Michael Spivak's Calculus on Manifolds - A partial solutions manual

George Lydakis

June 1, 2024

Contents

Chapter 1

Functions on Euclidean Space

1.1 Norm and Inner Product

Exercise 7
A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is **norm preserving** if $|T(x)| = |x|$ and **inner product preserving** if $\langle Tx, Ty \rangle = \langle x, y \rangle$.
a. Prove that T is norm preserving if and only if T is inner product preserving.
b. Prove that such a linear transformation T is 1-1 and T^{-1} is of the same sort.

Solution.
a. Let $x, y \in \mathbb{R}^n$. Firstly, we observe the following relationship between the standard Euclidean norm and inner product:

$$\langle x+y, x+y \rangle = \langle x, x \rangle + 2\langle x, y \rangle + \langle y, y \rangle \implies ||x+y||^2 = ||x||^2 + ||y||^2 + 2\langle x, y \rangle$$

$$\implies \langle x, y \rangle = \frac{||x+y||^2 - ||x||^2 - ||y||^2}{2}$$

Thus, we have that if T is norm-preserving, then for any two $x, y \in \mathbb{R}^n$ it holds that

$$||T(x)|| = ||x||, ||T(y)|| = ||y||, ||T(x+y)|| = ||x+y||$$

Therefore:

$$\langle T(x), T(y) \rangle = \frac{||T(x) + T(y)||^2 - ||T(x)||^2 - ||T(y)||^2}{2} =$$

$$= \frac{||T(x+y)||^2 - ||x||^2 - ||y||^2}{2} = \frac{||x+y||^2 - ||x||^2 - ||y||^2}{2} = \langle x, y \rangle$$

, therefore T is also inner product preserving. For the other direction, if T is inner product preserving, then more specifically for any $x \in \mathbb{R}^n$:

$$||T(x)||^2 = \langle T(x), T(x) \rangle = \langle x, x \rangle = ||x||^2$$

, which means T is norm preserving.
b. Firstly, recall that for linear operators it suffices to show that they are injective in order to show that they are invertible. Therefore, we only need to examine $T(x) = 0$. $T(x) = 0$ by the definition of the inner product is equivalent to $||T(x)|| = 0$. But then because T is norm preserving, we have that $||x|| = ||T(x)|| = 0$, and by the same equivalency as before this means $x = 0$, therefore T is injective. This means that it is also invertible, i.e. T^{-1} is well defined.
Now, for any $x \in \mathbb{R}^n$, we have that $T^{-1}(x) = y$, such that $T(y) = x$. Since T is norm preserving we have that:

$$||T(y)|| = ||y|| \implies ||x|| = ||T^{-1}(x)||$$

, therefore T^{-1} is also norm preserving.

Exercise 8
If $x, y \in \mathbb{R}^n$ are non-zero, the **angle** between x and y, denoted $\angle(x, y)$, is defined as $\arccos(\frac{\langle x,y\rangle}{||x||\cdot||y||})$, which makes sense by the Cauchy-Schwarz inequality. The linear transformation T is **angle preserving** if T is 1-1, and for $x, y \neq 0$ we have $\angle(Tx, Ty) = \angle(x, y)$.
a. Prove that if T is norm preserving, then T is angle preserving.
b. If there is a basis $x_1, \ldots, x_n \in \mathbb{R}^n$ and numbers $\lambda_1, \ldots \lambda_n$ such that $T(x_i) = \lambda_i x_i$, prove that if T is angle preserving then all $|\lambda_i|$ are equal.
c. What are all angle preserving $T : \mathbb{R}^n \to \mathbb{R}^n$?

Solution.
a. Suppose T is norm preserving. Then, as we saw in exercise 7, it is also inner product preserving. Then, for $x, y \in \mathbb{R}^n$:

$$\angle(T(x), T(y)) = \arccos(\frac{\langle T(x), T(y)\rangle}{||T(x)|| \cdot ||T(y)||}) = \arccos(\frac{\langle x, y\rangle}{||x|| \cdot ||y||}) = \angle(x, y)$$

b. In order to simplify our computations, we will first convert the basis $x_1, \ldots, x_n$ to a basis $y_1, \ldots, y_n$ such that $y_i = \frac{x_i}{||x_i||}$, i.e. the new basis consists of unit length vectors. Now consider any two y_i, y_i vectors of this basis. These are clearly non-zero, and also linearly independent, which means that $y_i + y_j, y_i - y_j$ are non-zero as well. Since T is angle preserving, it must hold that:

$$\angle(T(y_i + y_j), T(y_i - y_j)) = \angle(y_i + y_j, y_i - y_j)$$

$$\implies \arccos(\frac{\langle T(y_i + y_j), T(y_i - y_j)\rangle}{||T(y_i - y_j)|| \cdot ||T(y_i + y_j)||}) = \arccos(\frac{\langle y_i + y_j, y_i - y_j\rangle}{||y_i + y_j|| \cdot ||y_i - y_j||}$$

$$\implies \frac{\langle \lambda_i y_i + \lambda_j y_j, \lambda_i y_i - \lambda_j y_j\rangle}{||\lambda_i y_i - \lambda_j y_j|| \cdot ||\lambda_i y_i + \lambda_j y_j||} = \frac{1 - \langle y_i, y_j\rangle + \langle y_j, y_i\rangle - 1}{||y_i + y_j|| \cdot ||y_i - y_j||}$$

$$\implies \langle \lambda_i y_i + \lambda_j y_j, \lambda_i y_i - \lambda_j y_j\rangle = 0 \implies \lambda_i^2 - \lambda_i\lambda_j\langle y_i, y_j\rangle + \lambda_i\lambda_j\langle y_i, y_j\rangle - \lambda_j^2 = 0$$

$$\implies \lambda_i^2 = \lambda_j^2 \implies |\lambda_i| = |\lambda_j|$$

, where throughout the derivation we used the fact that $||y_i|| = ||y_j|| = 1$.
c. By the Polar Decomposition, we know that $T = S\sqrt{T^*T}$ for some isometry S. If T is angle preserving, then for any two non-zero x, y it must be the case that:

$$\frac{\langle S\sqrt{T^*T}(x), S\sqrt{T^*T}(y)\rangle}{||S\sqrt{T^*T}(x)|| \cdot ||S\sqrt{T^*T}(y)||} = \frac{\langle x, y\rangle}{||x|| \cdot ||y||} \implies \frac{\langle\sqrt{T^*T}(x), \sqrt{T^*T}(y)\rangle}{||\sqrt{T^*T}(x)|| \cdot ||\sqrt{T^*T}(y)||} = \frac{\langle x, y\rangle}{||x|| \cdot ||y||}$$

, where we omitted the arccos for brevity and used the properties of isometries. Observe that this means that $\sqrt{T^*T}$ is angle preserving. Additionally, this operator is positive and self-adjoint, which means that by the Real Spectral Theorem there exists an orthonormal basis of $\mathbb{R}^n$ with respect to which its matrix is diagonal. Every entry on the diagonal of this matrix must be positive, since $\sqrt{T^*T}$ is positive. Since these entries are precisely the eigenvalues of $\sqrt{T^*T}$, and since it is angle preserving, by part (b) we obtain that they must all be equal in absolute values, and consequently equal (since they are positive). Therefore, $T = SaI = aS$, for some isometry S and some positive number a.
Conversely, if $T = aS$ for some isometry S and a positive number a, the properties of isometries and the definition of angle preserving operators yields that T is indeed angle preserving (we have but to observe that a^2 will appear both on the numerator and denominator of the argument of arccos).

Exercise 9
If $0 \leq \theta \leq \pi$, let $T \in \mathbb{R}^2 \to \mathbb{R}^2$ have the matrix $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$. Show that T is angle preserving and if $x \neq 0$, then $\angle(x, Tx) = \theta$.

Solution.
From Linear Algebra we know that this matrix corresponds to an isometry in $\mathbb{R}^2$. By part (c) of exercise 8 above, T is angle preserving, since it equals an isometry times (a multiple of) the identity. For a non-zero $x = (x_1, x_2)$ we have that:

$$\angle(x, Tx) = \arccos\left(\frac{\langle x, Tx\rangle}{||x|| \cdot ||Tx||}\right) = \arccos\left(\frac{\langle (x_1, x_2), (\cos\theta x_1 + \sin\theta x_2, -\sin\theta x_1 + \cos\theta x_2)\rangle}{||x|| \cdot ||x||}\right)$$

$$= \arccos\left(\frac{\cos\theta x_1^2 + \sin\theta x_1 x_2 - \sin\theta x_1 x_2 + \cos\theta x_2^2}{||x||^2}\right) = \arccos\left(\frac{\cos\theta ||x||^2}{||x||^2}\right) = \arccos(\cos\theta) = \theta$$

Exercise 10
If $T : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation, show that there is a number M such that $||T(h)|| \leq M||h||$ for $h \in \mathbb{R}^m$. *Hint:* Estimate $||T(h)||$ in terms of $||h||$ and the entries in the matrix of T.

