Definition: *Suppose* $f : D \to \mathbb{R}$, *where* $D \subseteq \mathbb{R}^N$. *Then* f *is differentiable at* $x_o \in D$ *if there exists* $\lambda \in \mathbb{R}^N$ *such that* $f(x_o + h) = f(x_o) + \lambda \bullet h + o(h)$ *for all* $h \in \mathbb{R}^N$ *such that* $x_o + h \in D$.

Example: Let $f(x) := \|x\|^2 = x \bullet x$ and $D := \mathbb{R}^N$. Then $f(x_o + h) = f(x_o) + 2x_o \bullet h + \|h\|^2$. Thus with $\lambda := 2x_o$, we have $\|h\|^2 = o(h)$ so f is differentiable at every $x_o \in D$. In particular, $\lambda = 0$ when $x_o = 0$.

Example: Let $f(x) := \|x\| = \sqrt{x \bullet x}$ and $D := \mathbb{R}^N$. If $x_o \neq 0$ then

$$\begin{aligned}
f(x_o + h) - f(x_o) &= \frac{(x_o + h) \bullet (x_o + h) - x_o \bullet x_o}{\sqrt{(x_o + h) \bullet (x_o + h)} + \sqrt{x_o \bullet x_o}} \\
&= \frac{2x_o \bullet h + \|h\|^2}{\sqrt{(x_o + h) \bullet (x_o + h)} + \sqrt{x_o \bullet x_o}} \\
&= \frac{2x_o \bullet h + \|h\|^2}{\|x_o + h\| + \|x_o\|} \\
&= (2x_o \bullet h + \|h\|^2) \left(\frac{1}{2\|x_o\|} - \frac{1}{2\|x_o\|} + \frac{1}{\|x_o + h\| + \|x_o\|} \right) \\
&= \frac{2x_o \bullet h + \|h\|^2}{2\|x_o\|} - (2x_o \bullet h + \|h\|^2) \left(\frac{\|x_o + h\| - \|x_o\|}{2\|x_o\|(\|x_o + h\| + \|x_o\|)} \right) \\
&= \frac{x_o \bullet h}{\|x_o\|} + \frac{\|h\|^2}{2\|x_o\|} - (2x_o \bullet h + \|h\|^2) \left(\frac{\|x_o + h\| - \|x_o\|}{2\|x_o\|(\|x_o + h\| + \|x_o\|)} \right) \\
&= \frac{x_o}{\|x_o\|} \bullet h + \frac{\|h\|^2}{2\|x_o\|} - (2x_o \bullet h + \|h\|^2) \left(\frac{\|x_o + h\| - \|x_o\|}{2\|x_o\|(\|x_o + h\| + \|x_o\|)} \right).
\end{aligned}$$

We let $\lambda := \dfrac{x_o}{\|x_o\|}$ and seek to show that the rest on the right-hand side is $o(h)$.

Exercise: Prove that $o(h) + o(h) = o(h)$ (!).

We look at the third term on the right, divided by $\|h\|$:

$$\left| -(2x_o \bullet h + \|h\|^2) \left(\frac{\|x_o + h\| - \|x_o\|}{2\|x_o\|(\|x_o + h\| + \|x_o\|)} \right) \right| \frac{1}{\|h\|} \leq \frac{(2\|x_o\|\|h\| + \|h\|^2)\|h\|}{2\|x_o\|^2\|h\|} = \frac{(2\|x_o\|\|h\| + \|h\|^2)}{2\|x_o\|^2} \to 0.$$

Exercise: Explain the $\leq$. Steps are combined. Stating appropriate Theorems is how to do the explanation.

We have shown that $\|x\|$ is differentiable at x_o if $x_o \neq 0$. Next we show that $\|x\|$ is *not* differentiable at $x_o = 0$.

Suppose it were. Then $\|h\| = 0 + \lambda \bullet h + o(h)$. We cannot have $\lambda = 0$ because then $\|h\| = o(h)$, which leads to the contradiction $1 = 0$. We choose $h_n = -\lambda/n$ $(\neq 0)$. Then

$$\frac{1}{n}\|\lambda\| = \|h_n\| = \lambda \bullet h_n + o(h_n) = -\frac{1}{n}\|\lambda\|^2 + o(-\lambda/n), \quad \text{or} \quad \|\lambda\| = -\|\lambda\|^2 + n \cdot o(\lambda/n) \to -\|\lambda\|^2$$

as $n \to \infty$. This is a contradiction: a positive number is not negative.

Exercise: Verify that $n \cdot o(\lambda/n) \to 0$ as $n \to \infty$.

The example $f(x) = \|x\|$ is an example of a continuous function that is not everywhere differentiable.

