Chapter 1
Introduction to Differentiable Functions

Summary *We introduce differentiable functions, directional and partial derivatives, graphs and level sets of functions of several variables.*

In this concise chapter we introduce continuous and differentiable functions between arbitrary finite dimensional spaces. We pay particular attention to notation, as appropriate notation is often the difference between simple and complicated presentations of several-variable calculus. Once this is in place many of our calculations follow the same lines as in the one dimensional calculus. We do not include proofs but, for readers familiar with analysis, we provide suggestions that lead to proofs along the lines that apply in the one variable calculus.

The following extremely simple example illustrates the type of calculation we will be executing frequently and the reader should practice similar examples until they become routine and the intermediate step is unnecessary.

Example 1.1 Let

$$f : \mathbb{R}^3 \longrightarrow \mathbb{R}, \;\; f(x, y, z) = xe^y + y^2 z^3.$$

The *partial derivative* of f with respect to x, $\dfrac{\partial f}{\partial x}$ or f_x, is obtained by treating y and z as constants and differentiating with respect x in the usual one variable way. Thus if $A = e^y$ and $B = y^2 z^3$ then $f(x, y, z) = Ax + B$ and

$$\frac{\partial f}{\partial x} = \frac{d}{dx}(Ax + B) = A = e^y.$$

Similarly if $C = x$ and $D = z^3$ then $f(x, y, z) = Ce^y + Dy^2$ and

$$\frac{\partial f}{\partial y} = \frac{d}{dy}(Ce^y + Dy^2) = Ce^y + 2Dy = xe^y + 2yz^3,$$

and, if $E = xe^y$ and $F = y^2$, then $f(x, y, z) = E + Fz^3$ and

S. Dineen, *Multivariate Calculus and Geometry*,
Springer Undergraduate Mathematics Series, DOI: 10.1007/978-1-4471-6419-7_1,

$$\frac{\partial f}{\partial z} = \frac{d}{dz}(E + Fz^3) = 3Fz^2 = 3y^2z^2.$$

We now recall concepts and notation from linear algebra. First we define the distance between vectors in $\mathbb{R}^n$. This will enable us to define convergent sequences, open and closed sets, continuous and differentiable functions, and state the fundamental existence theorem for maxima and minima.

If $X = (x_1, \ldots, x_n) \in \mathbb{R}^n$ let $\|X\| = (x_1^2 + \cdots + x_n^2)^{1/2}$ and call $\|X\|$ the *length* (or *norm*) of X. If X and $Y = (y_1, \ldots, y_n)$ are vectors in $\mathbb{R}^n$ then $\|X - Y\|$ is the distance between X and Y. The *inner product* (or *dot product* or *scalar product*) of X and Y, $X \cdot Y$ or $\langle X, Y\rangle$, is defined as

$$\langle X, Y\rangle = X \cdot Y = \sum_{i=1}^{n} x_i y_i.$$

We have $\|X\|^2 = \langle X, X\rangle$ and two vectors X and Y are *perpendicular* if and only if their inner product is zero.

For $1 \le j \le n$, let $\mathbf{e}_j = (0, \ldots, 1, 0, \ldots)$, where 1 lies in the jth position. The set $(\mathbf{e}_j)_{j=1}^n$ is a *basis*, the standard *unit vector basis*,[1] for $\mathbb{R}^n$. If $X = (x_1, \ldots, x_n) \in \mathbb{R}^n$ then

$$X = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n = \sum_{i=1}^{n} x_i\mathbf{e}_i = \sum_{j=1}^{n}\langle X, \mathbf{e}_j\rangle\mathbf{e}_j.$$

A mapping $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is linear if

$$T(aX + bY) = aT(X) + bT(Y)$$

for all $X, Y \in \mathbb{R}^n$ and all $a, b \in \mathbb{R}$. For $1 \le i \le m$ and $1 \le j \le n$ let $a_{i,j} = \langle T(\mathbf{e}_j), \mathbf{e}_i\rangle$. If $X = (x_1, \ldots, x_n)$ then, interchanging the order of summation, we obtain

$$\begin{aligned}
T(X) = T\Big(\sum_{j=1}^{n} x_j\mathbf{e}_j\Big) &= \sum_{j=1}^{n} x_j T(\mathbf{e}_j)\\
&= \sum_{j=1}^{n} x_j \Big(\sum_{i=1}^{m}\langle T(\mathbf{e}_j), \mathbf{e}_i\rangle\mathbf{e}_i\Big)\\
&= \sum_{i=1}^{m}\Big\{\sum_{j=1}^{n} x_j\langle T(\mathbf{e}_j), \mathbf{e}_i\rangle\Big\}\mathbf{e}_i
\end{aligned}$$