Solution.
We have seen in Hubbard and Hubbard that the following analogue of the Cauchy-Schwarz inequality holds for any matrix A and any vector b:

$$||Ab|| \leq ||A|| \cdot ||b||$$

, where $||A||$ indicates the Frobenius norm of A, i.e. $||A|| = \sqrt{\sum_{i,j} A_{i,j}^2}$. Additionally, we know that for any linear transformation T, $T(h) = \mathcal{M}(T)\mathcal{M}(h)$, where $\mathcal{M}$ indicates the matrices with respect to the standard bases of $\mathbb{R}^m, \mathbb{R}^n$. Therefore, we have that:

$$||T(h)|| = ||\mathcal{M}(T)\mathcal{M}(h)|| \leq ||\mathcal{M}(T)|| \cdot ||h||$$

.
If we set $M = ||\mathcal{M}(T)||$ we obtain the inequality requested by the exercise.

Exercise 13
If $x, y \in \mathbb{R}^n$, then x and y are called perpendicular (or orthogonal) if $\langle x, y\rangle = 0$. If x, y are perpendicular, prove that $||x + y||^2 = ||x||^2 + ||y||^2$.

Solution.
By the properties of the inner product and the definition of the associated norm we have that:

$$||x + y||^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle = ||x||^2 + 0 + 0 + ||y||^2 = ||x||^2 + ||y||^2$$

Exercise 26
Let $A = \{(x, y) \in \mathbb{R}^2 : x > 0, 0 < y < x^2\}$.
(a) Show that every straight line through (0, 0) contains an interval around (0, 0) which is in $\mathbb{R}^2 - A$.
(b) Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x) = 0$ if $x \notin A$ and $f(x) = 1$ if $x \in A$. For $h \in \mathbb{R}^2$ define $g_h : \mathbb{R} \to \mathbb{R}$ by $g_h(t) = f(th)$. Show that each g_h is continuous at 0, but f is not continuous at (0, 0).

Solution.
(a) Let us first consider the case of non-vertical lines, i.e. lines of the form $y = ax$. The points lying on this line are of the form (x, ax). We are thus interested in examining when is it the case that $(x, ax) \notin A$. This happens precisely whenever $x \leq 0$ or $0 \geq ax$ or $ax \geq x^2$. There are two cases:

- $a \leq 0$: then the second inequality holds for all non-negative x. The first inequality holds trivially for all negative x. Thus it suffices to pick e.g. the line segment from $(-a, -a^2)$ to (a, a^2), and this will lie entirely in $\mathbb{R}^2 - A$. Geometrically, this line has a negative or zero slope, and as such it never intersects the area between the parabola $y = x^2$ and the horizontal axis.

- $a > 0$: Then the first inequality again holds trivially for all non-positive x. For positive x, we divide both sides of the third inequality with x to obtain $a \geq x$. Thus for all $0 < x \leq a$, the points (x, ax) lie in $\mathbb{R}^2 - A$. Pick, therefore, the line segment from $(-a, -a^2)$ to (a, a^2). Geometrically, we have found the intersection of the line with the parabola $y = x^2$ and, for aesthetic purposes, made the line segment symmetric around $(0, 0)$.

In either case, the line segment from $(-a, -a^2)$ to (a, a^2) lies in $\mathbb{R}^2 - A$.
Now we examine the case of the vertical line $x = 0$. Here the line segment $(0, -1)$ to $(0, 1)$ works trivially: indeed, any point with $x = 0$ satisfies the first of the inequalities, $x \leq 0$. We have thus found an interval of the desired form for all lines through the origin.
(b) The value of f at $(0, 0)$ is $f((0, 0)) = 0$ since $(0, 0) \notin A$. Consider the sequence of points $i \to p_i = (\frac{1}{i}, \frac{1}{i^2} - \frac{1}{2i^3})$. As $i \to \infty$, both the x- and the y-coordinate here tend to 0. Therefore, this sequence of points in $\mathbb{R}^2$ tends to $(0, 0)$. For f to be continuous, it must be the case that $\lim_{i \to \infty} f(p_i) = f((0, 0)) = 0$. Each p_i clearly has an x coordinate x_i which satisfies $x_i > 0$. Furthermore, its y-coordinate is $y_i = \frac{1}{i^2} - \frac{1}{2i^3} = \frac{2i-1}{2i^3}$. For any integer $i \geq 1$, this is clearly positive. Additionally, $y_i - x_i^2 = -\frac{1}{2i^3} < 0 \implies y_i < x_i^2$. This means that all p_i are in A, and thus $f(p_i) = 1$. Obviously, the limit of this constant sequence of $f(p_i)$ as $i \to \infty$ is $1 \neq 0$, therefore f is not continuous at the origin.
Now consider any g_h such that $g_h(t) = f(th), h \in \mathbb{R}^2$. th defines a line in $\mathbb{R}^2$ that passes through the origin. From (a) we know that there exists a line segment around which lies entirely in $\mathbb{R}^2 - A$, and thus f evaluated on any point p of it is $f(p) = 0$. Let p_1, p_2 be the endpoints of this line segment. Then $p_1 = t_1 h, p_2 = t_2 h$ and WLOG $t_1 < t_2$ (we proved that this is a non-degenerate line segment, thus t_1 cannot equal t_2). But then for all $t \in [t_1, t_2]$, $g_h(t) = f(th) = 0$. Clearly, this implies that g_h is continuous at 0, since it is constant on an interval around it (and on 0 itself).

Chapter 2

Differentiation

2.1 Basic Definitions

Exercise 4
Let g be a continuous real-valued function on the unit circle $\{x \in \mathbb{R}^2 \ : \ ||x|| = 1$ such that $g(0,1) = g(1,0) = 0$ and $g(-x) = -g(x)$. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by:

$$f(x) = \begin{cases} ||x|| \cdot g\Big(\frac{x}{||x||}\Big), \ x \neq 0 \\ 0, \ x = 0 \end{cases}$$

(a) If $x \in \mathbb{R}^2$ and $h : \mathbb{R} \to \mathbb{R}$ is defined by $h(t) = f(tx)$, show that h is differentiable.
(b) Show that f is not differentiable at $(0,0)$ unless $g = 0$.
Hint: first show that $D_f(0,0)$ would have to be 0 by considering (h,k) with $k = 0$ and then with $h = 0$.

Solution.
(a) Let us consider the following cases:

- $x = 0$: in this case, $h(t) = f(0) = 0$ for all $t \in \mathbb{R}$, which is clearly a differentiable function.
- $x \neq 0$: in this case, the expansion of h depends on t. For $t > 0$, $h(t) = f(tx) = ||tx|| \cdot g\Big(\frac{tx}{||tx||}\Big) = t||x|| \cdot g\Big(\frac{x}{||x||}\Big)$. For $t = 0$, $h(0) = f(0) = 0$. For $t < 0, h(t) = f(tx) = ||tx|| \cdot g\Big(\frac{tx}{||tx||}\Big) = -t||x|| \cdot g\Big(-\frac{x}{||x||}\Big) = t||x||g\Big(\frac{x}{||x||}\Big)$, where we used the fact that $g(-x) = -g(x)$.

  Clearly, for $t > 0$ or $t < 0$, h is differentiable as a polynomial function (everything that depends on x is constant). Therefore, the only point of interest is $t = 0$. We apply the definition of the derivative, starting with the case $u \to 0^+$:

$$\lim_{u \to 0^+} \frac{h(0+u) - h(0)}{u} = \lim_{u \to 0^+} \frac{u||x|| \cdot g\Big(\frac{x}{||x||}\Big)}{u} = ||x|| \cdot g\Big(\frac{x}{||x||}\Big)$$

  Now, for $u \to 0^-$:

$$\lim_{u \to 0^-} \frac{h(0+u) - h(0)}{u} = \lim_{u \to 0^-} \frac{u||x|| \cdot g\Big(\frac{x}{||x||}\Big)}{u} = ||x|| \cdot g\Big(\frac{x}{||x||}\Big)$$

  , which means that this limit does indeed exist, and thus $h'(0) = ||x|| \cdot g\Big(\frac{x}{||x||}\Big)$. Consequently, h is indeed differentiable.

