Corollary. Under the hypotheses of the Theorem, the forward image of a relatively open set in D is open in $\mathbb{E}^m$.

10.1 Linear Transformations and Norms.

A. Brief review of linear algebra.

1. Definition. A *linear transformation* $L: \mathbb{E}^n \to \mathbb{E}^m$ is a function with the property that for every $\mathbb{x}, \mathbb{y} \in \mathbb{E}^n$, and scalars α, β,

$$L(\alpha\mathbb{x} + \beta\mathbb{y}) = \alpha L(\mathbb{x}) + \beta L(\mathbb{y}).$$

We denote the collection of all such linear transformations by $\mathcal{L}(\mathbb{E}^n, \mathbb{E}^m)$.

2. Definition. A *basis,* $\mathcal{B}$ in $\mathbb{E}^n$ is a collection of vectors $\mathcal{B} = \{\mathbb{b}_1, \mathbb{b}_2, \dots, \mathbb{b}_n\}$ that is linearly independent. If $\mathcal{B}$ is a basis then every vector $\mathbb{x} \in \mathbb{E}^n$ can be written uniquely as

$$\mathbb{x} = \sum_{j=1}^{n} \alpha_j\, \mathbb{b}_j$$

where the coefficients $\alpha_j = \mathbb{x} \cdot \mathbb{v}_j$ where $\{\mathbb{v}_j\}_{j=1}^n$ is the unique collection of vectors *biorthogonal* to $\mathcal{B}$, that is, such that $\mathbb{b}_j \cdot \mathbb{v}_k = 1$ if $j = k$ and 0 otherwise.

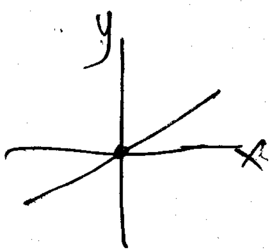
A basis $\mathcal{B}$ is *orthonormal* if $\mathbb{b}_j = \mathbb{v}_j$ for all j.

$$B = \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_n \\ | & | & & | \end{bmatrix} \qquad V = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$$

$$V = (B^{-1})^T$$

$$\begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots 1 \end{bmatrix} = \begin{bmatrix} — & v_1 & — \\ — & v_2 & — \\ & \vdots & \\ — & v_n & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_n \\ | & | & & | \end{bmatrix}$$

$\mathbb{R} \to \mathbb{R}$



3. <u>Remark</u>. (a) $L: \mathbb{E}^1 \to \mathbb{E}^1$ is linear if and only if $L(x) = \alpha\, x$ for some real number α.

(b) $L: \mathbb{E}^1 \to \mathbb{E}^m$ is linear if and only if $L(x) = (\alpha_1 x, \alpha_2 x, \ldots, \alpha_m x) = \mathbb{a}\, x$ for some fixed $\mathbb{a} \in \mathbb{E}^m$.

L represents a parametrized line, with direction $\vec{a}$

(c) $L: \mathbb{E}^n \to \mathbb{E}^1$ is linear if and only if $L(\mathbb{x}) = \mathbb{a} \cdot \mathbb{x} = \sum_{j=1}^{n} \alpha_j\, x_j$. In this case, $\alpha_j = L(\mathbb{e}_j)$ where $\mathbb{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ is the j^{th} element of the *standard basis* $\{\mathbb{e}_j\}_{j=1}^{n}$.

$\vec{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{E}^n$ (1 in the jth position)

$\vec{x} \in \mathbb{E}^n$: $\vec{x} = \sum_{j=1}^{n} x_j \vec{e}_j$

$\vec{x} = (x_1, \ldots, x_n)$. $L(\vec{x}) = \sum_{j=1}^{n} x_j \underbrace{L(\vec{e}_j)}_{\alpha_j}$

4. Theorem. Every linear transformation $L\colon \mathbb{E}^n \to \mathbb{E}^m$ can be represented as an $m \times n$ matrix.