[1] We use the same notation for the standard basis in $\mathbb{R}^n$ and $\mathbb{R}^m$. The context tells the dimension of the space involved. Otherwise we would be using more and more unwieldy notation.

$$= \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n} a_{i,j} x_j \right\} \mathbf{e}_i .$$

This shows that $T(X) = A(X)$ where the $m \times n$ matrix $A = (a_{i,j})_{1 \le i \le m, 1 \le j \le n}$ operates on the *column* vector X by matrix multiplication. We may now identify the space of linear mappings from $\mathbb{R}^n$ into $\mathbb{R}^m$ with the space of $m \times n$ matrices, $\mathbb{M}_{m,n}$. To present in a reasonable form the product rule and chain rule (see below) we identify $\mathbb{R}^n$ with $\mathbb{M}_{n,1}$, that is the points in $\mathbb{R}^n$ are considered to be *column* vectors. The reader should however note that, for typographical convenience, this convention is often ignored and points in $\mathbb{R}^n$ are written as *row* vectors. However, in taking derivatives the correct convention should be followed to avoid meaningless expressions.

A subset U of $\mathbb{R}^n$ is *open* if for each $X_0 \in U$ there exists $\varepsilon > 0$ such that[2]

$$\{X \in \mathbb{R}^n : \|X - X_0\| < \varepsilon\} \subset U.$$

A subset A of $\mathbb{R}^n$ is *closed* if its complement is open and a set B is *bounded* if there exists $M \in \mathbb{R}$ such that $\|x\| \le M$ for all $x \in B$. Thus all points in an open set A are surrounded by points from A while any point that can be reached from a closed set B belongs to B. A crucial role in all aspects of calculus, analysis and geometry is played by sets which are both closed and bounded; such sets are said to be *compact*.

Example 1.2 The closed solid sphere $\{(x, y, z) : x^2 + y^2 + z^2 \le 1\}$ and its boundary $\{(x, y, z) : x^2 + y^2 + z^2 = 1\}$ are both compact subsets of $\mathbb{R}^3$ while the solid sphere without boundary $\{(x, y, z) : x^2 + y^2 + z^2 < 1\}$ is an open bounded subset of $\mathbb{R}^3$. A line is a closed unbounded set while every open subset of $\mathbb{R}^3$ is a union of open spheres.

If $(X_k)_{k=1}^{\infty}$ is a sequence of vectors in $\mathbb{R}^n$ and $Y \in \mathbb{R}^n$ then we say that X_k converges to Y as k tends to infinity and write $X_k \longrightarrow Y$ as $k \longrightarrow \infty$ and $\lim_{k\to\infty} X_k = Y$ if

$$\|X_k - Y\| \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty.$$

Convergence in $\mathbb{R}^n$ is thus reduced to convergence in $\mathbb{R}$. Moreover, if $X_k = (x_1^k, \ldots, x_n^k)$ and $Y = (y_1, \ldots, y_n)$ then

$$\lim_{k\to\infty} X_k = Y \Longleftrightarrow \lim_{k\to\infty} x_i^k = y_i, \quad 1 \le i \le n.$$

A function $F: U \subset \mathbb{R}^n \to \mathbb{R}^m$ is *continuous* at $X_0 \in U$ if for each sequence $(X_k)_{k=1}^{\infty}$ in U

[2] The ε chosen will depend on X_0 and to be rigorous we should indicate this dependence in some way, e.g. by writing ε_{X_0}. This would lead to unnecessarily complicated notation. We hope that this simplification will not be the source of any confusion.

$$\lim_{k\to\infty} X_k = X_0 \Longrightarrow \lim_{k\to\infty} F(X_k) = F(X_0).$$

When this is the case we write $\lim_{X\to X_0} F(X) = F(X_0)$. If F is continuous at all points in U we say F is continuous on U.