(b) Consider approaching $(0,0)$ with a sequence of points $i \to (\frac{1}{i}, 0)$. Observe, firstly, that for any such point, $f(\frac{1}{i}, 0) = ||(\frac{1}{i}, 0)|| \cdot g\Big(\frac{(\frac{1}{i},0)}{||(\frac{1}{i},0)||}\Big) = \frac{1}{i} \cdot g(1,0) = 0$. A similar argument shows that $f(0, \frac{1}{i}) = 0$. But this means precisely that both partial derivatives of f at (0, 0) are zero.
We know that if f were differentiable at $(0,0)$, the unique linear transformation which would be the derivative would equal the Jacobian matrix at 0, which here is a row of two zeros.
Now suppose that g is not zero everywhere, i.e. there exists $x = (x_1, x_2), ||x|| = 1$ such that $g(x_1, x_2) = c \neq 0$. Consider the sequence $i \to (\frac{x_1}{i}, \frac{x_2}{i})$, which tends to $(0,0)$. If f is differentiable at $(0,0)$, the directional derivative that corresponds to this vector would have to be 0 (due to the Jacobian being the zero matrix). However, recall that in part (a) we found that for any x, more particularly for the x we examine here, $h'(0) = ||x|| \cdot g\Big(\frac{x}{||x||}\Big) = g(x) = c \neq 0$. But the directional derivative of f along x at $(0,0)$ is:

$$\lim_{u \to 0} \frac{f(u(x_1, x_2)) - 0}{u} = h'(0)$$

, for an h defined by this x. Clearly, this is non-zero, and as such the derivative of f at (0, 0) cannot exist unless $g = 0$, in which case f is trivially differentiable at (0, 0) as the zero function.

Exercise 9
Two functions $f, g : \mathbb{R} \to \mathbb{R}$ are equal up to n-th order at a if:

$$\lim_{h \to 0} \frac{f(a+h) - g(a+h)}{h^n} = 0$$

(a) Show that f is differentiable at a if and only if there is a function g of the form $g(x) = a_0 + a_1(x-a)$ such that f, g are equal up to first order at a.
(b) If $f'(a), \ldots, f^(n)(a)$ exist, show that f and the function g defined by

$$g(x) = \sum_{i=0}^{n} \frac{f^(i)(a)}{i!}(x-a)^i$$

are equal up to n-th order at a. *Hint*: The limit

$$\lim_{x \to a} \frac{f(x) - \sum_{i=0}^{n-1} \frac{f^{(i)}(a)}{i!}(x-a)^i}{(x-a)^n}$$

may be evaluated by L'Hospital's rule.

Solution.
(a) $\Longrightarrow$: Suppose first that f is differentiable at a, in which case $f'(a)$ is well defined as the familiar limit. Let then $g(x) = f(a) + f'(a)(x-a)$. Then:

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

As $h \to 0$, the first fraction here equals $f'(a)$. The same holds for the second fraction. Therefore, one can apply the addition rule for limits to obtain that the initial limit is zero, and as such f, g are equal up to first order at a.
$\Longleftarrow$: Conversely, suppose that f is equal up to first order at a to a function of the form $g(x) = a_0 + a_1(x-a)$. First of all, observe the following:

$$f(a+h) - g(a+h) = \frac{f(a+h) - g(a+h)}{h} \cdot h$$

, for non-zero h. We have written $f(a+h) - g(a+h)$ as a product of two functions whose limits at 0 are both 0. Therefore, it also holds that $\lim_{h \to 0}(f(a+h) - g(a+h)) = 0$. It is trivial to see that the limit of

$g(a+h)$ as h approaches 0 is $g(a) = a_0$. Therefore, one can obtain the corollary that $\lim_{h\to 0} f(a+h) = a_0$. Because f is continuous at a, we obtain that $f(a) = a_0$. Now:

$$\lim_{h\to 0} \frac{f(a+h) - g(a+h)}{h} = \lim_{h\to 0} \frac{f(a+h) - a_0 - a_1 h}{h} = \lim_{h\to 0} \frac{f(a+h) - f(a) - a_1 h}{h}$$
$$= \lim_{h\to 0} \Big(\frac{f(a+h) - f(a)}{h} - a_1 \Big)$$

We know this limit equals 0, which by the addition (subtraction) rule for limits implies also that the limit of the first fraction must equal a_1, and this precisely means that f is differentiable at a.
(b) In order to show that f, g are equal up to n-th order at a we have to show that:

$$\lim_{h\to 0} \frac{f(a+h) - g(a+h)}{h^n} = 0$$

By performing a change of variables where $x = h + a$, the limit can be rewritten as:

$$\lim_{x\to a} \frac{f(x) - g(x)}{(x-a)^n} = 0$$

We then write:

$$\lim_{x\to a} \frac{f(x) - g(x)}{(x-a)^n} = \lim_{x\to a} \frac{f(x) - \sum_{i=0}^{n} \frac{f^{((i)}(a)}{i!}(x-a)^i}{(x-a)^n}$$

The denominator here tends to 0 as $x \to a$. The same holds for the numerator, since all terms of the sum for $i > 0$ tend to 0, while for $i = 0$ the corresponding term tends to $f(a)$, which is what $f(x)$ tends to as well. Because both the numerator and denominator are differentiable, one can apply L'Hospital's rule. Observe that this will lower the degree of the denominator by 1, while for the numerator the 0-th term of the sum will become 0 (since we are differentiating a constant).
Now the 1-th term of the sum tends to $f'(a)$, which is what $f'(x)$ tends to as well. We can therefore apply L'Hospital's rule again, and in fact we can do this n times. At the n-th application the denominator is now constant. The numerator will be:

$$f^{(n)}(x) - \frac{f^{(n)}(a)}{n!} \cdot 1 \cdot 2 \cdot \ldots \cdot n = f^{(n)}(x) - f^{(n)}(a)$$

, since again we repeatedly differentiated $(x-a)^n$ and all other terms of the sum have by then vanished. But clearly, due to the existence of all derivatives up to n for f, this tends to 0. Consequently, by L'Hospital's rule, all previous limits are zero as well, meaning that the original limit, which is the definition of equality up to n-th order of f, g, is also 0.

2.2 Basic Theorems

Exercise 12
A function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ is **bilinear** if for $x, x_1, x_2 \in \mathbb{R}^n, y, y_1, y_2 \in \mathbb{R}^m$ and $a \in \mathbb{R}$ we have

$$f(ax, y) = af(x, y) = f(x, ay),$$

$$f(x_1 + x_2, y) = f(x_1, y) + f(x_2, y),$$
$$f(x, y_1 + y_2) = f(x, y_1) + f(x, y_2)$$

(a) Prove that if f is bilinear, then
$$\lim_{(h,k)\to 0} \frac{||f(h, k)||}{||(h, k)||} = 0$$
(b) Prove that $Df(a, b)(x, y) = f(a, y) + f(x, b)$.
(c) Show that the formula for $Dp(a, b)$ in Theorem 2-3 ($Dp(a, b)(x, y) = bx+ay, p(x, y) = x\cdot y, x, y \in \mathbb{R}$) is a special case of (b).