Let $\mathcal{B} = \{\vec{b}_1, \dots, \vec{b}_n\}$ be a basis for $\mathbb{E}^n$

and $\mathcal{C} = \{\vec{c}_1, \dots, \vec{c}_m\}$ a basis for $\mathbb{E}^m$

Given $\vec{x} \in \mathbb{E}^n$, $\vec{x} = \sum_{j=1}^{n} \alpha_j \vec{b}_j$

$$L(\vec{x}) = L\left(\sum_{j=1}^{n} \alpha_j \vec{b}_j\right) = \sum_{j=1}^{n} \alpha_j \underbrace{L(\vec{b}_j)}_{\in \mathbb{E}^m}$$

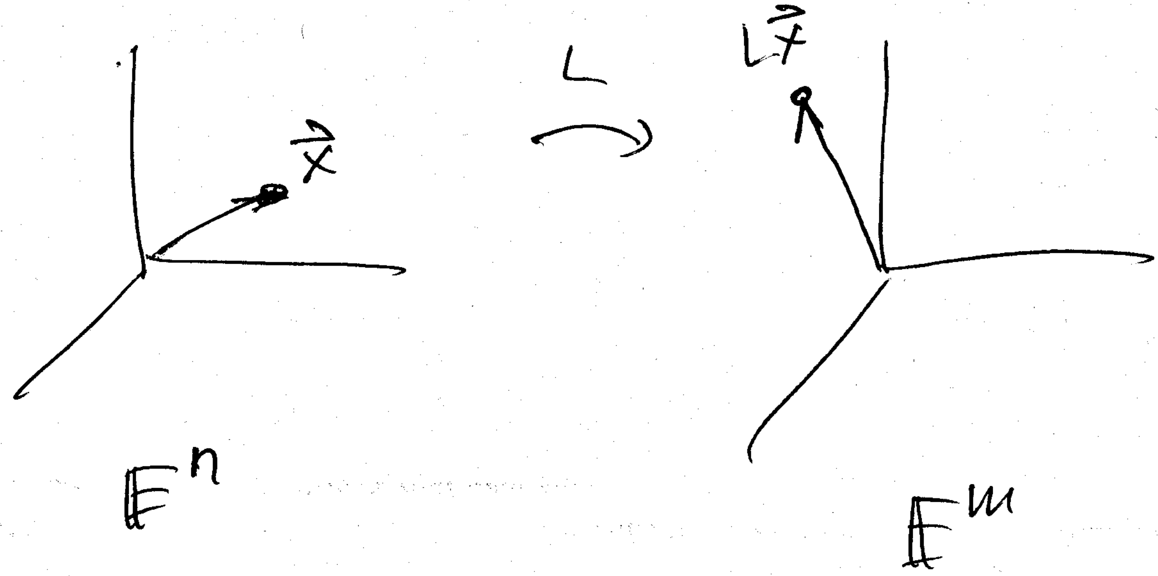
$\therefore L(\vec{b}_j) = \sum_{k=1}^{m} t_{j,k} \vec{c}_k$

