Matrix representation of $D\boldsymbol{f}(\boldsymbol{x}_0)$, Partial derivatives

In this section we study the matrix representation of $D\boldsymbol{f}(\boldsymbol{x}_0)$.

Jacobian matrix

Definition 1. *Let* $\boldsymbol{f}: \mathbb{R}^N \mapsto \mathbb{R}^M$ *be differentiable at* $\boldsymbol{x}_0 \in \mathbb{R}^N$*. Then the matrix representation of its derivative* $D\boldsymbol{f}(\boldsymbol{x}_0)$ *is called the Jacobian matrix of* $\boldsymbol{f}$ *at* $\boldsymbol{x}_0$*.*

Remark 2. There doesn't seem to be a universally accepted notation for this matrix. We will use the notation $\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)(\boldsymbol{x}_0)$ to denote this $M \times N$ matrix.

Example 3. Let $\boldsymbol{f}$ be linear with matrix representation A. Then at any $\boldsymbol{x}_0 \in \mathbb{R}^N$, the Jacobian matrix is A.

Proof. We have seen that when $\boldsymbol{f}$ is a linear function, its derivative $D\boldsymbol{f}(\boldsymbol{x}_0) = \boldsymbol{f}$. Therefore they share the matrix representation. □

We have seen that, even for very simple functions, establishing its differentiability is quite nontrivial. One particular source of this difficulty comes from not knowing $D\boldsymbol{f}(\boldsymbol{x}_0)$. Therefore it would be convenient to have a method obtaining possible candidates for the derivative without establishing differentiability first.

To see how this could be done, we notice that, when $\boldsymbol{f}$ is linear, we have

$$\begin{pmatrix} f_1(x_1, \ldots, x_N) \\ \vdots \\ f_M(x_1, \ldots, x_N) \end{pmatrix} = \begin{pmatrix} a_{11} x_1 + \cdots + a_{1N} x_N \\ \vdots \\ a_{M1} x_1 + \cdots + a_{MN} x_N \end{pmatrix} \tag{1}$$

and each a_{ij} can be obtained as follows: Fix x_k for all $k \neq j$ and treat $f_i(\cdots, x_j, \cdots)$ as a function of x_j alone. Then

$$f_i(\cdots, x_j, \cdots) = a_{ij} x_j + [\text{terms not involving } x_j] \tag{2}$$

and we would have

$$a_{ij} = f_i'. \tag{3}$$

This leads to the notion of partial derivatives.

Partial derivatives

Theorem 4. *Let* $\boldsymbol{f}: \mathbb{R}^N \mapsto \mathbb{R}^M$ *be differentiable at* $\boldsymbol{x}_0 \in \mathbb{R}^N$*. Let* A *be its Jacobian matrix there. Then the limits*

$$\lim_{h \to 0, h \neq 0} \frac{f_i(\boldsymbol{x}_0 + h\,\boldsymbol{e}_j) - f_i(\boldsymbol{x}_0)}{h} \tag{4}$$

exist for all $i = 1, \ldots, M$ *and* $j = 1, \ldots, N$*, and equals* a_{ij}*.*

Proof. Since $\boldsymbol{f}$ is differentiable at $\boldsymbol{x}_0$, we have

$$\lim_{\boldsymbol{x} \longrightarrow \boldsymbol{x}_0} \frac{\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}_0) - D\boldsymbol{f}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)\|}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} = 0. \tag{5}$$

Taking $\boldsymbol{x} = \boldsymbol{x}_0 + h\, \boldsymbol{e}_j$ we have

$$\lim_{h \to 0, h \neq 0} \frac{\| \boldsymbol{f}(\boldsymbol{x}_0 + h\, \boldsymbol{e}_j) - \boldsymbol{f}(\boldsymbol{x}_0) - h\, (A\, \boldsymbol{e}_j) \|}{|h|} = 0. \tag{6}$$

This implies

$$\lim_{h \to 0, h \neq 0} \frac{| f_i(\boldsymbol{x}_0 + h\, \boldsymbol{e}_j) - f_i(\boldsymbol{x}_0) - a_{ij}\, h |}{h} = 0 \tag{7}$$

or equivalently

$$\lim_{h \to 0, h \neq 0} \frac{f_i(\boldsymbol{x}_0 + h\, \boldsymbol{e}_j) - f_i(\boldsymbol{x}_0)}{h} = a_{ij}. \tag{8}$$