We could also define, in different ways, a length function on $\mathbb{M}_{m,n}$, for instance by identifying $\mathbb{M}_{m,n}$ with $\mathbb{R}^{mn}$. Any such standard definition will be equivalent to the following: if $A^k \in \mathbb{M}_{m,n}$ for all k, $A^k = (a^k_{i,j})_{i,j}$ and $A = (a_{i,j})_{i,j}$ then

$$\lim_{k\to\infty} A^k = A \Longleftrightarrow \lim_{k\to\infty} a^k_{i,j} = a_{i,j}$$

for all (i, j), $1 \le i \le m$, $1 \le j \le n$. We may now define for positive integers l, m and n continuous mappings from $U \subset \mathbb{R}^l \longrightarrow \mathbb{M}_{m,n}$.

A function $f{:}A \subset \mathbb{R}^n \to \mathbb{R}$ has a *maximum* on A if there exists $X_1 \in A$ such that $f(X) \le f(X_1)$ for all $X \in A$. We call $f(X_1)$ the maximum of f on A and say that f achieves its maximum on A at X_1. The maximum, if it exists, is unique but may be achieved at more than one point. A point X_1 in A is called a *local maximum* of f on A if there exists $\delta > 0$ such that f achieves its maximum on $A \cap \{X : \|X - X_1\| < \delta\}$ at X_1. If, in addition, $f(X) < f(X_1)$ whenever $X \neq X_1$ we call X_1 a *strict local maximum*. Isolated local maxima are strict, i.e. if for some $\delta > 0$, X_1 is the only local maximum of f in $A \cap \{X : \|X - X_1\| < \delta\}$ then X_1 is a strict local maximum. In particular, if the set of local maxima of f is *finite* then all local maxima are strict local maxima. The analogous definitions of minimum, local minimum and strict local minimum are obtained by reversing the above inequalities.

Compact sets and continuity feature in the following *fundamental existence theorem for maxima and minima.*

Theorem 1.3 *A continuous real-valued function defined on a compact subset of* $\mathbb{R}^n$ *has a maximum and a minimum.*

The practical problem of finding maxima and minima often requires a degree of smoothness finer than continuity called differentiability. Continuity and differentiability of most functions we encounter can be verified by using functions from $\mathbb{R}$ into $\mathbb{R}$, addition, multiplication, composition of functions and linear mappings.

Definition 1.4 Let $F{:}U \subset \mathbb{R}^n \to \mathbb{R}^m$, U open, and let $X_0 \in U$. We say that F is differentiable at $X_0 \in U$ if there exists a function $A : U \longrightarrow \mathbb{M}_{m,n}$ which is continuous at X_0 such that

$$F(X) = F(X_0) + A(X)(X - X_0) \tag{1.1}$$

for all $X \in U$.

If $f{:}(a, b) \subset \mathbb{R} \longrightarrow \mathbb{R}$ is differentiable at $x_0 \in (a, b)$ in the classical sense, that is if

$$f'(x_0) = \lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

is finite, then letting

$$A(x) = \begin{cases} (f(x) - f(x_0))/(x - x_0), & \text{if } x \neq x_0, \\ f'(x_0), & \text{if } x = x_0, \end{cases}.$$

we see that f is also differentiable in the sense of Definition 1.4. By using (1.1) we see that the converse is also true.

The function A in Definition 1.4 is not necessary unique, However, if the matrix-valued functions A and B satisfy $F(X) = F(X_0) + A(X)(X - X_0)$ and $F(X) = F(X_0) + B(X)(X - X_0)$ and both are continuous at X_0 then

$$\lim_{X \to X_0} A(X) = \lim_{X \to X_0} B(X)$$

and we call the common value the derivative of F at X_0 and denote it by both $DF(X_0)$ and $F'(X_0)$. If f is a scalar-valued function we also write $\nabla f(X)$ in place of $f'(X)$ and call ∇f the *gradient* of f. Gradient is just another word for *slope* and, in one variable calculus the derivative is the slope of the tangent line to the graph of the function. Note that if $F\colon \mathbb{R}^n \to \mathbb{R}^m$ then $F'(X)$ is an $m \times n$ matrix (i.e. the order is reversed) and is identified with a linear mapping from $\mathbb{R}^n$ into $\mathbb{R}^m$. The term

$$F(X_0) + F'(X_0)(X - X_0)$$

is our *linear approximation* to $F(X)$ near X_0.