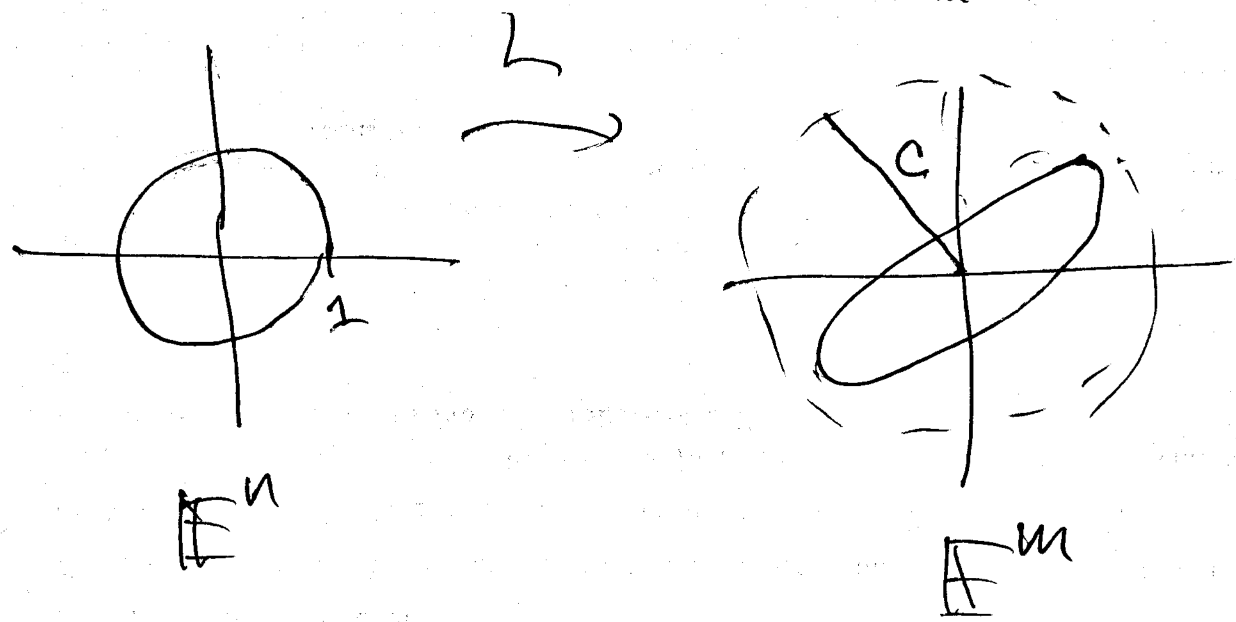
$$L(\vec{x}) = \sum_{j=1}^{n} \sum_{k=1}^{m} t_{j,k} \alpha_j \vec{c}_k = \sum_{k=1}^{m} \underbrace{\left[\sum_{j=1}^{n} t_{j,k} \alpha_j\right]}\vec{c}_k$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \longmapsto \begin{bmatrix} \ \\ \ \\ \ \end{bmatrix} = \underbrace{\begin{bmatrix} t_{11} & t_{21} & t_{31} & \cdots & t_{n1} \\ t_{12} & t_{22} & \cdots & & t_{n2} \\ & & \vdots & & \\ t_{1m} & & \cdots & & t_{nm} \end{bmatrix}}_{[L]_{\mathcal{B} \to \mathcal{C}}} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

Given $L \in \mathcal{L}(\mathbb{E}^n, \mathbb{E}^m)$ then $\exists C$ s.t. $\forall \vec{x}$

$$\|L\vec{x}\|_{\mathbb{E}^m} \le C \|\vec{x}\|_{\mathbb{E}^n}$$





B. The operator norm.

1. Definition. The *operator norm* (or just the *norm*) of $L \in \mathcal{L}(\mathbb{E}^n, \mathbb{E}^m)$ is defined by
$\|L\| = \inf\{C > 0 : \|L\mathbb{x}\| \leq C\|\mathbb{x}\| \; for\ all\ \mathbb{x} \in \mathbb{E}^n\}$
or by

$$\|L\| = \left\{\sup \frac{\|L\mathbb{x}\|}{\|\mathbb{x}\|} : \mathbb{x} \in \mathbb{E}^n\right\}$$

2. Remark. (a) $\|L\mathbb{x}\| \leq C\|\mathbb{x}\|$ for all $\mathbb{x} \in \mathbb{E}^n$ says that the transformation L magnifies the norm of a given $\mathbb{x} \in \mathbb{E}^n$ by a factor of no more than C. The norm $\|L\|$ is the smallest such factor.

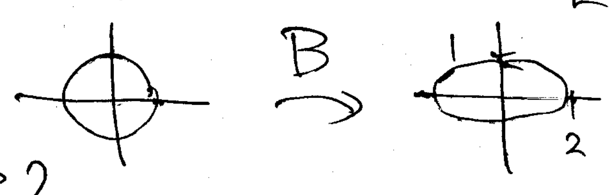
(b) The two quantities appearing in the definition of operator norm are the same. The proof of this is an exercise, but some of your work is done in Exercise 10.1.

