Proof. Let $\varepsilon > 0$. By the Fourier theorem we know that

$$\|f - \sum_{n=-N}^{N} \hat{f}(n)e_n\|_2 \leq \varepsilon$$

if N is large enough (depending on ε). In particular, by the triangle inequality this implies that

$$\|f\|_2 - \varepsilon \leq \| \sum_{n=-N}^{N} \hat{f}(n)e_n\|_2 \leq \|f\|_2 + \varepsilon.$$

On the other hand, by Corollary 5 we have

$$\| \sum_{n=-N}^{N} \hat{f}(n)e_n\|_2 = (\sum_{n=-N}^{N} |\hat{f}(n)|^2)^{1/2}$$

and hence

$$(\|f\|_2 - \varepsilon)^2 \leq \sum_{n=-N}^{N} |\hat{f}(n)|^2 \leq (\|f\|_2 + \varepsilon)^2.$$

Taking lim sup, we obtain

$$(\|f\|_2 - \varepsilon)^2 \leq \limsup_{N\to\infty} \sum_{n=-N}^{N} |\hat{f}(n)|^2 \leq (\|f\|_2 + \varepsilon)^2.$$

Since ε is arbitrary, we thus obtain by the squeeze test that

$$\limsup_{N\to\infty} \sum_{n=-N}^{N} |\hat{f}(n)|^2 = \|f\|_2^2$$

and the claim follows. □

There are many other properties of the Fourier transform, but we will not develop them here. In the exercises you will see a small number of applications of the Fourier and Plancherel theorems.

Exercise 16.5.1. Let f be a function in $C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$, and define the *trigonometric Fourier coefficients* a_n, b_n for $n = 0, 1, 2, 3, \ldots$ by

$$a_n := 2\int_{[0,1]} f(x)\cos(2\pi nx)\, dx; \quad b_n := 2\int_{[0,1]} f(x)\sin(2\pi nx)\, dx.$$

(a) Show that the series

$$\frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n\cos(2\pi nx) + b_n\sin(2\pi nx))$$

converges in L^2 metric to f. (Hint: use the Fourier theorem, and break up the exponentials into sines and cosines. Combine the positive n terms with the negative n terms.)

(b) Show that if $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are absolutely convergent, then the above series actually converges uniformly to f, and not just in L^2 metric. (Hint: use Theorem 16.5.3).

Exercise 16.5.2. Let $f(x)$ be the function defined by $f(x) = (1-2x)^2$ when $x \in [0, 1)$, and extended to be $\mathbf{Z}$-periodic for the rest of the real line.

(a) Using Exercise 16.5.1, show that the series

$$\frac{1}{3} + \sum_{n=1}^{\infty} \frac{4}{\pi^2 n^2}\cos(2\pi nx)$$

converges uniformly to f.

(b) Conclude that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$. (Hint: evaluate the above series at $x = 0$.)

(c) Conclude that $\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}$. (Hint: expand the cosines in terms of exponentials, and use Plancherel's theorem.)

Exercise 16.5.3. If $f \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$ and P is a trigonometric polynomial, show that

$$\widehat{f * P}(n) = \hat{f}(n)c_n = \hat{f}(n)\hat{P}(n)$$

for all integers n. More generally, if $f, g \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$, show that

$$\widehat{f * g}(n) = \hat{f}(n)\hat{g}(n)$$

for all integers n. (A fancy way of saying this is that the Fourier transform *intertwines* convolution and multiplication).

Exercise 16.5.4. Let $f \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$ be a function which is differentiable, and whose derivative f' is also continuous. Show that f' also lies in $C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$, and that $\widehat{f'}(n) = in\hat{f}(n)$ for all integers n.

