \in \mathbb{R}^{n \times n} : \det(X) \neq 0\}$ and the function $f : S \to \mathbb{R}$ equals $f(X) = \det(X)$. Evaluate the differential.

$$f(X+H) - f(X) = \det(X+H) - \det(X) = \det(X(I_n + X^{-1}H)) - \det(X)$$
$$= \det(X)(\det(I_n + X^{-1}H) - 1).$$

$$\det(I_n + X^{-1}H) = \prod_{i=1}^{n} [1 + \lambda_i(X^{-1}H)] =$$

$$= 1 + \sum_{i=1}^{n} \lambda_i(X^{-1}H) + \left( \sum_{1 \le i \le j \le n} \lambda_i(X^{-1}H)\lambda_j(X^{-1}H) + \dots \right) =$$

$\sum x_{ij} \mathrm{Adj}(x)_{ji}$ $$= 1 + \mathrm{Tr}\left(X^{-1}H\right) + o(\|H\|).$$ $\mathrm{Adj} = |x| \cdot X^{-1}$

Then, final formula for differential... $d(\det(X)) = \det(X)\langle X^{-\top}, dX \rangle$

For non-invertible matrices — Jacobi formula: $d(\det(X)) = \left\langle \mathrm{Adj}(X)^{\top}, dX \right\rangle$

(because $\frac{\partial f}{\partial x_{ij}} = (-1)^{(i+j)} M_{ij} = \mathrm{Adj}(X)_{ji}$.)

# Frobenius norm

Find the gradient $\nabla f(X)$ and the differential $df(X)$ of function $f(X)$

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

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

# Frobenius norm

Find the gradient $\nabla f(X)$ and the differential $df(X)$ of function $f(X)$

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

where $A \in \mathbb{R}^{m \times k}$, $B \in \mathbb{R}^{m \times n}$. $\|X\|_F = \langle X, X \rangle^{\frac{1}{2}}$

$$d\|X\| = \frac{1}{2} \frac{2\langle X, dX \rangle}{\|X\|}$$

# Frobenius norm

Find the gradient $\nabla f(X)$ and the differential $df(X)$ of function $f(X)$

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

where $A \in \mathbb{R}^{m \times k}$, $B \in \mathbb{R}^{m \times n}$. $\|X\|_F = \langle X, X \rangle^{\frac{1}{2}}$

$\langle \cdot, dX \rangle$

$$d(\|X\|_F) = \left\langle \frac{X}{\|X\|_F}, dX \right\rangle$$

$$\left\langle \frac{AX-B}{\|\cdot\|}, \underbrace{d(AX-B)}_{A\,dX} \right\rangle$$

# Frobenius norm

Find the gradient $\nabla f(X)$ and the differential $df(X)$ of function $f(X)$

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

where $A \in \mathbb{R}^{m \times k}$, $B \in \mathbb{R}^{m \times n}$. $\|X\|_F = \langle X, X \rangle^{\frac{1}{2}}$

$$d(\|X\|_F) = \left\langle \frac{X}{\|X\|_F}, dX \right\rangle$$

$$df(X) = d(\|AX - B\|_F) = \left\langle \underbrace{\frac{A^\top (AX - B)}{\|AX - B\|}}_{\nabla f}, dX \right\rangle$$

# Uncomfortable Trace

Find $\nabla f(X)$, $df(X)$ of $f(X)$

$$f(X) = \text{Tr}\left(AXBX^{-1}\right), \quad X \in \mathbb{R}^{n \times n}, \det(X) \neq 0,$$

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

# Uncomfortable Trace

Find $\nabla f(X)$, $df(X)$ of $f(X)$

$$f(X) = \mathrm{Tr}\left(\underline{AXBX^{-1}}\right), \quad X \in \mathbb{R}^{n \times n}, \det(X) \neq 0,$$

$\langle A^T, - \_\_\_ - \rangle$

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

We can represent a trace as the inner product

$$f(X) = \langle I_n, AXBX^{-1} \rangle.$$

# Uncomfortable Trace

Find $\nabla f(X)$, $df(X)$ of $f(X)$

$$f(X) = \mathrm{Tr}\left(AXBX^{-1}\right), \quad X \in \mathbb{R}^{n\times n}, \det(X) \neq 0,$$

where $A, B \in \mathbb{R}^{n\times n}$.
We can represent a trace as the inner product

$\langle \cdot , dX \rangle$

$$f(X) = \langle I_n, AXBX^{-1} \rangle.$$

$$\begin{aligned}
df(X) = &\quad \langle I_n, d(AXBX^{-1})\rangle = \langle I_n, Ad(XBX^{-1})\rangle \\
= &\quad \langle I_n, A(dX)BX^{-1} + AXd(BX^{-1})\rangle \\
= &\quad \langle I_n, A(dX)BX^{-1} + AXB\cdot(-X^{-1}(dX)X^{-1})\rangle \\
= &\quad \mathrm{Tr}\left(A(dX)BX^{-1}\right) - \mathrm{Tr}\left(AXBX^{-1}(dX)X^{-1}\right)
\end{aligned}$$

# Uncomfortable Trace

Find $\nabla f(X)$, $df(X)$ of $f(X)$

$$f(X) = \text{Tr}\left(AXBX^{-1}\right), \quad X \in \mathbb{R}^{n \times n}, \det(X) \neq 0,$$