(c) Given $L \in \mathcal{L}(\mathbb{E}^n, \mathbb{E}^m)$, $\|L\|$ is always finite, and in fact if L is represented by the matrix $[L] = (l_{i,j})_{m\times n}$ then

$$\|L\| \leq \left(\sum_{i=1}^{m}\sum_{j=1}^{n} l_{i,j}^2\right)^{1/2}$$

however, usually strict inequality holds.

e.g. $L: \mathbb{E}^2 \to \mathbb{E}^2$ represented by $B = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$
(wrt standard basis)



$\|B\| = 2$

But $(2^2 + 1^2)^{1/2} = \sqrt{5} > 2$.

(d) In fact, $||L||$ is independent of the matrix representation of L. Also $\det(L) = \det([L])$ is also independent of the matrix that is used to represent L.

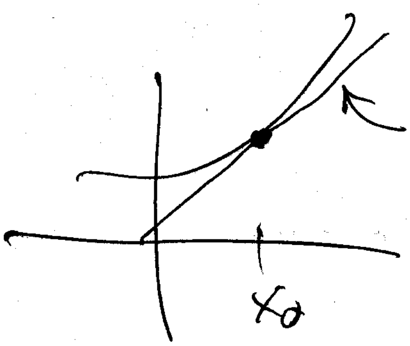
$f: \mathbb{E}^n \to \mathbb{E}^m$

$$f(x_1, \dots, x_n) = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{bmatrix} \qquad f_i : \mathbb{E}^n \to \mathbb{R} = \mathbb{E}^1$$

Q: What is f'?

$f: \mathbb{R} \to \mathbb{R}$.

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0+h) - f(x_0)}{h}$$



$y = f(x_0) + f'(x_0)(x - x_0)$

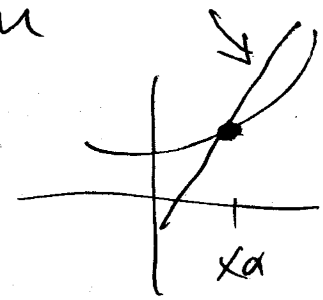
best straight-line approx to $f(x)$ at $x = x_0$.

$$\lim_{h \to 0} \left| \frac{f(x_0+h) - f(x_0)}{h} - f'(x_0) \right| = 0$$

$$\lim_{h \to 0} \frac{|f(x_0+h) - f(x_0) - f'(x_0)h|}{|h|} = 0$$

on $\forall \varepsilon > 0 \ \exists \delta > 0$ s.t. $|h| < \delta$ then

$|f(x_0+h) - f(x_0) - f'(x_0)h| < \varepsilon |h|$.



For any $m \in \mathbb{R}$

$|f(x_0+h) - f(x_0) - mh| \to 0$ as $h \to 0$.

But for a unique m_0 $|f(x_0+h) - f(x_0) - m_0 h| \to 0$ faster than h.

meaning $\dfrac{|f(x_0+h)-f(x_0)-m_0 h|}{|h|} \to 0$ as $h \to 0$.

What is m_0? $f'(x_0)$.

We say $|f(x_0+h)-f(x_0)-f'(x_0)h| = o(|h|)$ as $h \to 0$.

$F(x) = o(G(x))$ as $x \to 0$ means $\dfrac{F(x)}{G(x)} \to 0$ as $x \to 0$

$F(x) = O(G(x))$ as $x \to 0$ means $\dfrac{F(x)}{G(x)}$ is bounded as $x \to 0$

$F(x) \leq C G(x)$ $\forall x$ small.

10.2 Differentiable Functions.

A. The derivative.

1. Motivation. (a) Recall that a function $f: \mathbb{E}^1 \to \mathbb{E}^1$ is differentiable at x_0 in its domain, with derivative $f'(x_0)$ if

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = f'(x_0)$$

Rewriting this in terms of the definition of the limit gives: For every $\epsilon > 0$ there is a $\delta > 0$ such that if $|h| < \delta$ then

$$\left| \frac{f(x_0 + h) - f(x_0)}{h} - f'(x_0) \right| < \epsilon$$

or, rewriting again

$$|f(x_0 + h) - f(x_0) - f'(x_0)h| < \epsilon|h|$$

(b) If we define the linear transformation $A \in \mathcal{L}(\mathbb{E}^1, \mathbb{E}^1)$ by $A(h) = f'(x_0)h$ for all $h \in \mathbb{E}^1$, Then we can rewrite above as