Exercise 16.5.5. Let $f, g \in C(\mathbf{R}/\mathbf{Z}; \mathbf{C})$. Prove the *Parseval identity*

$$\Re \int_0^1 f(x)\overline{g(x)}\, dx = \Re \sum_{n \in \mathbf{Z}} \hat{f}(n)\overline{\hat{g}(n)}.$$

(Hint: apply the Plancherel theorem to $f + g$ and $f - g$, and subtract the two.) Then conclude that the real parts can be removed, thus

$$\int_0^1 f(x)\overline{g(x)}\, dx = \sum_{n \in \mathbf{Z}} \hat{f}(n)\overline{\hat{g}(n)}.$$

(Hint: apply the first identity with f replaced by if.)

Exercise 16.5.6. In this exercise we shall develop the theory of Fourier series for functions of any fixed period L.

Let $L > 0$, and let $f : \mathbf{R} \to \mathbf{C}$ be a complex-valued function which is continuous and L-periodic. Define the numbers c_n for every integer n by

$$c_n := \frac{1}{L} \int_{[0,L]} f(x) e^{-2\pi i n x/L}\, dx.$$

(a) Show that the series

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x/L}$$

converges in L^2 metric to f. More precisely, show that

$$\lim_{N \to \infty} \int_{[0,L]} |f(x) - \sum_{n=-N}^{N} c_n e^{2\pi i n x/L}|^2\, dx = 0.$$

(Hint: apply the Fourier theorem to the function $f(Lx)$.)

(b) If the series $\sum_{n=-\infty}^{\infty} |c_n|$ is absolutely convergent, show that

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x/L}$$

converges uniformly to f.

(c) Show that

$$\frac{1}{L}\int_{[0,L]} |f(x)|^2\, dx = \sum_{n=-\infty}^{\infty} |c_n|^2.$$

(Hint: apply the Plancherel theorem to the function $f(Lx)$.)

Chapter 17

Several variable differential calculus

17.1 Linear transformations

We shall now switch to a different topic, namely that of differentiation in several variable calculus. More precisely, we shall be dealing with maps $f : \mathbf{R}^n \to \mathbf{R}^m$ from one Euclidean space to another, and trying to understand what the derivative of such a map is.

Before we do so, however, we need to recall some notions from linear algebra, most importantly that of a linear transformation and a matrix. We shall be rather brief here; a more thorough treatment of this material can be found in any linear algebra text.

Definition 17.1.1 (Row vectors). Let $n \geq 1$ be an integer. We refer to elements of $\mathbf{R}^n$ as *n-dimensional row vectors*. A typical n-dimensional row vector may take the form $x = (x_1, x_2, \ldots, x_n)$, which we abbreviate as $(x_i)_{1 \leq i \leq n}$; the quantities $x_1, x_2, \ldots, x_n$ are of course real numbers. If $(x_i)_{1\leq i\leq n}$ and $(y_i)_{1\leq i\leq n}$ are n-dimensional row vectors, we can define their vector sum by

$$(x_i)_{1\leq i\leq n} + (y_i)_{1\leq i\leq n} = (x_i + y_i)_{1\leq i\leq n},$$

and also if $c \in \mathbf{R}$ is any scalar, we can define the scalar product $c(x_i)_{1\leq i\leq n}$ by

$$c(x_i)_{1\leq i\leq n} := (cx_i)_{1\leq i\leq n}.$$

Of course one has similar operations on $\mathbf{R}^m$ as well. However, if $n \neq m$, then we do not define any operation of vector addition between vectors in $\mathbf{R}^n$ and vectors in $\mathbf{R}^m$ (e.g., $(2,3,4)+(5,6)$ is undefined). We also refer to the vector $(0,\ldots,0)$ in $\mathbf{R}^n$ as the *zero vector* and also denote it by 0. (Strictly speaking, we should denote the zero vector of $\mathbf{R}^n$ by $0_{\mathbf{R}^n}$, as they are technically distinct from each other and from the number zero, but we shall not take care to make this distinction). We abbreviate $(-1)x$ as $-x$.