Solution.
(a) We begin by observing that any $h \in \mathbb{R}^n, k \in \mathbb{R}^m$ can be written as linear combinations of the respective standard bases:
$$h = \sum_{i=1}^{n} a_i e_i, k = \sum_{j=1}^{m} b_j f_j$$

By the properties of bilinearity, we then have that:

$$f(h, k) = f(\sum_{i=1}^{n} a_i e_i, \sum_{j=1}^{m} b_j f_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j f(e_i, f_j)$$

Each $f(e_i, f_j)$ is a p-dimensional vector. Consequently, by applying the triangle inequality in $\mathbb{R}^p$ we obtain that:
$$||f(h, k)|| \leq \sum_{i=1}^{n}\sum_{j=1}^{m} |a_i b_j| \cdot ||f(e_i, f_j)||$$

Call $M = \max_{i,j}\{||f(e_i, f_j)||\}$, in which case $||f(h, k)|| \leq M \sum_{i=1}^{n}\sum_{j=1}^{m} |a_i b_j|$. Now consider the following vectors $x, y \in \mathbb{R}^{n\cdot m}$:

$$x = \begin{array}{c} \left\{\begin{array}{c} \\ \\ \\ \end{array}\right. m \\ \left\{\begin{array}{c} \\ \\ \\ \end{array}\right. m \\ \left\{\begin{array}{c} \\ \\ \\ \end{array}\right. m \end{array} \begin{pmatrix} |a_1| \\ |a_1| \\ \vdots \\ |a_2| \\ |a_2| \\ \vdots \\ |a_n| \\ |a_n| \\ \vdots \\ |a_n| \end{pmatrix}, y = \begin{pmatrix} |b_1| \\ |b_2| \\ \vdots \\ |b_m| \\ |b_1| \\ |b_2| \\ \vdots \\ |b_m| \\ |b_1| \\ |b_2| \\ \vdots \\ |b_m| \end{pmatrix}$$

From this we have that $||x|| = \sqrt{ma_1^2 + ma_2^2 + \ldots + ma_n^2} = \sqrt{m}\sqrt{\sum_{i=1}^{n} a_i^2} = \sqrt{m}||h||$, and similarly that $||y|| = \sqrt{n}||k||$. We further observe that $\langle x, y\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} |a_i b_j|$. Then, by applying the Cauchy-Schwarz inequality:
$$||f(h, k)|| \leq M||x||\cdot||y|| = M\sqrt{mn}||h|| \cdot ||k||$$

Then:

$$\frac{||f(h,k)||}{||(h,k)||} = \frac{||f(h,k)||}{\sqrt{a_1^2 + \ldots + a_n^2 + b_1^2 + \ldots + b_m^2}} = \frac{||f(h,k)||}{\sqrt{||h||^2 + ||k||^2}} \le M\sqrt{mn}\frac{||h|| \cdot ||k||}{\sqrt{||h||^2 + ||k||^2}}$$

Now we have that:

$$0 \le ||h||^2 - 2||h|| \cdot ||k|| + ||k||^2 \le ||h||^2 - ||h|| \cdot ||k|| + ||k||^2$$

$$\implies ||h|| \cdot ||k|| \le ||h||^2 + ||k||^2 \implies \frac{||h|| \cdot ||k||}{\sqrt{||h||^2 + ||k||^2}} \le \sqrt{||h||^2 + ||k||^2}$$

This leads us to conclude that:

$$\frac{||f(h,k)||}{||(h,k)||} \le M\sqrt{mn}\sqrt{||h||^2 + ||k||^2}$$

Observe that the RHS here tends to 0 as $(h,k) \to 0$. Therefore, the same holds for the LHS, which means that we safely conclude that $\lim_{(h,k)\to 0} \frac{||f(h,k)||}{||(h,k)||}$.
(b) Take any $a, x \in \mathbb{R}^n, b, y \in \mathbb{R}^m$. We need to show that:

$$\lim_{(x,y)\to 0} \frac{1}{||(x,y)||}(f(a+x, b+y) - f(a,b) - (f(a,y) + f(x,b))) = 0$$

By applying the properties of bilinearity again we can rewrite the limit as:

$$\lim_{(x,y)\to 0} \frac{1}{||(x,y)||}(f(a,b) + f(a,y) + f(x,b) + f(x,y) - f(a,b) - f(a,y) - f(x,b))$$

$$= \lim_{(x,y)\to 0} \frac{f(x,y)}{||(x,y)||}$$

The limit we computed in (a) that involves the norm of the numerator here also guarantees that this limit above is zero as well. We have thus shown that $Df(a,b)(x,y) = f(a,y) + f(x,b)$.
(c) Consider the function $p(x,y) = x \cdot y$. For any $a, x_1, x_2, y_1, y_2 \in \mathbb{R}$ we have that:

$$p(x_1 + x_2, y_1) = (x_1 + x_2) \cdot y_1 = p(x_1, y_1) + p(x_2, y_1)$$

$$p(x_1, y_1 + y_2) = x_1 \cdot (y_1 + y_2) = p(x_1, y_1) + p(x_1, y_2)$$

$$p(ax_1, y_1) = a \cdot x_1 \cdot y_1 = a \cdot p(x_1, y_1)$$

Hence, this is a bilinear function and we can apply part (b) to obtain that:

$$Dp(a,b)(x,y) = p(a,y) + p(x,b) = a \cdot y + x \cdot b$$

, which agrees with what was previously proven.

Exercise 13
Define $IP : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by $IP(x,y) = \langle x, y \rangle$.
(a) Find $D(I)(a,b)$ and $(IP)'(a,b)$.
(b) If $f, g : \mathbb{R} \to \mathbb{R}^n$ are differentiable and $h\mathbb{R} \to \mathbb{R}$ is defined by $h(t) = \langle f(t), g(t) \rangle$, show that

$$h'(a) = \langle f'(a)^T, g(a) \rangle + \langle f(a), g'(a)^T \rangle$$

(Note that $f'(a)$ is a $n \times 1$ matrix; its transpose $f'(a)^T$ is a $1 \times n$ matrix, which we consider as a member of $\mathbb{R}^n$.
(c) If $f : \mathbb{R} \to \mathbb{R}^n$ is differentiable and $|f(t)| = 1$ for all t, show that $\langle f'(t)^T, f(t) \rangle = 0$.
(d) Exhibit a differentiable function $f : \mathbb{R} \to \mathbb{R}$ such that the function $|f|$ defined by $|f|(t) = |f(t)|$ is not differentiable.

Solution.
(a) Observe that IP is a bilinear function (easily shown by applying the properties of the inner product. Then, by applying the result of exercise 12 we obtain that:

$$D(IP)(a,b)(x,y) = IP(a,y) + IP(x,b) = \langle a, y\rangle + \langle x, b\rangle$$

To compute the Jacobian matrix $IP'(a,b)$, we have that:

$$(IP)'(a,b) = \begin{pmatrix} D(IP)(a,b)(e_1,0) & \dots & D(IP)(a,b)(e_n,0) & D(IP)(a,b)(0,e_1)\dots & D(IP)(a,b)(0,e_n)\end{pmatrix}$$
$$= \begin{pmatrix} b_1 & b_2 \dots b_n & a_1 & a_2 & \dots & a_n \end{pmatrix}$$

(b) We have already proved this in Hubbard & Hubbard (main text as well as exercise 6, paragraph 1.8). Nevertheless, we'll provide another proof based on part (a).
First, let $k : \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}^n, k(t) = (f(t), g(t))$. Then we have that $h(a) = IP(k(a))$. By the composition rule of differentiation (note that IP, k are both differentiable) we have that:

$$h'(a) = D(IP)(k(a))D(k)(a)$$

Here it is the case that $D(k)(a) = (f'(a)^T, g'(a)^T)$ (interpeted as a column vector). Thus:

$$h(a) = D(IP)(k(a))(f'(a)^T, g'(a)^T) = \langle f(a), g'(a)^T\rangle + \langle f'(a)^T, g(a)\rangle$$

(c) Note that for such an $f, h(t) = \langle f(t), f(t)\rangle = ||f(t)||^2 = 1$. This is therefore a constant function of t, and as such $h'(t) = 0$. By part (b):

$$0 = h'(t) = \langle f'(t)^T, f(t)\rangle + \langle f(t), f'(t)^T\rangle = 2\langle f'(t), f(t)\rangle \implies \langle f'(t)^T, f(t)\rangle = 0$$

Geometrically, this means that since f's values always lie on the unit hypersphere and f is differentiable, then the tangent vector on each $f(t)$ is perpendicular to $f(t)$ (which confirms the well-known result for the unit circle and its tangents if $n = 2$).
(d) One has but to consider the function $f(t) = t$, which has $f'(t) = 1$, yet $|f|(t) = |t|$, which is not differentiable at 0.