$$|f(x_0 + h) - f(x_0) - A(h)| < \epsilon|h|$$

(c) Now clearly, for any linear transformation $A \in \mathcal{L}(\mathbb{E}^1, \mathbb{E}^1)$, or equivalently any number m, the quantity $|f(x_0 + h) - f(x_0) - A(h)| \to 0$ as $h \to 0$ (assuming f is continuous at x_0). However, the definition of differentiability says that in fact,

$$\frac{|f(x_0 + h) - f(x_0) - A(h)|}{h} \to 0$$

or in other words that $|f(x_0 + h) - f(x_0) - A(h)|$ goes to zero *faster than* h. There is only one transformation A that satisfies this criterion.

(d) We conclude that (i) the derivative $f'(x_0)$ can be thought of as a linear transformation, (ii) this linear transformation has the property that the difference between it and $f(x_0 + h) - f(x_0)$ goes to zero faster than h goes to zero, and (iii) it is the only linear transformation that does so.

(usually we assume D is open and $\vec{x} \in D$.)

2. Definition. Let $\mathbb{f}: D \to \mathbb{E}^m$ for some $D \subseteq \mathbb{E}^n$ and let $\mathbb{x} \in D$ be a cluster point of D. Then $\mathbb{f}$ is *differentiable* at $\mathbb{x}$ with derivative $\mathbb{f}'(\mathbb{x}) \in \mathcal{L}(\mathbb{E}^n, \mathbb{E}^m)$ if

$$\lim_{\mathbb{h} \to \mathbb{o}} \frac{\|\mathbb{f}(\mathbb{x}+\mathbb{h}) - \mathbb{f}(\mathbb{x}) - \mathbb{f}'(\mathbb{x})(\mathbb{h})\|}{\|\mathbb{h}\|} = 0$$

3. Theorem. (10.2.2) If $\mathbb{f}$ is differentiable at $\mathbb{x}_0 \in D$ then $\mathbb{f}$ is continuous at $\mathbb{x}_0$.

Pf: To show f is continuous at $\vec{x}_0$, we must show that $\lim_{\vec{x} \to \vec{x}_0} f(\vec{x}) = f(\vec{x}_0)$, or $\lim_{\vec{x} \to \vec{x}_0} f(\vec{x}) - f(\vec{x}_0) = \vec{0}$ or $\lim_{\vec{x} \to \vec{x}_0} \|f(\vec{x}) - f(\vec{x}_0)\| = 0$.

But letting $\vec{h} = \vec{x} - \vec{x}_0$, we have

$$\|f(\vec{x}) - f(\vec{x}_0)\| = \|f(\vec{x}_0 + \vec{h}) - f(\vec{x}_0)\|$$
$$\leq \|f(\vec{x}_0 + \vec{h}) - f(\vec{x}_0) - f'(\vec{x}_0)\vec{h}\| + \|f'(\vec{x}_0)\vec{h}\|$$
$$\leq \|f(\vec{x}_0 + \vec{h}) - f(\vec{x}_0) - f'(\vec{x}_0)\vec{h}\| + \underbrace{\|f'(\vec{x}_0)\|}\,\|\vec{h}\|$$
$$\to 0 \text{ as } \vec{h} \to \vec{0}.$$

↑ operator norm of $f'(\vec{x}_0)$.

B. Computing $\mathbb{f}'(\mathbb{x})$.

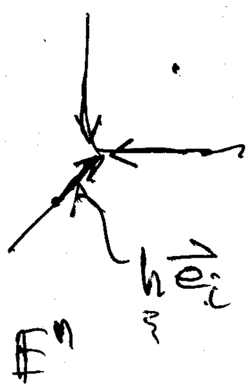
1. Remark. (a) If $\mathbb{f}'(\mathbb{x}) \in \mathcal{L}(\mathbb{E}^n, \mathbb{E}^m)$ then it has a representation as a $m \times n$ matrix with respect to the standard basis. What is that matrix?