If F is differentiable then F is continuous. The sum, difference and scalar multiple of differentiable functions are differentiable and the product of scalar-valued and vector-valued differentiable functions are differentiable. We have (see also Exercise 1.7) the following formulae whenever $F, G : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ are differentiable, $a, b \in \mathbb{R}$, $A \in \mathbb{R}^m$, $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is linear and $(A + T)(X) = A + T(X)$ for all $X \in \mathbb{R}^m$:

(i) $(aF + bG)'(X) = aF'(X) + bG'(X)$,

(ii) $(f \cdot G)'(X) = f(X) \cdot G'(X) + G(X) \circ \nabla f(X)$,

(iii) $(A + T)'(X) = T$.

Part (ii) above is the product rule for differentiation. Note that $\cdot$ denotes scalar multiplication, while $\circ$ denotes matrix multiplication of the $m \times 1$ matrix $G(X)$ and the $1 \times n$ matrix $\nabla f(X)$. Part (iii) tells us that the derivative of a linear function is constant and equal to itself at all points while the derivative of a constant function is 0 (in the appropriate space).

The composition of differentiable functions is again differentiable and the *chain rule*, (1.2), gives the derivative of the composition in an elegant form. Let

$$F : U\ (\text{open}) \subset \mathbb{R}^n \to \mathbb{R}^m, \ \ G : V\ (\text{open}) \subset \mathbb{R}^m \to \mathbb{R}^p$$

denote differentiable functions. If $X \in U$ and $F(X) \in V$ then $(G \circ F)'(X)$ exists and

$$D(G \circ F)(X) = (G \circ F)'(X) = G'(F(X)) \circ F'(X) = DG(F(X)) \circ DF(X). \quad (1.2)$$

If $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ is differentiable at $X = (x_1, \ldots, x_n) \in U$ then $\nabla f(X)$ is a $1 \times n$ matrix, that is a row matrix with n entries. If $\mathbf{e}_i \in \mathbb{R}^n$ and $g_i(x) = X + x\mathbf{e}_i$ for all $x \in \mathbb{R}$ then g_i is differentiable and $g_i'(x) = [g_i'(x)](\mathbf{e}_1) = \mathbf{e}_i$ at all $x \in \mathbb{R}$. Since

$$f \circ g_i(x) = f(X + x\mathbf{e}_i) = f(x_1, \ldots, x_{i-1}, x_i + x, \ldots, x_n)$$

the composition $f \circ g_i : \mathbb{R} \longrightarrow \mathbb{R}$ is differentiable and

$$\begin{aligned} \frac{d(f \circ g_i)}{dx}(X) &= \lim_{x \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + x, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i, \ldots, x_n)}{x} \\ &= \frac{\partial f}{\partial x_i}(X) = \nabla f(g(0))(g_i'(0)) = \nabla f(X)(\mathbf{e}_i). \end{aligned}$$

This shows that

$$\nabla f(X) = Df(X) = f'(X) = \left(\frac{\partial f}{\partial x_1}(X), \ldots, \frac{\partial f}{\partial x_n}(X)\right)$$

and, if $\omega = \sum_{j=1}^n \omega_j \mathbf{e}_j \in \mathbb{R}^n$,

$$\nabla f(X)(\omega) = \nabla f(X)\Big(\sum_{j=1}^n \omega_j \mathbf{e}_j\Big) = \sum_{j=1}^n \omega_j \frac{\partial f}{\partial x_j}(X).$$

If $F: U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ then $F(X)$ has m coordinates $f_1(X), \ldots, f_m(X)$ and we often write $F = (f_1, \ldots, f_m)$ where each f_i is a real-valued function of n variables. Hence, $F = \sum_{i=1}^m f_i \mathbf{e}_i$ and F is differentiable if and only if each f_i is differentiable. If $\omega = \sum_{j=1}^n \omega_j \mathbf{e}_j \in \mathbb{R}^n$ then,