Thus ends the proof. □

Definition 5. (Partial derivatives) *Let $\boldsymbol{f}: \mathbb{R}^N \mapsto \mathbb{R}^M$ be differentiable at $\boldsymbol{x}_0$. Its j-th partial derivative is the vector*

$$\frac{\partial \boldsymbol{f}}{\partial x_j} := \lim_{h \to 0, h \neq 0} \frac{\boldsymbol{f}(\boldsymbol{x}_0 + h\, \boldsymbol{e}_j) - \boldsymbol{f}(\boldsymbol{x}_0)}{h}. \tag{9}$$

Thus we see that the Jacobian matrix consists of partial derivatives:

$$A = \left(\frac{\partial f_i}{\partial x_j} \right). \tag{10}$$

Example 6. Let $f: \mathbb{R}^2 \mapsto \mathbb{R}$ be given by

$$f(x, y) := a\, x^2 + b\, x\, y + c\, y^2. \tag{11}$$

Find the partial derivatives.

Solution. We have

$$\frac{\partial f}{\partial x}(x, y) = 2\, a\, x + b\, y; \qquad \frac{\partial f}{\partial y}(x, y) = 2\, c\, y + b\, x. \tag{12}$$

The pros and cons of partial derivatives

Why don't we simply use existence of partial derivatives for differentiability for $\boldsymbol{f}: \mathbb{R}^N \mapsto \mathbb{R}^M$?

Example 7. Let $f: \mathbb{R}^2 \mapsto \mathbb{R}$ be such that $f(x, y) = \begin{cases} 1 & x = 0 \text{ or } y = 0 \\ 0 & \text{elsewhere.} \end{cases}$. Then $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ exists at $(0, 0)$ but obviously f is not even continuous there.

Theorem 8. *Let $\boldsymbol{f}: \mathbb{R}^N \mapsto \mathbb{R}^M$ and $\boldsymbol{x}_0 \in \mathbb{R}^N$. Further assume there is $r > 0$ such that all the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exists in $B(\boldsymbol{x}_0, r)$ and is continuous at $\boldsymbol{x}_0$. Then $\boldsymbol{f}$ is differentiable at $\boldsymbol{x}_0$ with Jacobian matrix $\left(\frac{\partial f_i}{\partial x_j} \right)$.*

Proof. Let $\boldsymbol{x} \in B(\boldsymbol{x}_0, r)$. Denote

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \text{ and } \boldsymbol{x}_0 = \begin{pmatrix} x_{01} \\ \vdots \\ x_{0N} \end{pmatrix}. \tag{13}$$

Denoting $\boldsymbol{x}$ by $\boldsymbol{x}_N$. Now define $N-1$ new vectors/points:

$$\boldsymbol{x}_1 := \begin{pmatrix} x_1 \\ x_{02} \\ \vdots \\ x_{0N} \end{pmatrix}, \quad \boldsymbol{x}_2 := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{0N} \end{pmatrix}, \quad \cdots \quad , \boldsymbol{x}_{N-1} := \begin{pmatrix} x_1 \\ \vdots \\ x_{N-1} \\ x_{0N} \end{pmatrix}. \tag{14}$$

We claim that all $\boldsymbol{x}_j \in B(\boldsymbol{x}_0, r)$ and furthermore the line segments connecting each pair of points $\boldsymbol{x}_{j-1}, \boldsymbol{x}_j$ lies inside $B(\boldsymbol{x}_0, r)$ for all $j = 1, \ldots, N$.