Exercise 14
Let $E_i, i = 1, \dots, k$ be Euclidean spaces of various dimensions. A function $f : E_1 \times \dots \times E_k \to \mathbb{R}^p$ is called **multilinear** if for each choice of $x_j \in E_j, j \neq i$ the function $g : E_i \to \mathbb{R}^p$ defined by $g(x) = f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_k)$ is a linear transformation.
(a) If f is multilinear and $i \neq j$, show that for $h = (h_1, \dots, h_k)$, with $h_l \in E_l$, we have

$$\lim_{h\to 0} \frac{||f(a_1, \dots, h_i, \dots, h_j, \dots, a_k)||}{||h||} = 0$$

Hint: If $g(x,y) = f(a_1, \dots, x, \dots, y, \dots, a_k)$, then g is bilinear.
(b) Prove that

$$Df(a_1, \dots, a_k)(x_1, \dots, x_k) = \sum_{i=1}^{k} f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_k)$$

Solution.
(a) First we suppose that $j = i + 1$. As indicated in the hint, it is easy to show that the function $g(h_i, h_j) = f(a_1, \dots, h_i, h_j, \dots, a_k)$ is bilinear (a simple application of the definition of bilinearity and the definition of multilinearity). In addition, we have that $||h|| = \sqrt{||h_i||^2 + ||h_j||^2 + z}, z \geq 0$. Thus:

$$\frac{||f(a_1, \dots, h_i, h_j, \dots a_k)||}{||h||} = \frac{||g(h_i, h_j)||}{\sqrt{||h_i||^2 + ||h_j||^2 + z}} \leq \frac{||g(h_i, h_j)||}{\sqrt{||h_i||^2 + ||h_j||^2}} = \frac{||g(h_i, h_j)||}{||(h_i, h_j)||}$$

Precisely because g is bilinear, by exercise 12 we know that the RHS tends to 0 as $h \to 0$ (since in that case $(h_i, h_j) \to 0$ as well), and because of the inequality the LHS tends to 0 as well.

We will now apply a "finite induction" on the difference $\Delta = j - i$. We have already shown that the required statement holds for $\Delta = 1$. Suppose that it holds for $\Delta = m \geq 1$. We need to show that it holds for $m + 1$ as well. To do this, if $j = i + m$ we have the following:

$$||f(a_1, \ldots, h_i, \ldots, h_{j-1}, h_j, a_k)|| = ||f(a_1, \ldots, h_i, \ldots, h_{j-1}, \sum_{l=1}^{\dim E_j} b_l e_l, \ldots, a_k)||$$

$$= ||\sum_{l=1}^{\dim E_j} b_l f(a_1, \ldots, h_i, \ldots, h_{j-1}, e_l, \ldots, a_k)|| \leq \sum_{l=1}^{\dim E_j} |b_l| \cdot ||f(a_1, \ldots, h_i, \ldots, h_{j-1}, e_l, \ldots, a_k||,$$

by applying the triangle inequality in the j-th Euclidean space, and by using its standard basis e_l. Now we observe that the inductive hypothesis guarantees that each term of the sum, when divided by $||h||$ tends to 0 as $h \to 0$, since it only contains Δ arguments that are not constant. Furthermore, each $|b_l|$ will also tend to 0, since $h \to 0$, which we know guarantees $h_j \to 0$ as well, which we know that in turn guarantees that each of its components tends to 0. We've therefore proven that the required limit is 0 for $\Delta = m$ as well, thus completing the proof for any "distance" of the indices $i, j, i \neq j$).
(b) We examine the quantity inside the limit definition of the derivative:

$$\lim_{h \to 0} \frac{1}{||h||}(f(a_1 + h_1, a_2 + h_2, \ldots, a_k + h_k) - f(a_1, \ldots a_k) - \sum_{i=1}^{k} f(a_1, \ldots, a_{i-1}, h_i, a_{i+1}, \ldots, a_k))$$

Let us work with the numerator by applying the additivity property that arises from the definition of multilinearity:

$$f(a_1 + h_1, \ldots, a_k + h_k) - f(a_1, \ldots, a_k) - \sum_{i=1}^{k} f(a_1, \ldots, a_{i-1}, h_i, a_{i+1}, \ldots, a_k)$$

$$= \sum_{I \in \mathcal{P}(\{1, \ldots, k\})} f(C_1, \ldots, C_K) - f(a_1, \ldots, a_k) - \sum_{i=1}^{k} f(a_1, \ldots, a_{i-1}, h_i, a_{i+1}, \ldots, a_k),$$

where $C_j = h_j$ if $j \in I$ and a_j otherwise (notice that this just an abbreviated way of writing the full sum that results from multilinearity). Now we observe the following:

- The term $f(C_1, \ldots, C_k)$ when $I = \emptyset$ cancels out with $f(a_1, \ldots, a_k)$.
- All terms $f(C_1, \ldots, C_k)$ when $I = \{j\}, j = 1, \ldots, k$ cancel out with the terms of the proposed derivative.
- All other terms feature *at least two* arguments C_i, C_j that equal h_i, h_j. From part (a), the quotient of each of those over $||h||$ tends to 0 as $h \to 0$. But then by a simple application of the triangle inequality, the norm of $f(a_1 + h_1, \ldots, a_k + h_k) - f(a_1, \ldots, a_k) - \sum_{i=1}^{k} f(a_1, \ldots, a_{i-1}, h_i, a_{i+1}, \ldots, a_k)$ over $||h||$ also tends to 0 as $h \to 0$.

From this we conclude that the desired limit from which we started is also 0, which is precisely the definition of the derivative for f, thus completing the proof.

Exercise 15
Regard an $n \times n$ matrix as a point in the n-fold product $\mathbb{R}^n \times \ldots \times \mathbb{R}^n$ by considering each row as a member of $\mathbb{R}^n$.
(a) Prove that $\det : \mathbb{R}^n \times \ldots \times \mathbb{R}^n \to \mathbb{R}$ is differentiable and

$$D(\det)(a_1, \ldots, a_n)(x_1, \ldots, x_n) = \sum_{i=1}^{n} \det \begin{pmatrix} a_1 \\ \vdots \\ x_i \\ \vdots \\ a_n \end{pmatrix}$$

(b) If $a_{ij} : \mathbb{R} \to \mathbb{R}$ are differentiable and $f(t) = \det(a_{ij}(t))$, show that

$$f'(t) = \sum_{j=1}^{n} \det \begin{pmatrix} a_{11}(t) & \ldots & a_{1n}(t) \\ \vdots & \vdots & \vdots \\ a'_{j1}(t) & \ldots & a'_{jn}(t) \\ \vdots & \vdots & \vdots \\ a_{n1}(t) & \ldots & a_{nn}(t) \end{pmatrix}$$

(c) If $\det(a_{ij}(t)) \neq 0$ for all t and $b_1, \ldots, b_n : \mathbb{R} \to \mathbb{R}$ are differentiable, let $s_1, \ldots, s_n : \mathbb{R} \to \mathbb{R}$ be the functions such that $s_1(t), \ldots, s_n(t)$ are the solutions of the equations

$$\sum_{j=1}^{n} a_{ij}(t) s_j(t) = b_i(t), i = 1, 2, \ldots, n$$

Show that s_i is differentiable and find $s_i'(t)$.

(a) We recall from Linear Algebra that the determinant of a matrix is a linear function with respect to each of its rows when all other rows are kept constant. But then this means that det as it is defined here is a multilinear function in $\mathbb{R}^n \times \ldots \times \mathbb{R}^n$. Therefore, exercise 14 is directly applicable here: det is differentiable, and in fact it must be the case that: $D(\det)(a_1, \ldots, a_n)(x_1, \ldots, x_n) = \sum_{i=1}^n \det(a_1, \ldots, x_i, \ldots, a_n)$, which because we are treating rows as vectors of $\mathbb{R}^n$ will equal precisely the sum of determinants as they are given in the exercise (stacking $a_1, \ldots, x_i, \ldots, a_n$ row-wise).
(b) Here we have a collection of functions a_{ij} each of which gives one element of the matrix $f(t)$ for a certain t. The resulting f will therefore be differentiable as a Cartesian product and composition of differentiable functions. To compute f', we first define $g : \mathbb{R} \to \mathbb{R}^n \times \ldots \times \mathbb{R}^n$ as:

$$g(t) = (a_{11}(t), a_{12}(t), \ldots, a_{nn}(t))$$

We have that:

$$g'(t) = \begin{pmatrix} a'_{11}(t) \\ a'_{12}(t) \\ \vdots \\ a'_{nn}(t) \end{pmatrix},$$

whose dimensions make sense as a linear transformation from $\mathbb{R}$ to $\mathbb{R}^n \times \ldots \times \mathbb{R}^n$. By part(a) we additionally have that:

$$\det{}'(a_1, \ldots, a_n) = \begin{pmatrix} D_1 \det(a_1, \ldots, a_n) & \ldots & D_{n \cdot n} \det(a_1, \ldots, a_n) \end{pmatrix}$$