To extend the idea of differentiability to vector-valued functions we really proceed by differentiating one component at a time. This works, whereas the idea of doing differentiation one variable at a time does *not* work. However, there is some utility to using the language of Linear Algebra in this context. The main concept we use is "linear transformation," the same as "linear mapping" and "linear function." By this we mean a function $L : V \to W$, where V and W are vector spaces (linear spaces) with the same scalars that has this property: L "preserves" the linear operations, namely addition and scalar multiplication. In symbols, we require that $L(ax+by) = aL(x)+bL(y)$ for all scalars a, b and all vectors x and y in V. We have seen several examples of linear mappings. To give an example, we need to specify the vector spaces and the linear mapping. We may have to do some work to prove that the mapping is indeed linear.

Examples: We let V be the collection of all sequences that have a limit. To verify that this is a vector space we have to show that we have good operations of addition and scalar multiplication. Thus if x and y are two sequences with limits, does $x + y$ (here $(x + y)_n := x_n + y_n$) belong to V? Yes it does, by one of the Limit Theorems. If c is a scalar, does cx belong to V? Here, $(cx)_n := cx_n$. Again yes by another Limit Theorem. Some other details need to be checked, but V is indeed a vector space.

Here are two examples of linear mappings that map V to V. We define Lx by $(Lx)_n := x_{n+1}$ This is called the "Left Shift." The other example is the "Right Shift," defined by $(Rx)_1 := 0$ and $(Rx)_n := x_{n-1}$ for $n > 1$.

Exercise: Verify that L and R are linear mappings that have domain V and range (space) V. Verify that L is onto and R is one-to-one. Calculate RL and LR.

Another example of a linear mapping from $\mathbb{R}^N$ to $\mathbb{R}$ is $x \mapsto \lambda \bullet x$, where $\lambda \in \mathbb{R}^N$.

We need a linear mapping $\Lambda : \mathbb{R}^N \to \mathbb{R}^M$. This we can get by matrix multiplication: $\Lambda(x) := Ax$, where A is an $N \times M$ matrix with entries a_{ij}. By a very strong convention, i in this context always refers to a *row* of the matrix A and j always refers to a *column* of the matrix A. For example, in the matrix $A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$, $a_{12} = 2$ and $a_{23} = 6$. The matrix obtained by interchanging the rows and columns of an $M \times N$ matrix A is denoted A^T and called the *transpose of* A. For our example we have $A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$. The first column of A is $\begin{pmatrix} 1 \\ 4 \end{pmatrix} = Ae_1$, and the second row of A is $\begin{pmatrix} 4 & 5 & 6 \end{pmatrix} = e_2^T A$. Note that (again, strong convention) when e_j or e_i^T appears in a matrix product, it is assumed to have the size appropriate to the matrix product.

Definition: *Suppose $f : D \to \mathbb{R}^M$, where $D \subseteq \mathbb{R}^N$. Then f is differentiable at $x_o \in D$ if there exists a linear mapping $\Lambda : \mathbb{R}^N \to \mathbb{R}^M$ such that $f(x_o + h) = f(x_o) + \Lambda(h) + o(h)$ for all $h \in \mathbb{R}^N$ such that $x_o + h \in D$. If f is differentiable at $x_o \in D$ then Λ is given by the matrix with entries $\left.\dfrac{\partial f_i}{\partial x_j}\right|_{x=x_o}$.*

We note that this Definition includes the definition when $M = 1$, so we may as well drop the other one and just use this one.

The quantity $\Lambda(h)$ is usually written Λh, to connote a matrix product. Now that you know that the existence of the partial derivatives does not necessarily mean that the function is differentiable, we can simply refer to Λ as $f'(x_o)$, so the definition can also read:

$$f(x_o + h) = f(x_o) + f'(x_o)h + o(h), \quad \text{where} \quad f'(x_o) = \left(\left.\frac{\partial f_i}{\partial x_j}\right|_{x=x_o} \right)_{M \times N}.$$

Notice that the i-th row of $f'(x_o)$ is the "derivative" of f_i, the real-valued function that is the i-th component of the vector-valued function f.

Exercise: Calculate the "derivative" of $f(x) := Ax$, where A is an $N \times M$ matrix with entries a_{ij}. Prove that f is differentiable at every $x_o \in \mathbb{R}^N$.

Exercise: Calculate the "derivative" of $f(x, y) := xy^T$, where x and y are in $\mathbb{R}^N$. This is tricky! What is the domain of f? The range of f? You may need to redefine "differentiable" so that $\mathbb{R}^N$ and $\mathbb{R}^M$ are replaced by "finite-dimensional vector space with real scalars!"