The operations of vector addition and scalar multiplication obey a number of basic properties:

Lemma 17.1.2 ($\mathbf{R}^n$ is a vector space). *Let x, y, z be vectors in $\mathbf{R}^n$, and let c, d be real numbers. Then we have the commutativity property $x + y = y + x$, the additive associativity property $(x + y) + z = x + (y + z)$, the additive identity property $x + 0 = 0 + x = x$, the additive inverse property $x + (-x) = (-x) + x = 0$, the multiplicative associativity property $(cd)x = c(dx)$, the distributivity properties $c(x + y) = cx + cy$ and $(c + d)x = cx + dx$, and the multiplicative identity property $1x = x$.*

Proof. See Exercise 17.1.1. □

Definition 17.1.3 (Transpose). If $(x_i)_{1\le i\le n} = (x_1, x_2, \ldots, x_n)$ is an n-dimensional row vector, we can define its *transpose* $(x_i)^T_{1\le i\le n}$ by

$$(x_i)^T_{1\le i\le n} = (x_1, x_2, \ldots, x_n)^T := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

We refer to objects such as $(x_i)^T_{1\le i\le n}$ as *n-dimensional column vectors*.

Remark 17.1.4. There is no functional difference between a row vector and a column vector (e.g., one can add and scalar multiply column vectors just as well as we can row vectors), however we shall (rather annoyingly) need to transpose our row vectors into column vectors in order to be consistent with the conventions of matrix multiplication, which we will see later. Note that we view row vectors and column vectors as residing in different spaces; thus for instance we will not define the sum of a row vector with a column vector, even when they have the same number of elements.

Definition 17.1.5 (Standard basis row vectors). We identify n special vectors in $\mathbf{R}^n$, the *standard basis row vectors* $e_1, \ldots, e_n$. For each $1 \leq j \leq n$, e_j is the vector which has 0 in all entries except for the j^{th} entry, which is equal to 1.

For instance, in $\mathbf{R}^3$, we have $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, and $e_3 = (0, 0, 1)$. Note that if $x = (x_j)_{1 \leq j \leq n}$ is a vector in $\mathbf{R}^n$, then

$$x = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n = \sum_{j=1}^{n} x_j e_j,$$

or in other words every vector in $\mathbf{R}^n$ is a *linear combination* of the standard basis vectors $e_1, \ldots, e_n$. (The notation $\sum_{j=1}^{n} x_j e_j$ is unambiguous because the operation of vector addition is both commutative and associative). Of course, just as every row vector is a linear combination of standard basis row vectors, every column vector is a linear combination of standard basis column vectors:

$$x^T = x_1 e_1^T + x_2 e_2^T + \ldots + x_n e_n^T = \sum_{j=1}^{n} x_j e_j^T.$$

There are (many) other ways to create a basis for $\mathbf{R}^n$, but this is a topic for a linear algebra text and will not be discussed here.

Definition 17.1.6 (Linear transformations). A *linear transformation $T : \mathbf{R}^n \to \mathbf{R}^m$* is any function from one Euclidean space $\mathbf{R}^n$ to another $\mathbf{R}^m$ which obeys the following two axioms:

(a) (Additivity) For every $x, x' \in \mathbf{R}^n$, we have $T(x + x') = Tx + Tx'$.

(b) (Homogeneity) For every $x \in \mathbf{R}^n$ and every $c \in \mathbf{R}$, we have $T(cx) = cTx$.

Example 17.1.7. The *dilation operator* $T_1 : \mathbf{R}^3 \to \mathbf{R}^3$ defined by $T_1 x := 5x$ (i.e., it dilates each vector x by a factor of 5) is a linear transformation, since $5(x+x') = 5x + 5x'$ for all $x, x' \in \mathbf{R}^3$ and $5(cx) = c(5x)$ for all $x \in \mathbf{R}^3$ and $x \in \mathbf{R}$.

Example 17.1.8. The *rotation operator* $T_2 : \mathbf{R}^2 \to \mathbf{R}^2$ defined by a clockwise rotation by $\pi/2$ radians around the origin (so that $T_2(1,0) = (0,1)$, $T_2(0,1) = (-1,0)$, etc.) is a linear transformation; this can best be seen geometrically rather than analytically.