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

We can represent a trace as the inner product

$$f(X) = \langle I_n, AXBX^{-1} \rangle.$$

$$\begin{aligned}
df(X) = & \quad \langle I_n, d(AXBX^{-1}) \rangle = \langle I_n, Ad(XBX^{-1}) \rangle \\
= & \quad \langle I_n, A(dX)BX^{-1} + AXd(BX^{-1}) \rangle \\
= & \quad \langle I_n, A(dX)BX^{-1} + AXB \cdot (-X^{-1}(dX)X^{-1}) \rangle \\
= & \quad \text{Tr}\left(A(dX)BX^{-1}\right) - \text{Tr}\left(AXBX^{-1}(dX)X^{-1}\right) \\
= & \quad \text{Tr}\left(BX^{-1}A(dX)\right) - \text{Tr}\left(X^{-1}AXBX^{-1}(dX)\right) \\
= & \quad \langle A^\top X^{-\top} B^\top - X^{-\top} B^\top X^\top A^\top X^{-\top}, dX \rangle.
\end{aligned}$$

# More simple Trace (try by yourself)

Find $\nabla f(X)$, $df(X)$ of $f(X)$

$$f(X) = \mathrm{Tr}\left(AXBX^{-1}\right), \quad X \in \mathbb{R}^{n \times n}, \det(X) \neq 0,$$

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

# Log Det. Concavity

Find $\nabla f(X)$, $df(X)$, $d^2 f(X)$ of $f(X)$

$$f(X) = \ln(\det(X)), \quad X \in \mathbb{S}^n_{++}$$

# Log Det. Concavity

Find $\nabla f(X)$, $df(X)$, $d^2 f(X)$ of $f(X)$

$$f(X) = \ln(\det(X)), \quad X \in \mathbb{S}^n_{++}$$

$$df(X) = d(\ln\det(X)) = \frac{d(\det(X))}{\det(X)} = \frac{\det(X)\langle X^{-\top}, dX\rangle}{\det(X)} \overset{X \in \mathbb{S}^n_{++}}{=} \langle X^{-1}, dX\rangle.$$

# Log Det. Concavity

Find $\nabla f(X)$, $df(X)$, $d^2 f(X)$ of $f(X)$

$$f(X) = \ln(\det(X)), \quad X \in \mathbb{S}^n_{++}$$

$$df(X) = d(\ln\det(X)) = \frac{d(\det(X))}{\det(X)} = \frac{\det(X)\langle X^{-\top}, dX\rangle}{\det(X)} \overset{X\in\mathbb{S}^n_{++}}{=} \langle X^{-1}, dX\rangle.$$

$$d^2 f(X) = \langle d(X^{-1}), dX_1\rangle = -\langle X^{-1}(dX)X^{-1}, dX_1\rangle.$$

# Log Det. Concavity

Find $\nabla f(X)$, $df(X)$, $d^2 f(X)$ of $f(X)$

$$f(X) = \ln(\det(X)), \quad X \in \mathbb{S}^n_{++}$$

$$df(X) = d(\ln\det(X)) = \frac{d(\det(X))}{\det(X)} = \frac{\det(X)\langle X^{-\top}, dX\rangle}{\det(X)} \overset{X\in\mathbb{S}^n_{++}}{=} \langle X^{-1}, dX\rangle.$$

$$d^2 f(X) = \langle d(X^{-1}), dX_1\rangle = -\langle X^{-1}(dX)X^{-1}, dX_1\rangle.$$

$$d^2 f(X)[H, H] = -\langle X^{-1}HX^{-1}, H\rangle = -\mathrm{Tr}\left(X^{-1}HX^{-1}H\right) =$$

$X^{-\frac{1}{2}} =$

# Log Det. Concavity

Find $\nabla f(X)$, $df(X)$, $d^2 f(X)$ of $f(X)$

$$f(X) = \ln(\det(X)), \quad X \in \mathbb{S}^n_{++}$$

$$df(X) = d(\ln\det(X)) = \frac{d(\det(X))}{\det(X)} = \frac{\det(X)\langle X^{-\top}, dX\rangle}{\det(X)} \overset{X\in\mathbb{S}^n_{++}}{=} \langle X^{-1}, dX\rangle.$$

$$d^2 f(X) = \langle d(X^{-1}), dX_1\rangle = -\langle X^{-1}(dX)X^{-1}, dX_1\rangle.$$

$$d^2 f(X)[H, H] = -\langle X^{-1}HX^{-1}, H\rangle = -\mathrm{Tr}\left(X^{-1}HX^{-1}H\right) =$$
$$= -\mathrm{Tr}\left(X^{-1/2}X^{-1/2}HX^{-1/2}X^{-1/2}H\right) =$$
$$= -\mathrm{Tr}\left(X^{-1/2}HX^{-1/2}X^{-1/2}HX^{-1/2}\right) =$$
$$= -\langle X^{-1/2}HX^{-1/2}, X^{-1/2}HX^{-1/2}\rangle = -\|X^{-1/2}HX^{-1/2}\|_F^2 \leq 0.$$

# Det outside)

Find $\nabla f(X)$, $df(X)$ of $f(X)$

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

where $A, X, B$ have proper dimensions, and $AX^{-1}B$ is invertible.