$$= \left(\sum_{i=1}^n \det \begin{pmatrix} a_1 \\ \vdots \\ e_1 \\ \vdots \\ a_n \end{pmatrix} \quad \sum_{i=1}^n \det \begin{pmatrix} a_1 \\ \vdots \\ e_2 \\ \vdots \\ a_n \end{pmatrix} \quad \ldots \quad \sum_{i=1}^n \det \begin{pmatrix} a_1 \\ \vdots \\ e_{n \cdot n} \\ \vdots \\ a_n \end{pmatrix} \right),$$

where in each sum the corresponding e_j is placed on each row of the matrix whose determinant is computed. With this definition of g we have that $f(t) = \det(g(t))$, and then by the chain rule:

$$f'(t) = \det'(g(t))g'(t)$$

$$= \left(\sum_{i=1}^n \det \begin{pmatrix} a_1 \\ \vdots \\ e_1 \\ \vdots \\ a_n \end{pmatrix} \quad \sum_{i=1}^n \det \begin{pmatrix} a_1 \\ \vdots \\ e_2 \\ \vdots \\ a_n \end{pmatrix} \quad \dots \quad \sum_{i=1}^n \det \begin{pmatrix} a_1 \\ \vdots \\ e_{n \cdot n} \\ \vdots \\ a_n \end{pmatrix} \right) \cdot \begin{pmatrix} a'_{11}(t) \\ a'_{12}(t) \\ \vdots \\ a'_{nn}(t) \end{pmatrix},$$

where here we use a_i as shorthand notation for a row formed by stacking a_{ij} column-wise for all j. By computing the inner product one ends up with a sum of sums. Each inner sum is a sum of determinants of matrices with one row that features a single "1" and $n-1$ zeros. This is multiplied by an element of the form $a'_{ij}(t)$ that can be inserted into the matrix by the per-row linearity of the determinant. Lastly, we can again use the per-row additivity of the determinant to gather together all matrices that have one of e_j in the same row (i.e. all of the matrices whose j-th row equals one of the basis vectors multiplied by a_{ji} for some i). By rewriting everything in this way, we are left with a single sum of determinants of matrices that looks as follows:

$$f'(t) = \sum_{j=1}^n \det \begin{pmatrix} a_{11}(t) & \dots & a_{1n}(t) \\ \vdots & \dots & \vdots \\ a'_{j1}(t) & \dots & a'_{jn}(t) \\ \vdots & \dots & \vdots \\ a_{n1}(t) & \dots & a_{nn}(t) \end{pmatrix},$$

since the "gathering" described above was such that for each j, all $a'_{ji}(t)$ ended up in the same row.
(c) By creating two vectors $s(t), b(t)$ out of all $s_i(t), b_i(t)$ and naming $A(t)$ the matrix resulting from $a_{ij}(t)$, the given equations can be rewritten as $A(t)^T s(t) = b(t)$. Since $\det A(t) = \det A(t)^T \neq 0, A(t)^T$ is invertible and this equation has a unique solution. We can now use Cramer's rule to obtain a solution for $s(t)$:

$$s_j(t) = \frac{\det \begin{pmatrix} a_{11}(t) & \dots & b_1(t) & \dots & a_{n1}(t) \\ a_{12}(t) & \dots & b_2(t) & \dots & a_{n2}(t) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{1n}(t) & \dots & b_n(t) & \dots & a_{nn}(t) \end{pmatrix}}{\det A(t)},$$

where $b_i(t)$ are placed in the j-th column. This consists of a composition and quotient of differentiable functions, and as such is also differentiable. Now we can use parts (a) and (b) to obtain a solution for $s'_j(t)$. We name $g(t) : \mathbb{R} \to \mathbb{R}$ the numerator, in which case from part (b):

$$g'(t) = \sum_{j=1}^n \det \begin{pmatrix} a_{11}(t) & \dots & b_1(t) & \dots & a_{n1}(t) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a'_{1j}(t) & \dots & b'_j(t) & \dots & a'_{nj}(t) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{1n}(t) & \dots & b_n(t) & \dots & a_{nn}(t) \end{pmatrix}$$

We name $h(t) : \mathbb{R} \to \mathbb{R}$ the denominator, in which case again from part (b):

$$h'(t) = \sum_{j=1}^n \det \begin{pmatrix} a_{11}(t) & \dots & a_{1n}(t) \\ \vdots & \dots & \vdots \\ a'_{j1}(t) & \dots & a'_{jn}(t) \\ \vdots & \dots & \vdots \\ a_{n1}(t) & \dots & a_{nn}(t) \end{pmatrix}$$

We then use the quotient rule of differentiation to obtain that:

$$s_j'(t) = \frac{g'(t)h(t) - h'(t)g(t)}{h^2(t)}$$

, which we will not expand for the sake of brevity.

Exercise 16
Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable and has a differentiable inverse $f^{-1} : \mathbb{R}^n \to \mathbb{R}^n$. Show that $(f^{-1})'(a) = [f'(f^{-1}(a))]^{-1}$.
Hint: $f \circ f^{-1}(x) = x$.

Solution.
As indicated in the hint, we write:

$$f \circ f^{-1}(x) = x$$

We know that $(x)'(a) = a$, i.e. that for $g(x) = x, Dg = I$. Furthermore, both f, f^{-1} are differentiable, thus by applying the composition rule of derivation on the LHS, we obtain:

$$[D(f \circ f^{-1})](a)(v) = [D(f)](f^{-1}(a))[D(f^{-1})](v)$$

By using the equation $f \circ f^{-1}(x) = x$, we obtain that the corresponding derivatives must also be equal. In terms of (Jacobian) matrices, this would mean that:

$$f'(f^{-1}(a))(f^{-1})'(a) = I,$$

and since we are dealing with square matrices, the two matrices on the LHS constitute a pair of a matrix and its inverse, meaning that we can write:

$$(f^{-1})'(a) = (f'(f^{-1}(a)))^{-1}$$

2.3 Partial Derivatives

Exercise 22
If $f : \mathbb{R}^2 \to \mathbb{R}$ and $D_2 f = 0$, show that f is independent of the second variable. If $D_1 f = D_2 f = 0$, show that f is constant.

Solution.
Fix any $x \in \mathbb{R}$. We need to show that for any two $y_1, y_2 \in \mathbb{R}, f(x, y_1) = f(x, y_2)$ (in other words, that f cannot change if its first argument is kept constant). Let $g : \mathbb{R} \to \mathbb{R}, g_x(z) = f(x, z)$. Since $D_2 f =$ everywhere, g_x is differentiable and by the definition of partial derivative, $g_x(z) = 0$. From calculus 1 we know that this means that g_x is constant. which is equivalent to $f(x, y_1) = f(x, y_2)$ for any two y_1, y_2.
In the second case, suppose $D_1 f = D_2 f = 0$ everywhere. Suppose that f is not constant, in other words that $f(x_1, y_1) \neq f(x_2, y_2)$ for at least two $z_1 = (x_1, y_1), z_2 = (x_2, y_2), z_1 \neq z_2$.
Consider then the function $g : \mathbb{R} \to \mathbb{R}, g(t) = f(t(z_2 - z_1) + z_1)$. We note that $g(0) = f(z_1), g(1) = f(z_2)$, hence that $g(0) \neq g(1)$. Furthermore, because both partial derivatives of f are 0 everywhere, they are more specifically continuous everywhere. Since $\mathbb{R}^2$ is open, we know that this implies that f is actually differentiable in all of $\mathbb{R}^2$. Therefore Df is well defined. By the chain rule we now have that:

$$[Dg(a)](t) = [D_f(t(z_2 - z_1) + z_1)] \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix} \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix} = 0$$

Therefore $g'(t) = 0$ everywhere (we can simplify the notation since g is a one-variable function). But then by the Mean Value Theorem applied on $(0, 1)$ we'd have that for some $x \in (0, 1)$:

$$g'(x) = \frac{g(1) - g(0)}{1} \neq 0$$

, which is clearly a contradiction. Therefore f must be constant.

Exercise 23
Let $A = \{(x, y) \in \mathbb{R}^2 : x < 0, \text{ or } x \geq 0 \text{ and } y \neq 0\}$.
(a) If $f : A \to \mathbb{R}$ and $D_1 f = D_2 f = 0$, show that f is constant.
Hint: Note that any two points in A can be connected by a sequence of lines each parallel to one of the axes.
(b) Find a function $f : A \to \mathbb{R}$ such that $D_2 f = 0$ but f is not independent of the second variable.