$$F'(X)\Big(\sum_{j=1}^n \omega_j \mathbf{e}_j\Big) = \sum_{i=1}^m \Big\{\nabla f_i(X)\Big(\sum_{j=1}^n \omega_j \mathbf{e}_j\Big)\Big\}\mathbf{e}_i = \sum_{i=1}^m \Big\{\sum_{j=1}^n \omega_j \frac{\partial f_i}{\partial x_j}(X)\Big\}\mathbf{e}_i$$

and hence, for $1 \le j \le n, 1 \le i \le n$, we have $\langle F'(X)\mathbf{e}_j, \mathbf{e}_i \rangle = \dfrac{\partial f_i}{\partial x_j}(X)$ and $\dfrac{\partial f_i}{\partial x_j}(X)$ is the (i, j) entry in the $m \times n$ matrix $DF(X) = F'(X)$.

If $\mathbf{v} \in \mathbb{R}^n$ then the function $G : x \in \mathbb{R} \longrightarrow X + x\mathbf{v} \in \mathbb{R}^n$ is differentiable at any $X \in \mathbb{R}^n$ and, as above, $G'(x) = G'(x)(\mathbf{e}_1) = \mathbf{v}$ for all $x \in \mathbb{R}$. If $F : U \text{ (open)} \in \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at X then $(F \circ G)'(0) = F'(G(0)) \circ G'(0) =$

$F'(X)(\mathbf{v})$. We write $D_{\mathbf{v}}F(X)$ and $F_{\mathbf{v}}(X)$ in place of $D(F \circ G)(0)$ and call $D_{\mathbf{v}}F(X)$ the *directional derivative* of F at X in the direction $\mathbf{v}$. Since

$$F(X + x\mathbf{v}) = F(G(x)) = F(G(0)) + A(x)x = F(X) + A(x)x$$

where $\lim_{x\to 0} A(x) = D_{\mathbf{v}}F(X)$ we have

$$D_{\mathbf{v}}F(X) = \lim_{x\to 0} \frac{F(X + x\mathbf{v}) - F(X)}{x}. \tag{1.3}$$

If $\mathbf{v} = \mathbf{e}_i$ we write $\dfrac{\partial F}{\partial x_i}(X)$ and $F_{x_i}(X)$ in place of $D_{\mathbf{e}_i}F(X)$ and call these the (first-order) *partial derivatives* of F.

If F is differentiable at X then all (first-order) directional and partial derivatives of F exist and we have shown

$$F'(X) = \begin{pmatrix} \nabla f_1(X) \\ \vdots \\ \nabla f_m(X) \end{pmatrix} = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(X) & \dots & \dfrac{\partial f_1}{\partial x_n}(X) \\ \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(X) & \dots & \dfrac{\partial f_m}{\partial x_n}(X) \end{pmatrix} = \big(F_{x_1}(X), \dots, F_{x_n}(X)\big).$$

If $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$ this shows that

$$D_{\mathbf{v}}F(X) = F'(X)(\mathbf{v}) = \sum_{i=1}^{n} v_i F_{x_i}(X) = \sum_{i=1}^{n} v_i \frac{\partial F}{\partial x_i}(X).$$

If all first-order partial derivatives of F exist and are continuous then F is differentiable. This criterion enables us to see, literally at a glance, the sets on which most functions have derivatives, partial derivatives and directional derivatives. Consider, for instance, the function

$$F(x, y, z) = \big(\sin(xyz), x^2 - y^2, \exp(xy)\big).$$

We have

$$F(x, y, z) = \sin(xyz)\mathbf{e}_1 + (x^2 - y^2)\mathbf{e}_2 + \exp(xy)\mathbf{e}_3$$

and it suffices to consider separately the three $\mathbb{R}$-valued functions $\sin(xyz)$, $x^2 - y^2$ and $\exp(xy)$. Since linear functions are continuous and the composition and product of continuous functions are continuous we see, on calculating the partial derivatives, as in Example 1.1, that all the first order partial derivatives exist and are continuous. Hence F is differentiable.

Higher-order directional and partial derivatives are defined in the usual way, i.e. by repeated differentiation. If $F : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is differentiable then each *row* in $F'(X)$ corresponds to a *coordinate function* of F while each *column* corresponds to an (independent) *coordinate variable*.