(b) Consider first a function $\mathbb{f}: \mathbb{E}^n \to \mathbb{E}^1$, that is a real-valued function of n variables. Let us write $\mathbb{f}(\mathbb{x}) = f(x_1, x_2, \ldots, x_n)$. In this case, for a given $\mathbb{x}_0 = (x_1^0, x_2^0, \ldots, x_n^0)$, $\mathbb{f}'(\mathbb{x}_0)$ is a linear transformation from $\mathbb{E}^n$ to $\mathbb{E}^1$ and hence can be written as

$$\mathbb{f}'(\mathbb{x}_0)\mathbb{h} = \mathbb{a} \cdot \mathbb{h}$$

for $\mathbb{h} \in \mathbb{E}^n$. And we have

$$\lim_{\mathbb{h} \to \mathbb{0}} \frac{\|\mathbb{f}(\mathbb{x}_0 + \mathbb{h}) - \mathbb{f}(\mathbb{x}_0) - \mathbb{a} \cdot \mathbb{h}\|}{\|\mathbb{h}\|} = 0$$



(c) Since the limit exists, we can approach zero from any direction. By letting $\mathbb{h} = h\mathbb{e}_i$, we get $\mathbb{a} \cdot \mathbb{h} = a_i h$ and writing the above limit in components we get

$$\lim_{h \to 0} \frac{f(x_1^0, \ldots, x_j^0 + h, \ldots, x_n^0) - f(x_1^0, \ldots, x_n^0)}{h} = a_i$$

But this is just the usual definition of the partial derivative. So we conclude

$$\mathbb{a} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right) = \nabla f(\mathbb{x}_0)$$

$\vec{a} = (a_1, a_2, \ldots, a_n)$ $\quad$ $\vec{a} \cdot h\vec{e}_i = h a_i$

$h\vec{e}_i = (0, \ldots, 0, h, 0, \ldots, 0)$

↑ ith

$$f: \mathbb{E}^n \to \mathbb{E}^m \qquad f(\vec{x}) = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{bmatrix}$$

$L: \mathbb{E}^n \to \mathbb{E}^m$ linear.

$$L(\vec{x}) = \begin{bmatrix} \vec{b}_1 \cdot \vec{x} \\ \vec{b}_2 \cdot \vec{x} \\ \vdots \\ \vec{b}_m \cdot \vec{x} \end{bmatrix} \quad \text{for } \vec{b}_i \in \mathbb{E}^n.$$

$$\| f(\vec{x}_0 + \vec{h}) - f(\vec{x}_0) - \underline{f'(\vec{x}_0)} \cdot \vec{h} \|$$

$$\left\| \begin{bmatrix} f_1(\vec{x}_0 + \vec{h}) \\ \vdots \\ f_m(\vec{x}_0 + \vec{h}) \end{bmatrix} - \begin{bmatrix} f_1(\vec{x}_0) \\ \vdots \\ f_m(\vec{x}_0) \end{bmatrix} - \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \cdots & & \frac{\partial f_2}{\partial x_n} \\ \frac{\partial f_m}{\partial x_1} & \cdots & & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} \right\|$$

So must have.

$$| f_i(\vec{x}_0 + \vec{h}) - f_i(\vec{x}_0) - \vec{b}_i \cdot \vec{h} | \to 0$$

faster than $\|\vec{h}\|$.

$$\text{but } \vec{b}_i = \begin{bmatrix} \frac{\partial f_i}{\partial x_1} \\ \frac{\partial f_i}{\partial x_2} \\ \vdots \\ \frac{\partial f_i}{\partial x_n} \end{bmatrix}$$

2. Theorem (10.2.3) Let $\mathbb{f}: D \to \mathbb{E}^m$, D an open subset of $\mathbb{E}^n$, be differentiable at $\mathbb{x} \in D$. Then the matrix of $\mathbb{f}'(\mathbb{x})$ with respect to the standard basis is given by

$$\mathbb{f}'(\mathbb{x}) = \left[\frac{\partial f_i}{\partial x_j}\right]_{m \times n}$$

Moreover, for any $\mathbb{v} \in \mathbb{E}^n$,

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

which is defined as the *directional derivative of* $\mathbb{f}$ *in the direction* $\mathbb{v}$ *at* $\mathbb{x}$.