Solution.
(a) A equals the plane minus the positive part of the horizontal axis and the origin, and is an open set, thus f is differentiable (by the same argument as in 22). Here we'd like to apply the same technique as in exercise 22. However, we observe that if we were to pick z_1, z_2 like we did in 22 and define g in the same way, it's not necessarily true that g will be well defined in all of $(0, 1)$. Indeed, if for example $z_1 = (1, -1), z_2 = (2, 1)$, then the line passing through z_1, z_2 intersects the positive part of the horizontal axis, and as such for some $t \in (0, 1), g(t) = f(t(z_2 - z_1) + z_1)$ cannot be defined.
We can use the provided hint to resolve this problem in the following way. Pick any two $z_1 = (x_1, y_1) \in A, z_2 = (x_2, y_2) \in A$, where WLOG, $y_2 \geq y_1$.

- If $y_2 = y_1$, then both of these are either positive or negative, and the line connecting z_1, z_2 is horizontal and lies entirely in A. Therefore exercise 22 is immediately applicable.

- If $y_2 > y_1$, we observe that we can connect z_1, z_2 with a sequence of line segments each parallel to one of the axes. We will not do this entirely rigorously, but the idea is that we start from z_1, draw a vertical line towards (x_1, y_2), and if we encounter the positive part of the horizontal axis while doing this, we pick $(x_1, \epsilon_1), \epsilon_1 \neq 0)$, and then draw a horizontal line from this point until the x-coordinate becomes negative. From that point on, we continue with a vertical line (until y becomes y_2) that is guaranteed to lie entirely within A (since x is now negative), and then finally draw a horizontal line towards z_2, which again lies entirely within A since $y_2 \neq 0$.

  For each of the created line segments, a g can be defined as in 22 based on the endpoints. Successive applications of the Mean Value Theorem yield that the values of f on the endpoints of each line segment must be equal. Then the transitive property of equality results in the fact that $f(z_2) = f(z_1)$, which contradicts the hypothesis. Therefore f is again constant.

(b) Consider the function $f : A \to \mathbb{R}$:

$$f(x, y) = \begin{cases} x^2, & x \geq 0, y > 0 \\ 0, & x < 0 \\ -x^2, & x \geq 0, y < 0 \end{cases}$$

Clearly, $D_2 f = 0$ everywhere. However, we observe that $f(1, 1) = 1, f(1, -1) = -1$, which are not equal despite the fact that the arguments differ only in the second variable. This means that f is not independent of the second variable.

Exercise 24
Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

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

(a) Show that $D_2 f(x, 0) = x$ for all x and $D_1 f(0, y) = -y$ for all y.
(b) Show that $D_{1,2} f(0, 0) \neq D_{2,1} f(0, 0)$.

Solution.
(a) First we examine the case $(x, 0) \neq (0, 0)$, that is, $x \neq 0$. We have then that:

$$D_2 f(x, y) = \frac{\partial xy\frac{x^2 - y^2}{x^2 + y^2}}{\partial y} = x\frac{x^2 - y^2}{x^2 + y^2} + xy\frac{-2y(x^2 + y^2) + (x^2 - y^2)2y}{(x^2 + y^2)^2}$$

$$= \frac{(x^3 - xy^2)(x^2+y^2) - 2xy^2(x^2+y^2) + 2xy^2(x^2-y^2)}{(x^2+y^2)^2}$$

$$= \frac{x^5 + x^3y^2 - x^3y^2 - xy^4 - 2x^3y^2 - 2xy^4 + 2x^3y^2 - 2xy^4}{(x^2+y^2)^2} = \frac{x^5 - 5xy^4}{(x^2+y^2)^2}$$

Evaluated for $y = 0$, this expression indeed yields that:

$$D_2 f(x,0) = \frac{x^5}{x^4} = x$$

To compute $D_2 f(0,0)$, we have that:

$$D_2 f(0,0) = \lim_{h \to 0} \frac{f(0, 0+h) - f(0,0)}{h} = \lim_{h \to 0} \frac{0-0}{h} = 0$$

, which is consistent with $D_f(x,0) = x$. For $(0,y) \neq (0,0)$, a similar computation will yield that $D_1 f(0,y) = -y$ (omitted for brevity). For $D_1 f(0,0)$, we have that:

$$D_1 f(0,0) = \lim_{h \to 0} \frac{f(0+h, 0) - f(0,0)}{h} = \lim_{h \to 0} \frac{0-0}{h} = 0$$

, again consistent with $D_1 f(0,y) = -y$.
(b) By using part (a) and the definition of second-order partial derivatives, we have that:

$$D_{1,2} f(0,0) = \lim_{h \to 0} \frac{D_1 f(0, 0+h) - D_1 f(0,0)}{h} = \lim_{h \to 0} \frac{-(0+h) - 0}{h} = -1$$

$$D_{2,1} f(0,0) = \lim_{h \to 0} \frac{D_2 f(0+h, 0) - f(0,0)}{h} = \lim_{h \to 0} \frac{0 + h - 0}{h} = 1$$

, so these are clearly not equal.

Exercise 25
Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} e^{-x^{-2}}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Show that f is a C^∞ function, and $f^{(i)}(0) = 0$ for all i.
Hint: The limit $f'(0) = \lim_{h \to 0} \frac{e^{-h^{-2}}}{h} = \lim_{h \to 0} \frac{1/h}{e^{h^{-2}}}$ can be evaluated by L' Hospital's rule. It is easy enough to find $f'(x)$ for $x \neq 0$, and $f''(0) = \lim_{h \to 0} f'(h)/h$ can then be found by L' Hospital's rule.

Solution.
For $x \neq 0$, we can see that $f^{(n)}(x)$ always exists since f is given by a "standard" formula. More precisely, taking any such derivative would result in an expression featuring $e^{-x^{-2}}$ and powers of x, which we know to infinitely differentiable. The question is thus whether $f^{(n)}(0)$ always exists. To begin, we have that:

$$f'(0) = \lim_{h \to 0} \frac{e^{-h^{-2}} - 0}{h}$$

, which, as given in the hint, can be evaluated with L' Hospital's rule, the result being 0. For the second derivative at 0, we first have that for $x \neq 0, f'(x) = \frac{2e^{-x^{-2}}}{x^3}$. Thus:

$$f''(0) = \lim_{h \to 0} \frac{f'(h) - f'(0)}{h} = \lim_{h \to 0} \frac{2e^{-h^{-2}}}{h^4}$$

Again by L' Hospital's rule this can be shown to be 0. Now we make the following observation. For any $n \geq 1, f^{(n)}(x)$ for $x \neq 0$ can be written as a sum of terms of the form $C\frac{e^{-x^{-2}}}{x^k}, k \geq 3$. This can be verified via the calculations above for $n = 1$, while for $n \geq 2$ we observe that, due to the product rule, the differentiation of either x^{-k} or $e^{x^{-2}}$ will increase the exponents of the denominators. To prove then that $f^{(n)}(0) = 0$ for all $n \geq 3$, that is, that $\lim_{h\to 0} \frac{f^{(n-1)}(h) - f^{(n-1)}(0)}{h}$, it suffices to show that $\lim_{h\to 0} \frac{Ce^{-h^{-2}}}{h^n} = 0$ for all $n \geq 3$. We have that:

$$L = \lim_{h\to 0} C\frac{e^{-h^{-2}}}{h^n} = \lim_{h\to 0} C\frac{1/h^n}{e^{h^{-2}}}$$

, and we can use L' Hospital's rule to obtain that:

$$L = \lim_{h\to 0} C\frac{-n/h^{n+1}}{-2h^{-3}e^{h^{-2}}} = \lim_{h\to 0} \frac{Cn}{2} \cdot \frac{1/h^{n-2}}{e^{h^{-2}}}$$

Recall that we have already shown this to be true when $n - 2 = 1$, i.e. when $n = 3$ and when $n - 2 = 2 \implies n = 4$. But then by induction it will also hold for any $n \geq 5$, since then $n - 2 \geq 3$ (our base case is $n = 3, 4$). By the arguments presented above, this also guarantees that $f^{(i)}(0) = 0$ for all i, and hence that f is C^∞.