Example 17.1.9. The *projection operator* $T_3 : \mathbf{R}^3 \to \mathbf{R}^2$ defined by $T_3(x,y,z) := (x,y)$ is a linear transformation (why?). The *inclusion operator* $T_4 : \mathbf{R}^2 \to \mathbf{R}^3$ defined by $T_4(x,y) := (x,y,0)$ is also a linear transformation (why?). Finally, the *identity operator* $I_n : \mathbf{R}^n \to \mathbf{R}^n$, defined for any n by $I_n x := x$ is also a linear transformation (why?).

As we shall shortly see, there is a connection between linear transformations and matrices.

Definition 17.1.10 (Matrices). An $m \times n$ *matrix* is an object A of the form

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix};$$

we shall abbreviate this as

$$A = (a_{ij})_{1 \le i \le m; 1 \le j \le n}.$$

In particular, n-dimensional row vectors are $1 \times n$ matrices, while n-dimensional column vectors are $n \times 1$ matrices.

Definition 17.1.11 (Matrix product). Given an $m \times n$ matrix A and an $n \times p$ matrix B, we can define the *matrix product* AB to be the $m \times p$ matrix defined as

$$(a_{ij})_{1 \le i \le m; 1 \le j \le n} (b_{jk})_{1 \le j \le n; 1 \le k \le p} := (\sum_{j=1}^{n} a_{ij} b_{jk})_{1 \le i \le m; 1 \le k \le p}.$$

In particular, if $x^T = (x_j)^T_{1\leq j\leq n}$ is an n-dimensional column vector, and $A = (a_{ij})_{1\leq i\leq m; 1\leq j\leq n}$ is an $m \times n$ matrix, then Ax^T is an m-dimensional column vector:

$$Ax^T = (\sum_{j=1}^{n} a_{ij}x_j)^T_{1\leq i\leq m}.$$

We now relate matrices to linear transformations. If A is an $m \times n$ matrix, we can define the transformation $L_A : \mathbf{R}^n \to \mathbf{R}^m$ by the formula

$$(L_A x)^T := Ax^T.$$

Example 17.1.12. If A is the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix},$$

and $x = (x_1, x_2, x_3)$ is a 3-dimensional row vector, then $L_A x$ is the 2-dimensional row vector defined by

$$(L_A x)^T = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 + 3x_3 \\ 4x_1 + 5x_2 + 6x_3 \end{pmatrix}$$

or in other words

$$L_A(x_1, x_2, x_3) = (x_1 + 2x_2 + 3x_3, 4x_1 + 5x_2 + 6x_3).$$

More generally, if

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

then we have

$$L_A(x_j)_{1\leq j\leq n} = (\sum_{j=1}^{n} a_{ij}x_j)_{1\leq i\leq m}.$$

For any $m \times n$ matrix A, the transformation L_A is automatically linear; one can easily verify that $L_A(x+y) = L_Ax + L_Ay$ and $L_A(cx) = c(L_Ax)$ for any n-dimensional row vectors x, y and any scalar c. (Why?)

Perhaps surprisingly, the converse is also true, i.e., every linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$ is given by a matrix:

Lemma 17.1.13. *Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. Then there exists exactly one $m \times n$ matrix A such that $T = L_A$.*

Proof. Suppose $T : \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation. Let $e_1, e_2, \ldots, e_n$ be the standard basis row vectors of $\mathbf{R}^n$. Then $Te_1, Te_2, \ldots, Te_n$ are vectors in $\mathbf{R}^m$. For each $1 \leq j \leq n$, we write Te_j in co-ordinates as

$$Te_j = (a_{1j}, a_{2j}, \ldots, a_{mj}) = (a_{ij})_{1 \leq i \leq m},$$

i.e., we define a_{ij} to be the i^{th} component of Te_j. Then for any n-dimensional row vector $x = (x_1, \ldots, x_n)$, we have

$$Tx = T(\sum_{j=1}^n x_j e_j),$$

which (since T is linear) is equal to

$$\begin{aligned}
&= \sum_{j=1}^n T(x_j e_j) \\
&= \sum_{j=1}^n x_j T e_j \\
&= \sum_{j=1}^n x_j (a_{ij})_{1 \leq i \leq m} \\
&= \sum_{j=1}^n (a_{ij} x_j)_{1 \leq i \leq m} \\
&= (\sum_{j=1}^n a_{ij} x_j)_{1 \leq i \leq m}.
\end{aligned}$$