Now consider the function

$$g_{ij}(t) := f_i(x_1, \ldots x_{j-1}, t, x_{j+1}, \ldots, x_{0N}). \tag{15}$$

Clearly g_{ij} is differentiable and

$$g_{ij}'(t) = \frac{\partial f_i}{\partial x_j}(x_1, \ldots x_{j-1}, t, x_{j+1}, \ldots, x_{0N}). \tag{16}$$

Now apply the single variable MVT to g_{ij} we have

$$g_{ij}(x_j) - g_{ij}(x_{0j}) = g'(\theta_{ij})\,(x_j - x_{0j}). \tag{17}$$

Writing in $\boldsymbol{f}$ this translates to

$$f_i(\boldsymbol{x}_j) - f_i(\boldsymbol{x}_{j-1}) = \frac{\partial f_i}{\partial x_j}(\boldsymbol{\xi}_{ij})\,(x_j - x_{0j}), \qquad i = 1, 2, \ldots, M; j = 1, 2, \ldots, N. \tag{18}$$

Here

$$\boldsymbol{\xi}_{ij} := \begin{pmatrix} x_1 \\ \vdots \\ x_{j-1} \\ \theta_{ij} \\ x_{0j+1} \\ \vdots \\ x_{0N} \end{pmatrix}. \tag{19}$$

Note that here each $\boldsymbol{\xi}_{ij}$ satisfies

$$\|\boldsymbol{\xi}_{ij} - \boldsymbol{x}_0\| \leqslant \|\boldsymbol{x} - \boldsymbol{x}_0\|. \tag{20}$$

Summing up, we have

$$\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}_0) = \tilde{A}\,(\boldsymbol{x} - \boldsymbol{x}_0) \tag{21}$$

where $\tilde{A}$ is the matrix with columns $\frac{\partial \boldsymbol{f}}{\partial x_j}(\boldsymbol{\xi}_j)$, that is

$$\tilde{A} = \left(\ \frac{\partial f_i}{\partial x_j}(\boldsymbol{\xi}_{ij}) \ \right). \tag{22}$$

Now letting

$$A = \left(\ \frac{\partial f_i}{\partial x_j}(\boldsymbol{x}_0) \ \right) \tag{23}$$

we have

$$\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}_0) - A(\boldsymbol{x} - \boldsymbol{x}_0) = \left(\tilde{A} - A\right)(\boldsymbol{x} - \boldsymbol{x}_0). \tag{24}$$

from which it follows that

$$\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}_0) - A(\boldsymbol{x} - \boldsymbol{x}_0)\| \leqslant \left\|\tilde{A} - A\right\|_F \|\boldsymbol{x} - \boldsymbol{x}_0\| \tag{25}$$

where

$$\|A\|_F := \left(\sum_{i,j} a_{ij}^2\right)^{1/2}. \tag{26}$$

Now continuity of partial derivatives easily gives

$$\lim_{\boldsymbol{x} \longrightarrow \boldsymbol{x}_0} \frac{\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}_0) - A(\boldsymbol{x} - \boldsymbol{x}_0)\|}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} = 0. \tag{27}$$

The details are left as exercise (see below). □

Exercise 1. Prove the claim: All $\boldsymbol{x}_j \in B(\boldsymbol{x}_0, r)$ and furthermore the line segments connecting each pair of points $\boldsymbol{x}_{j-1}, \boldsymbol{x}_j$ lies inside $B(\boldsymbol{x}_0, r)$ for all $j = 1, \ldots, N$.

Exercise 2. Prove the claim: Each $\boldsymbol{\xi}_j$ satisfies

$$\|\boldsymbol{\xi}_j - \boldsymbol{x}_0\| \leqslant \|\boldsymbol{x} - \boldsymbol{x}_0\|. \tag{28}$$

Exercise 3. Prove the estimate

$$\left\|\left(\tilde{A} - A\right)(\boldsymbol{x} - \boldsymbol{x}_0)\right\| \leqslant \left\|\tilde{A} - A\right\|_F \|\boldsymbol{x} - \boldsymbol{x}_0\| \tag{29}$$

where

$$\|A\|_F := \left(\sum_{i,j} a_{ij}^2\right)^{1/2}. \tag{30}$$

Exercise 4. Provide the details for the last part of the proof.

Problem 1. Critique the following idea trying to simplify the proof:

> Define
>
> $$g_i(t) := f_i(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0)). \tag{31}$$
>
> Now multivariable chain rule together with single variable MVT gives
>
> $$f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}_0) = g_i(1) - g_i(0) = \sum_{j=1}^{N} \frac{\partial f_i}{\partial x_j}(\boldsymbol{\xi}_{\boldsymbol{i}})(x_j - x_{0j}). \tag{32}$$

Is it correct?