by the way matrix multiplication is defined (which is why it was defined that way).

The above should be old hat, but what’s coming now is probably new. For \( T \in \mathcal{L}(\mathbb{R}^m; \mathbb{R}^n) \), we define the \textit{operator norm} of \( T \) by

\[
\|T\| = \inf \left\{ C > 0 : \|T(x)\| \leq C \|x\|, \forall x \in \mathbb{R}^m \right\}.
\]

Here \( \|T(x)\| \) is using the norm of \( \mathbb{R}^m \), \( \|x\| \) is using the norm of \( \mathbb{R}^n \). Basically what we’re saying is that if \( T \) starts with a small vector \( x \) but comes back with \( T(x) \) being really big, then \( T \) has a large norm.

**Theorem:** If \( T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \), then \( \|T\| < \infty \), i.e., there exists some finite number \( C \) so that \( \|T(x)\| \leq C \|x\| \) holds for all \( x \in \mathbb{R}^n \) (since the only way that the norm could be infinite is if no such \( C \) existed).

**Proof:** Let \( B \) be the matrix of \( T \). Then

\[
T(x) = Bx = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \cdot x \\ \vdots \\ b_m \cdot x \end{pmatrix},
\]

where \( b_i \) is the transpose of the \( i^{th} \) row of \( B \), by definition of matrix multiplication. So,

\[
\|T(x)\|^2 = (b_1 \cdot x)^2 + \cdots + (b_m \cdot x)^2 \\
\leq \|b_1\|^2 \|x\|^2 + \cdots + \|b_m\|^2 \|x\|^2 \\
= \left( \sum_{i=1}^m \sum_{j=1}^n b_{ij}^2 \right) \|x\|^2.
\]

Thus

\[
\sqrt{\sum_{i=1}^m \sum_{j=1}^n b_{ij}^2}
\]

serves as \( C \), showing that the norm is finite, and this in fact gives an upper bound for the norm. It certainly doesn’t in general equal the norm: just consider the identity from \( \mathbb{R}^2 \) to \( \mathbb{R}^2 \). The norm of this transformation is clearly 1, but our bound is \( \sqrt{2} \).

**Example:** If \( T : \mathbb{R}^2 \to \mathbb{R}^2 \) is given by

\[
T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ x_2 \end{pmatrix},
\]

\( \copyright 2012, \) T. Vogel
determine $\|T\|$. (Note: our upper bound is that it’s less than or equal to $\sqrt{5}$.)

**Answer:** Certainly $\|Tx\| = \sqrt{4x_1^2 + x_2^2} \leq \sqrt{4x_1^2 + 4x_2^2} = 2 \|x\|$ holds for every $x$, so $\|T\| \leq 2$. (Since 2 is an element of the set of which $\|T\|$ is the infimum.) On the other hand, $\left\| T \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\| = 2 = 2 \left\| \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\|$, so no number $C$ which is less than 2 can be in the set 
\{ $C > 0 : \|T(x)\| \leq C \|x\|, \forall x \in \mathbb{R}^m$ \}. Thus the infimum of the set must be exactly 2, which is therefore $\|T\|$.

**Proposition:** $\|T(x)\| \leq \|T\| \|x\|$ holds for all $T, x$. (Generalization of the Cauchy-Schwartz inequality.)

**Proof:** This is pretty much straight from the definition of operator norm. Since $\|T\|$ is defined as an infimum, there is a sequence $C_k \in \{ C > 0 : \forall x \in \mathbb{R}^n, \|T(x)\| \leq C \|x\| \}$ which converges to $\|T\|$. We have
$$\|T(x)\| \leq C_k \|x\|$$
holding for all $x, k$. Looking at any particular $x$ and taking $k$ to infinity, we get
$$\|T(x)\| \leq \|T\| \|x\|,$$
as desired.
Definition of differentiability (11.2)

First off, differentiability of a function \( f : \mathbb{R} \rightarrow \mathbb{R} \) is simply that the derivative exists. Then differentiability implies continuity. One might hope that differentiability of a function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is simply that the partials exist. However, this is a very weak condition.

**Example:** Define \( f : \mathbb{R}^2 \rightarrow \mathbb{R} \) by \( f(x, y) = 0 \) if \( x \) or \( y \) is zero, otherwise \( f = 1 \). This has partial derivatives at the origin:

\[
f_x(0, 0) = \lim_{h \to 0} \frac{f(h, 0) - f(0, 0)}{h} = 0,
\]
and similarly for \( y \). But \( f \) is not continuous at the origin. More impressive is the following example.