But if we let A be the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

then the previous vector is precisely $L_A x$. Thus $Tx = L_A x$ for all n-dimensional vectors x, and thus $T = L_A$.

Now we show that A is unique, i.e., there does not exist any other matrix

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix}$$

for which T is equal to L_B. Suppose for sake of contradiction that we could find such a matrix B which was different from A. Then we would have $L_A = L_B$. In particular, we have $L_A e_j = L_B e_j$ for every $1 \leq j \leq n$. But from the definition of L_A we see that

$$L_A e_j = (a_{ij})_{1 \leq i \leq m}$$

and

$$L_B e_j = (b_{ij})_{1 \leq i \leq m}$$

and thus we have $a_{ij} = b_{ij}$ for every $1 \leq i \leq m$ and $1 \leq j \leq m$, thus A and B are equal, a contradiction. □

Remark 17.1.14. Lemma 17.1.13 establishes a one-to-one correspondence between linear transformations and matrices, and is one of the fundamental reasons why matrices are so important in linear algebra. One may ask then why we bother dealing with linear transformations at all, and why we don't just work with matrices all the time. The reason is that sometimes one does not want to work with the standard basis $e_1, \dots, e_n$, but instead wants to use some other basis. In that case, the correspondence between linear transformations and matrices changes, and so it is still important to keep the notions of linear transformation and matrix distinct. More discussion on this somewhat subtle issue can be found in any linear algebra text.

Remark 17.1.15. If $T = L_A$, then A is sometimes called the *matrix representation of* T, and is sometimes denoted $A = [T]$. We shall avoid this notation here, however.

The composition of two linear transformations is again a linear transformation (Exercise 17.1.2). The next lemma shows that the operation of composing linear transformations is connected to that of matrix multiplication.

Lemma 17.1.16. *Let A be an $m \times n$ matrix, and let B be an $n \times p$ matrix. Then $L_A L_B = L_{AB}$.*

Proof. See Exercise 17.1.3. □

Exercise 17.1.1. Prove Lemma 17.1.2.

Exercise 17.1.2. If $T : \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation, and $S : \mathbf{R}^p \to \mathbf{R}^n$ is a linear transformation, show that the composition $TS : \mathbf{R}^p \to \mathbf{R}^m$ of the two transforms, defined by $TS(x) := T(S(x))$, is also a linear transformation. (Hint: expand $TS(x+y)$ and $TS(cx)$ carefully, using plenty of parentheses.)

Exercise 17.1.3. Prove Lemma 17.1.16.

Exercise 17.1.4. Let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. Show that there exists a number $M > 0$ such that $\|Tx\| \leq M\|x\|$ for all $x \in \mathbf{R}^n$. (Hint: use Lemma 17.1.13 to write T in terms of a matrix A, and then set M to be the sum of the absolute values of all the entries in A. Use the triangle inequality often - it's easier than messing around with square roots etc.) Conclude in particular that every linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$ is continuous.

17.2 Derivatives in several variable calculus

Now that we've reviewed some linear algebra, we turn now to our main topic of this chapter, which is that of understanding differentiation of functions of the form $f : \mathbf{R}^n \to \mathbf{R}^m$, i.e., functions from one Euclidean space to another. For instance, one might want to differentiate the function $f : \mathbf{R}^3 \to \mathbf{R}^4$ defined by

$$f(x, y, z) = (xy, yz, xz, xyz).$$

In single variable calculus, when one wants to differentiate a function $f : E \to \mathbf{R}$ at a point x_0, where E is a subset of $\mathbf{R}$ that contains x_0, this is given by

$$f'(x_0) := \lim_{x \to x_0; x \in E \backslash \{x_0\}} \frac{f(x) - f(x_0)}{x - x_0}.$$