Example 1.5 Let $F{:}\mathbb{R}^4 \to \mathbb{R}^3$ be defined by

$$F(x, y, z, w) = (x^2y, xyz, x^2 + y^2 + zw^2).$$

Then $F = (f_1, f_2, f_3)$ where $f_1(x, y, z, w) = x^2y$, $f_2(x, y, z, w) = xyz$ and $f_3(x, y, z, w) = x^2 + y^2 + zw^2$. Moreover, $\nabla f_1(x, y, z, w) = (2xy, x^2, 0, 0)$, $\nabla f_2(x, y, z, w) = (yz, xz, xy, 0)$ and $\nabla f_3(x, y, z, w) = (2x, 2y, w^2, 2zw)$. Hence

$$F'(x, y, z, w) = \begin{pmatrix} 2xy & x^2 & 0 & 0 \\ yz & xz & xy & 0 \\ 2x & 2y & w^2 & 2zw \end{pmatrix}.$$

If $X = (1, 2, -1, -2)$ and $\mathbf{v} = (0, 1, 2, -2)$ then

$$D_{\mathbf{v}}F(X) = F'(X) \circ {}^t\mathbf{v} = \begin{pmatrix} 4 & 1 & 0 & 0 \\ -2 & -1 & 2 & 0 \\ 2 & 4 & 4 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 2 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix}.$$

Associated with any function $F{:}U \subset \mathbb{R}^n \to \mathbb{R}^m$ we have two types of sets which play a special role in the development of the subject—graphs and level sets. The *graph* of F is the subset of $\mathbb{R}^{n+m}$ defined as follows

$$\begin{aligned} \text{graph}\,(F) &= \big\{(X, Y) : X \in U \text{ and } Y = F(X)\big\} \\ &= \big\{(X, F(X)) : X \in U\big\}. \end{aligned}$$

If $C = (c_1, \ldots, c_m)$ is a point in $\mathbb{R}^m$ we define the *level set* of F, $F^{-1}(C)$, by the formula

$$F^{-1}(C) = \big\{X \in U : F(X) = C\big\}.$$

In terms of the coordinate functions $f_1, \ldots, f_m$ of F we have

$$F(X) = C \iff f_i(X) = c_i \quad \text{for } i = 1, \ldots, m$$

and hence

$$F^{-1}(C) = \bigcap_{i=1}^{m} \big\{X \in U : f_i(X) = c_i\big\} = \bigcap_{i=1}^{m} f_i^{-1}(c_i).$$

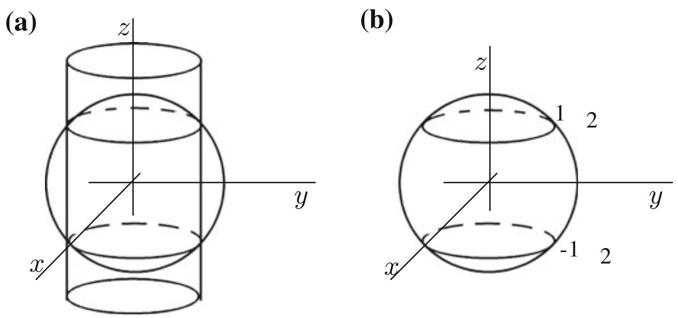

Fig. 1.1

Thus a level set of a vector-valued function is the finite intersection of level sets of real-valued functions. This is frequently useful in arriving at a geometrical interpretation of level sets as the following example shows.

Example 1.6 Let $F(x, y, z) = (x^2 + y^2 + z^2 - 1, 2x^2 + 2y^2 - 1)$. We have $F = (f_1, f_2)$ where $f_1(x, y, z) = x^2+y^2+z^2-1$ and $f_2(x, y, z) = 2x^2+2y^2-1$. The set $f_1^{-1}(0)$ is a *sphere* of radius 1 while $f_2^{-1}(0)$ is a *cylinder* parallel to the z–axis built on a circle with centre the origin and radius $1/\sqrt{2}$. If $\mathbf{0} = (0, 0)$ is the origin in $\mathbb{R}^2$ then

$$F^{-1}(\mathbf{0}) = f_1^{-1}(0) \cap f_2^{-1}(0)$$

is the intersection of a sphere and a cylinder in $\mathbb{R}^3$ (Fig. 1.1a).