**Example:** Define \( f \) by

\[
f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}.
\]

We’ve seen this before: the limit doesn’t exist at the origin, so \( f \) is not continuous there. However, the partials of \( f \) exist everywhere:

\[
f_x(x, y) = \begin{cases} \frac{(x^2 + y^2)y - 2x^2y}{(x^2 + y^2)^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases},
\]
where you have to use the definition to find the partial at the origin. You can find \( f_y(x, y) \) similarly. But we wouldn’t want to call \( f \) differentiable at the origin, since it’s not even continuous there. So, we need the proper generalization of derivative from \( \mathbb{R}^1 \).

Recall the differential approximation: that \( h f'(a) \) is a very good approximation of \( f(a + h) - f(a) \).

**Example from calc I:** \( \sqrt{68} \) can be approximated by taking \( f(x) = \sqrt{x} \), \( a = 64 \), \( h = 4 \), so

\[
\sqrt{68} - \sqrt{64} \approx 4f'(64) = 4 \left( \frac{1}{2} \right) 64^{-1/2} = \frac{1}{4},
\]
so our approximation is that

\[
\sqrt{68} \approx 8 \frac{1}{4}.
\]
Essentially what the differential approximation is saying is that the tangent line is a good approximation to the curve. (picture) If $\Delta f = f(a + h) - f(a)$ and $df = h f'(a)$, then we’re saying that $df$ is a good approximation to $\Delta f$. In fact, the difference goes to zero more rapidly than $h$:

$$\lim_{h \to 0} \frac{\Delta f - df}{h} = \lim_{h \to 0} \frac{f(a + h) - f(a) - h f'(a)}{h} = \lim_{h \to 0} \frac{f(a + h) - f(a) - f'(a) h}{h} = 0.$$  

This is the idea that we want to generalize to $\mathbb{R}^n$. Let’s look right now at $f : \mathbb{R}^2 \to \mathbb{R}$. We want to say that $f$ is differentiable at $(a, b)$ iff there exists a tangent plane at $(a, b, f(a, b))$ which in some sense is a sufficiently good approximation to the graph of the function near $(a, b)$. The general plane through that point is $z = f(a, b) + A(x - a) + B(y - b)$. For $f$ to be differentiable, we’ll want $A$ and $B$ so that the plane is a very good approximation to the graph near $(a, b)$. More specifically, look at a point $(a + h, b + k)$ near $(a, b)$. The point on the graph above this point is $(a + h, b + k, f(a + h, b + k))$. The point on the plane above this point is $(a + h, b + k, f(a, b) + Ah + Bk)$. Even if $f$ is merely continuous at $(a, b)$ we’d have

$$\lim_{(h, k) \to (0, 0)} f(a + h, b + k) - f(a, b) - Ah - Bk = 0$$

for any $A$ and $B$, so that’s not enough. Instead, we want $A$ and $B$ so that $f(a + h, b + k) - f(a, b) - Ah - Bk$ goes to zero faster than $\|(h, k)\|$.

**Definition (in special case that $f : \mathbb{R}^2 \to \mathbb{R}$):** $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at $(a, b)$ iff there exist numbers $A$ and $B$ so that

$$\lim_{(h, k) \to (0, 0)} \frac{f(a + h, b + k) - [f(a, b) + Ah + Bk]}{\|(h, k)\|} = 0.$$  

**Example:** Prove that $f(x, y) = x^2 + 2y^2$ is differentiable at $(1, 2)$.

**Proof:** We have $f(1 + h, 2 + k) = (1 + h)^2 + 2(2 + k)^2 = 9 + 2h + h^2 + 8k + 2k^2 = f(1, 2) + 2h + 8k + h^2 + 2k^2$. Take $A$ to be 2 and $B$ to be 8. Then

$$\frac{f(1 + h, 2 + k) - [f(1, 2) + 2h + 8k]}{\sqrt{h^2 + k^2}} = \frac{h^2 + 2k^2}{\sqrt{h^2 + k^2}}.$$  

Using the polar coordinate trick, replacing $h$ by $r \cos \theta$ and $k$ by $r \sin \theta$, the last fraction turns into $r \cos^2 \theta + 2r \sin^2 \theta$, which clearly goes to zero.