One could try to mimic this definition in the several variable case $f : E \to \mathbf{R}^m$, where E is now a subset of $\mathbf{R}^n$, however we encounter a difficulty in this case: the quantity $f(x) - f(x_0)$ will live in $\mathbf{R}^m$, and $x - x_0$ lives in $\mathbf{R}^n$, and we do not know how to divide an m-dimensional vector by an n-dimensional vector.

To get around this problem, we first rewrite the concept of derivative (in one dimension) in a way which does not involve division of vectors. Instead, we view differentiability at a point x_0 as an assertion that a function f is "approximately linear" near x_0.

Lemma 17.2.1. *Let E be a subset of $\mathbf{R}$, $f : E \to \mathbf{R}$ be a function, $x_0 \in E$, and $L \in \mathbf{R}$. Then the following two statements are equivalent.*

(a) *f is differentiable at x_0, and $f'(x_0) = L$.*

(b) *We have $\lim_{x \to x_0; x \in E - \{x_0\}} \frac{|f(x) - (f(x_0) + L(x - x_0))|}{|x - x_0|} = 0$.*

Proof. See Exercise 17.2.1. □

In light of the above lemma, we see that the derivative $f'(x_0)$ can be interpreted as the number L for which $|f(x) - (f(x_0) + L(x - x_0))|$ is small, in the sense that it tends to zero as x tends to x_0, even if we divide out by the very small number $|x - x_0|$. More informally, the derivative is the quantity L such that we have the approximation $f(x) - f(x_0) \approx L(x - x_0)$.

This does not seem too different from the usual notion of differentiation, but the point is that we are no longer explicitly dividing by $x - x_0$. (We are still dividing by $|x - x_0|$, but this will turn out to be OK). When we move to the several variable case $f : E \to \mathbf{R}^m$, where $E \subseteq \mathbf{R}^n$, we shall still want the derivative to be some quantity L such that $f(x) - f(x_0) \approx L(x - x_0)$. However, since $f(x) - f(x_0)$ is now an m-dimensional vector and $x - x_0$ is an n-dimensional vector, we no longer want L to be a scalar; we want it to be a linear transformation. More precisely:

Definition 17.2.2 (Differentiability). Let E be a subset of $\mathbf{R}^n$, $f : E \to \mathbf{R}^m$ be a function, $x_0 \in E$ be a point, and let $L : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. We say that f is *differentiable at* x_0 *with derivative* L if we have

$$\lim_{x \to x_0; x \in E - \{x_0\}} \frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} = 0.$$

Here $\|x\|$ is the length of x (as measured in the l^2 metric):

$$\|(x_1, x_2, \ldots, x_n)\| = (x_1^2 + x_2^2 + \ldots + x_n^2)^{1/2}.$$

Example 17.2.3. Let $f : \mathbf{R}^2 \to \mathbf{R}^2$ be the map $f(x, y) := (x^2, y^2)$, let x_0 be the point $x_0 := (1, 2)$, and let $L : \mathbf{R}^2 \to \mathbf{R}^2$ be the map $L(x, y) := (2x, 4y)$. We claim that f is differentiable at x_0 with derivative L. To see this, we compute

$$\lim_{(x,y) \to (1,2): (x,y) \neq (1,2)} \frac{\|f(x, y) - (f(1, 2) + L((x, y) - (1, 2)))\|}{\|(x, y) - (1, 2)\|}.$$

Making the change of variables $(x, y) = (1, 2) + (a, b)$, this becomes

$$\lim_{(a,b) \to (0,0): (a,b) \neq (0,0)} \frac{\|f(1 + a, 2 + b) - (f(1, 2) + L(a, b))\|}{\|(a, b)\|}.$$