For this particular example we obtain more information by solving the equations $f_1(x, y, z) = f_2(x, y, z) = 0$. We have $x^2 + y^2 = 1 - z^2 = 1/2$. Hence $z^2 = 1/2$, $z = \pm 1/\sqrt{2}$ and the level set consists of two circles on the unit sphere (Fig. 1.1b).

The relationship between graphs and level sets plays an important role in our study. The easy part of this relationship—every graph is a level set—is given in the next example while the difficult part—every (regular) level set is locally a graph—is the *implicit function theorem* (Chap. 2).

Example 1.7 Let $F: U \subset \mathbb{R}^n \to \mathbb{R}^m$. We define $G: U \times \mathbb{R}^m \to \mathbb{R}^m$ by $G(X, Y) = F(X) - Y$. If $\mathbf{0}$ is the origin in $\mathbb{R}^m$ then

$$\begin{aligned}(X, Y) \in G^{-1}(\mathbf{0}) &\iff G(X, Y) = \mathbf{0} \\ &\iff F(X) - Y = \mathbf{0} \\ &\iff (X, Y) \in \text{graph}\,(F).\end{aligned}$$

Hence $G^{-1}(\mathbf{0}) = \text{graph}\,(F)$ and every graph is a level set.

Exercises

1.1. Sketch the following sets

(a) $\left\{(x, y, z) : \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} < 1\right\}$
(b) $\{(x, y, z) : x^2 + y^2 = z^2\}$
(c) $\{(x, y, z) : x \geq 0,\ y \geq 0,\ z \geq 0\}$
(d) $\{(x, y, z) : x^2 + y^2 + z^2 - 2z = 0\}$
(e) $\{(x, y, z) : x^2 + y^2 + z^2 - 4z = 0,\ x^2 + y^2 = 4\}$.

From your sketches determine which of the sets are: open, closed, bounded, compact.

1.2. Find all first-order partial derivatives of

(a) $f(x, y, z) = (z^2 + x^2)\log(1 + x^2y^2)$
(b) $g(x, y, z) = xy\tan^{-1}(xz)$
(c) $h(x, y, z, w) = \dfrac{\sin(x^2 + y^2 + z^2 + w^2)}{1 + (x - y)^2}$.

1.3. If $F(x, y, z, w) = (x^2 - y^2, 2xy, zx, z^2w^2x^2)$ and $\mathbf{v} = (2, 1, -2, -1)$ find $F'(1, 2, -1, -2)$ and $D_{\mathbf{v}}F(1, 2, -1, -2)$.

1.4. Let $f(x, y, z) = x^2 - xy + yz^3 - 6z$. Find all points (x, y, z) such that $\nabla f(x, y, z) = (0, 0, 0)$.

1.5. If $f(x, y, z) = x^2e^y$ and $g(x, y, z) = y^2e^{xz}$ find ∇f, ∇g and $\nabla(fg)$. Verify that $\nabla(fg) = f\nabla g + g\nabla f$.

1.6. Let $P\colon (a, b) \subset \mathbb{R} \to \mathbb{R}^n$. Show that

$$\frac{d}{dt}\left(\|P(t)\|^2\right) = 2P(t) \circ P'(t) = 2\langle P(t), P'(t)\rangle$$

and deduce that if $\|P(t)\|$ does not depend on t then $P'(t) \perp P(t)$.

1.7. If $F(x, y, z) = (x^2, y^2 + z^2, xyz)$ and $G(x, y, z) = (e^x, y^2 - z^2, xyz)$ find $F'(x, y, z)$ and $G'(x, y, z)$. Let $H(x, y, z) = \langle F(x, y, z), G(x, y, z)\rangle$ where $\langle\ ,\ \rangle$ is the inner product in $\mathbb{R}^3$. Find $\nabla H(x, y, z)$ and verify that

$$\nabla H(x, y, z) = G(x, y, z) \circ F'(x, y, z) + F(x, y, z) \circ G'(x, y, z).$$

1.8. If $f(x, y, z) = (x^2 + y^2 + z^2)^{-1/2}$ show that

$$\nabla f(x, y, z) = -\frac{(x, y, z)}{(x^2 + y^2 + z^2)^{3/2}}.$$

1.9. If $F:U \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable and $F(P) \neq 0$ show that $\|F\|$ is differentiable at the point P. If $\mathbf{v} \in \mathbb{R}^n$ show that

$$\nabla_{\mathbf{v}}(\|F\|)(P) = \frac{\langle D_{\mathbf{v}} F, F \rangle}{\|F\|}(P).$$

1.10. Use Exercise 1.9 to give another proof of Exercise 1.8, i.e. first find $D_{e_i}\left(\frac{1}{\|x\|}\right)$ for $i = 1, 2, 3$.