Exercise 26
Let

$$f(x) = \begin{cases} e^{-(x-1)^{-2}} e^{-(x+1)^{-2}}, & x \in (-1, 1) \\ 0, & x \notin (-1, 1) \end{cases}$$

(a) Show that $f : \mathbb{R} \to \mathbb{R}$ is a C^∞ function which is positive on $(-1, 1)$ and 0 elsewhere.
(b) Show that there is a C^∞ function $g : \mathbb{R} \to [0, 1]$ such that $g(x) = 0$ for $x \leq 0$ and $g(x) = 1$ for $x \geq \epsilon$. *Hint*: If f is a C^∞ function which is positive on $(0, \epsilon)$ and 0 elsewhere, let $g(x) = \frac{\int_0^x f(y)dy}{\int_0^\epsilon f(y)dy}$.
(c) If $a \in \mathbb{R}^n$, define $g : \mathbb{R}^n \to \mathbb{R}$ by

$$g(x) = f([x^1 - a^1]/\epsilon) \cdot \ldots \cdot f([x^n - a^n]/\epsilon)$$

Show that g is a C^∞ function which is positive on

$$(a^1 - \epsilon, a^1 + \epsilon) \times \cdots \times (a^n - \epsilon, a^n + \epsilon)$$

and zero elsewhere.
(d) If $A \subset \mathbb{R}^n$ is open and $C \subset A$ is compact, show that there is a non-negative C^∞ function $f : A \to \mathbb{R}$ such that $f(x) > 0$ for $x \in C$ and $f = 0$ outside of some closed set contained in A.
(e) Show that we can choose such an f so that $f : A \to [0, 1]$ and $f(x) = 1$ for $x \in C$. *Hint*: If the function f of (d) satisfies $f(x) \geq \epsilon$ for $x \in C$, consider $g \circ f$, where g is the function of (b).

Solution.
(a) The second part of the required statement is trivially true: e raised to any power is always positive, and by definition f is zero outside of $(-1, 1)$. Additionally, it's trivially true that f is C^∞ outside of $[-1, 1]$ as a constant function, and in $(-1, 1)$ as a "standard" formula/composition. The only points that we need to check explicitly are -1 and 1. For each of them, the left and right n-th derivative is always 0 respectively (since $f(x) = 0, x \notin (-1, 1)$). We thus need to examine the right n-th derivative at -1 and the left n-th derivative at 1 only.
Let $g(x) = e^{-x^{-2}}$, in which case $f(x) = g(x - 1)g(x + 1)$. Notice also that this means that the n-th derivative of f consists of a sum of products of derivatives of g up to and including n-th order. For example, $f'(x) = g'(x - 1)g(x + 1) + g(x - 1)g'(x + 1)$ and

$$f''(x) = g''(x - 1)g(x + 1) + 2g'(x - 1)g'(x + 1) + g''(x + 1)g(x - 1)$$

Then we have that:

$$\lim_{h\to 0^+} \frac{f(-1 + h) - f(-1)}{h} = \lim_{h\to 0^+} \frac{g(-2 + h)g(h)}{h}$$

We observe that due to what we proved in exercise 25, this limit can be decomposed into the product of two limits, the second of which is zero and the first of which is a real number (continuity of g). Thus the

limit itself is zero and $f'(-1)_+ = 0$ (the right derivative of f at -1). Therefore $f'(-1) = 0$. With an exactly symmetrical argument we can show that $f'(1) = 0$.
We now also notice that if we take the limit definition of the n-th derivative at either -1 or 1, we will always end up with a numerator that is a sum of terms each of which is a product that consists of *at least one* partial derivative of g at 0, potentially more, and some terms of the form $g^{(n)}(-2+h)$ or $g^{(n)}(2-h)$. Again by using exercise 25, we "pair up" the denominator h of each of the terms of the sum with the partial of g at 0, observe that this is zero, and that all remaining terms of each product tend to real numbers. Thus the limit of the sum itself is zero, meaning that the n-th partial of f at both -1 and 1 is zero. This concludes the proof that f is C^∞.
(b) If we let $h(x) = f(\frac{2x}{\epsilon} - 1)$, this shifted and scaled version of f will also be C^∞ (this is easy to see by the fact that the composition rule simply introduces a multiplicative constant) and will be positive in $(0, \epsilon)$ and 0 outside of it. Notice that h is integrable as a continuous function. Now, as indicated in the hint, let $g(x) = \frac{\int_0^x h(y)dy}{\int_0^\epsilon h(y)dy}$. Then, g is clearly 0 for $x \leq 0$, whereas for $x \geq \epsilon$ the integration of f up to x will equal its integration up to ϵ only, since f is zero after 1. Furthermore, the Fundamental Theorem of Calculus directly relates g' to f, and as such shows that g is also C^∞.
(c) Clearly, if any $(x^i - a^i)/\epsilon$ is not in $(-1, 1)$, or equivalently if $x^i \notin (a^i - \epsilon, a_1 + \epsilon)$, the corresponding $f([x^1 - a^i]/\epsilon)$ will be zero, and thus g is also zero. On the other hand, if each $x^i \in (a^i - \epsilon, a^i + \epsilon)$ then the corresponding values of f are all positive, and hence $g(x)$ is also positive. We conclude that $g(x)$ is positive iff $x^i \in (a^i - \epsilon, a^i + \epsilon)$ for all i, and is zero otherwise. Now we need to show that all partial derivatives of all orders exist. It's clear from exercise 25 that each partial derivative of order 1 exists, since these are basically constant multiples of f'. By the product rule we observe that each $\frac{\partial^{(n)} g}{\partial x^{i_1} x^{i_2} \ldots x^{i_n}}$ is a sum of products of various orders of derivatives of f, and as such the above observation generalizes to the fact that each such partial derivative is itself differentiable, and therefore g is C^∞.
(d) In the trivial case where $C = \emptyset$, pick $f : A \to \mathbb{R}, f(x) = 0$ to satisfy the given constraints. Otherwise, C contains at least one point and is both closed and bounded. We now distinguish between two cases regarding A:

- If $A = \mathbb{R}^n$, which is the only non-empty set that can be both open and closed, then simply pick an open ball that contains C and along with its boundary let it be known as a closed set S. Then set $a \in \mathbb{R}^n = 0$ and ϵ to be the radius of this closed ball, and define g as we did in (c). It is now easy to see that this g satisfies the constraints required by the exercise.

- If $A \neq \mathbb{R}^n$, then we claim that for every $x \in C$, there exists ϵ such that $B_\epsilon(x) \in A$. Indeed, if this was not the case for some x, then A would by definition not be open. By picking then $\epsilon' < \epsilon$, we also have that the closed ball of radius ϵ' around x is fully contained in A (this will be important later on). Now, for each $x \in C$ find such ϵ' and define g_x to be as in (c). Notice also that the collection $\mathcal{O} = \{B_{\epsilon'(x)}(x), x \in C\}$ is an open cover of C. Because C is compact, we know that we can find a *finite collection* of sets $S_i \in \mathcal{O}$ that also covers C.

  This corresponds to a finite number, say N, of functions $g_1, g_2, \ldots g_N$ that are as in (c). It's easy enough to see that the function $g = \sum_{i=1}^N g_i$ is also C^∞ since each of them is C^∞. Additionally, one can see that g is non-negative as a sum of non-negative functions, and that in fact $g(x) > 0$ for $x \in C$, since each $x \in C$ belongs in at least one S_i, and then $g_i(x) > 0$. Lastly, we examine the behavior of g on points that do not belong in any S_i. In that case we see that each g_i must be zero on these points, and as such g is zero on them as well. But then g is zero on all points outside the closure of the finite union of open sets S_i, and this closure is contained in A (because we previously picked ϵ' such that not only the open ball $B'_\epsilon(x)$ is contained in A but also the corresponding closed one).

  This concludes the proof that g has the properties that were requested in this case as well.

(e) Pick an f that satisfies the constraints of (d), which we have now shown is always possible. f is C^∞, and therefore continuous. The set C is compact, and we therefore know that f achieves a minimum value, say $m > 0$ on it. Therefore $f(x) \geq m$ for all $x \in C$. We showed in (b) that for any $\epsilon > 0$, there exists a function $g : \mathbb{R} \to [0, 1]$ which is $C^\infty, g(x) = 0$ when $x \leq 0$ and $g(x) = 1$ when $x \geq \epsilon$. Pick then a g that corresponds to $\epsilon = m$. For our f here, we have that for $x \in C, f(x) \geq m$. Therefore the composition $g \circ f$

will have a value of 1 whenever $x \in C$, and will be C^∞ as a composition of C^∞ functions. Furthermore, since $f(x) = 0$ outside of some closed set contained in A and $g(x) = 0$ for $x \leq 0$, we have that for x that do not belong in this closed set, $g(f(x)) = 0$. These observations show that $g \circ f : A \to [0, 1]$ is C^∞, has a value of 1 for $x \in C$ and a value of 0 outside some closed set contained in A.