Substituting the formula for f and for L, this becomes

$$\lim_{(a,b) \to (0,0): (a,b) \neq (0,0)} \frac{\|((1 + a)^2, (2 + b)^2) - (1, 4) - (2a, 4b))\|}{\|(a, b)\|},$$

which simplifies to

$$\lim_{(a,b)\to(0,0):(a,b)\neq(0,0)} \frac{\|(a^2,b^2)\|}{\|(a,b)\|}.$$

We use the squeeze test. The expression $\frac{\|(a^2,b^2)\|}{\|(a,b)\|}$ is clearly non-negative. On the other hand, we have by the triangle inequality

$$\|(a^2,b^2)\| \leq \|(a^2,0)\| + \|(0,b^2)\| = a^2 + b^2$$

and hence

$$\frac{\|(a^2,b^2)\|}{\|(a,b)\|} \leq \sqrt{a^2+b^2}.$$

Since $\sqrt{a^2+b^2} \to 0$ as $(a,b) \to 0$, we thus see from the squeeze test that the above limit exists and is equal to 0. Thus f is differentiable at x_0 with derivative L.

As you can see, verifying that a function is differentiable from first principles can be somewhat tedious. Later on we shall find better ways to verify differentiability, and to compute derivatives.

Before we proceed further, we have to check a basic fact, which is that a function can have at most one derivative at any *interior* point of its domain:

Lemma 17.2.4 (Uniqueness of derivatives). *Let E be a subset of $\mathbf{R}^n$, $f : E \to \mathbf{R}^m$ be a function, $x_0 \in E$ be an interior point of E, and let $L_1 : \mathbf{R}^n \to \mathbf{R}^m$ and $L_2 : \mathbf{R}^n \to \mathbf{R}^m$ be linear transformations. Suppose that f is differentiable at x_0 with derivative L_1, and also differentiable at x_0 with derivative L_2. Then $L_1 = L_2$.*

Proof. See Exercise 17.2.2. □

Because of Lemma 17.2.4, we can now talk about *the* derivative of f at interior points x_0, and we will denote this derivative by $f'(x_0)$. Thus $f'(x_0)$ is the unique linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$ such that

$$\lim_{x\to x_0; x\neq x_0} \frac{\|f(x) - (f(x_0) + f'(x_0)(x-x_0))\|}{\|x-x_0\|} = 0.$$

Informally, this means that the derivative $f'(x_0)$ is the linear transformation such that we have

$$f(x) - f(x_0) \approx f'(x_0)(x - x_0)$$

or equivalently

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

(this is known as *Newton's approximation*; compare with Proposition 10.1.7).

Another consequence of Lemma 17.2.4 is that if you know that $f(x) = g(x)$ for all $x \in E$, and f, g are differentiable at x_0, then you also know that $f'(x_0) = g'(x_0)$ at every *interior* point of E. However, this is not necessarily true if x_0 is a boundary point of E; for instance, if E is just a single point $E = \{x_0\}$, merely knowing that $f(x_0) = g(x_0)$ does not imply that $f'(x_0) = g'(x_0)$. We will not deal with these boundary issues here, and only compute derivatives on the interior of the domain.

We will sometimes refer to f' as the *total derivative* of f, to distinguish this concept from that of partial and directional derivatives below. The total derivative f is also closely related to the *derivative matrix* Df, which we shall define in the next section.

Exercise 17.2.1. Prove Lemma 17.2.1.

Exercise 17.2.2. Prove Lemma 17.2.4. (Hint: prove by contradiction. If $L_1 \neq L_2$, then there exists a vector v such that $L_1 v \neq L_2 v$; this vector must be non-zero (why?). Now apply the definition of derivative, and try to specialize to the case where $x = x_0 + tv$ for some scalar t, to obtain a contradiction.)

17.3 Partial and directional derivatives

We now connect the notion of differentiability with that of partial and directional derivatives, which we now introduce.

Definition 17.3.1 (Directional derivative). Let E be a subset of $\mathbf{R}^n$, $f : E \to \mathbf{R}^m$ be a function, let x_0 be an interior point of E,