1.11. If

$$\begin{array}{c} (x, y, z) \xrightarrow{F} (xyz, x^2 + y^2, x^2 - y^2, z^2) \\ \| \\ (u, v, w, t) \xrightarrow{G} (u^2 + v^2, u^2 - v^2, w^2 - t^2, w^2 + t^2) \end{array}$$

$G = (G_1, G_2, G_3, G_4)$ and $H = (H_1, H_2, H_3, H_4) = G \circ F$ verify that

$$\frac{\partial H_2}{\partial x} = \frac{\partial G_2}{\partial u} \cdot \frac{\partial u}{\partial x} + \frac{\partial G_2}{\partial v} \cdot \frac{\partial v}{\partial x} + \frac{\partial G_2}{\partial w} \cdot \frac{\partial w}{\partial x} + \frac{\partial G_2}{\partial t} \cdot \frac{\partial t}{\partial x}$$

directly and also by using $H' = G' \circ F'$.

1.12. If the function $f:\mathbb{R}^n \to \mathbb{R}$ satisfies

$$\sum_{i=1}^{n} x_i^2 \frac{\partial^2 f}{\partial {x_i}^2} + \sum_{i=1}^{n} x_i \frac{\partial f}{\partial x_i} = 0$$

show that $h(x_1, \ldots, x_n) = f(\mathrm{e}^{x_1}, \ldots, \mathrm{e}^{x_n})$ satisfies

$$\sum_{i=1}^{n} \frac{\partial^2 h}{\partial {x_i}^2} = 0.$$

1.13. Let $f:\mathbb{R}^n \to \mathbb{R}$ be differentiable, let $X \in \mathbb{R}^n$ and $\Delta X = (\Delta x_1, \ldots, \Delta x_n) \in \mathbb{R}^n$. Show that there exists $g:\mathbb{R}^{2n} \to \mathbb{R}$ such that

$$f(X + \Delta X) = f(X) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(X)\Delta x_i + g(X, \Delta X)\|\Delta X\|$$

and $g(X, \Delta X) \to 0$ as $\Delta X \to 0$.

1.14. Let $f(x, y, z) = x^2y^2 + y^2z^2 + xyz$. By using the previous exercise and the values of f and its first-order derivatives at $(1, 1, 1)$ estimate $f(1.1, 1.05, 0.95)$. Find the error in your approximation and the error as a percentage of $f(1, 1, 1)$.

1.15. Identify geometrically and sketch the level set $F^{-1}(C)$ where $F(x, y, z) = (z^2 - x^2 - y^2, 2x - y)$ and $C = (1, 2)$.

1.16. If A is a subset of $\mathbb{R}^n$ show that A is closed if and only if for all $X_0 \in \mathbb{R}^n$ and $(X_k)_{k=1}^{\infty} \subset A$, $\lim_{k\to\infty} X_k = X_0$ implies $X_0 \in A$.

1.17. If A is a subset of $\mathbb{R}^n$ show that A is open if and only if for each $X_0 \in A$ and $(X_k)_{k=1}^{\infty} \subset \mathbb{R}^n$, $\lim_{k\to\infty} X_k = X_0$ there exists a positive integer k_0 such that $X_k \in A$ for all $k \geq k_0$.

1.18. Show that the sum, product and quotient of continuous (respectively differentiable) functions are continuous (respectively differentiable) and that differentiable functions are continuous.

1.19. Show that the level set

$$y^6 + y^2x^2 + x^2 + y^2 + 16z^2 - 8xyz - 2xy = 51$$

is a compact subset of $\mathbb{R}^3$.

1.20. If $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ is differentiable and $F : \mathbb{R}^3 \to \mathbb{R}^4$ is defined by $F(x, y, z) = (x, y, z, \phi(x, y, z))$ find F'.

1.21. If A is a symmetric $m \times n$ matrix and X and Y are eigenvectors corresponding to different eigenvalues show that $\langle X, Y \rangle = 0$.