Chapter 4
Functions of Several Variables: Higher derivatives

The Theorem of Hermann Amandus Schwarz

The general principle of higher derivatives involves, as we shall see in this chapter, a certain understanding of multilinear algebra. We shall cross the bridge of its elementary principles when we get to it.

But let us begin more modestly with second derivatives, where the situation is still simple. So let us assume that X is an open set in $V = \mathbb{R}^n$ and that $f\colon X \to W = \mathbb{R}^m$ is a differentiable function. Then, by definition, the derivative $f'(x)\colon V \to W$ exists for all $x \in X$ and is a linear map, that is, an element of $\mathrm{Hom}(V, W)$. Thus we have before us the differential form

$$(1) \qquad f'\colon X \to \mathrm{Hom}(V, W) \cong M_{mn}(\mathbb{R}) \cong \mathbb{R}^{mn}; \quad f'(x)\colon V \to W \text{ linear.}$$

The vector space $\mathrm{Hom}(V, W)$ is a Banach space with respect to the operator norm (cf. 1.33).

If f' is differentiable, then for each $x \in X$ the function $f''(x)\colon V \to \mathrm{Hom}(V, W)$ is a linear map $V \to \mathrm{Hom}(V, W)$, i.e. a member of $\mathrm{Hom}\big(V, \mathrm{Hom}(V, W)\big)$ which, after a selection of bases, we may consider as a linear map $\mathbb{R}^n \to \mathbb{R}^{mn}$; it is therefore determined by a $mn \times n$ matrix with mn^2 entries. The special case $n = 1$ was easy enough, because the first derivative of a curve is a curve and so the second derivative is again the derivative of a curve.

Let us first consider the special case $m = 1$, that is the case of a level function $f\colon X \to \mathbb{R}$ on an open subset X of $E \cong \mathbb{R}^n$. We recall that this is not such a great restriction since a function into $W \cong \mathbb{R}^m$ is not more than an m-tuple of level functions. Now the derivative $D_a f$ was identified with a vector, the gradient, via the inner product $(\cdot|\cdot)$ on E such that $(D_a f)(v) = (\mathrm{grad}_a f | v)$ for all $v \in E$ (cf. 6.68 ff.). The function $x \mapsto \mathrm{grad}_x f\colon X \to E = \mathbb{R}^n$ may be differentiable; if it is, then the second derivative at x, as the derivative of $x \mapsto \mathrm{grad}_x f = \big((\partial_1 f)(x), \ldots, (\partial_n f)(x)\big)$, is a linear map $H(x)\colon E \to E$ such that

$$\mathrm{grad}_{a+h} f = \mathrm{grad}_a f + H(x)(h) + \|h\| R(h) \text{ with } R(h) \to 0 \text{ for } h \to 0.$$

For all $u \in E$ this gives

$$(\mathrm{grad}_{a+h} f \mid u) = (\mathrm{grad}_a f \mid u) + (H(x)(h) \mid u) + \|h\| (\mathbb{R}(h) \mid u).$$

Thus as a member of $\mathrm{Hom}\big(E, \mathrm{Hom}(E, \mathbb{R})\big)$, for $v, w \in E$ we have $f''(x)(v)(w) = (H(x)v \mid w)$. This means that $f''(x)$ may be considered as the *bilinear form* $(v, w) \mapsto (H(x)v \mid w)$.

As $E = \mathbb{R}^n$, then the linear map $H(x)$ is canonically given by a matrix, namely, in view of Theorem 6.49, by the square matrix $\big((\partial_k\partial_j f)(x)\big)_{j,k=1,\dots,n}$. This means that we have to deal with *mixed partial derivatives* $x \mapsto (\partial_k\partial_j f)(x) : X \to \mathbb{R}$. In general, we do not have

$$(\partial_k\partial_j f)(x) = (\partial_j\partial_k f)(x).$$

However, there are, fortunately, situations where we can make this conclusion. It should be clear at this early stage that this is an important result.

COMMUTING PARTIAL DERIVATIVES

Theorem 4.1. (H. A. Schwarz) *For a twice* continuously *differentiable function* $f\colon X \to \mathbb{R}$ *on an open set of* $\mathbb{R}^n$ *we have*

$$(\forall j,\ k = 1, \dots, n;\ x \in X)\ (\partial_k\partial_j f)(x) = (\partial_j\partial_k f)(x). \tag{2}$$

The matrix of $f''(x)$ *is symmetric for all* $x \in X$. □

From the formulation of this theorem we see that the core of the matter concerns two variables, since for the computation of the j-th and k-th variable all other $n-2$ variables remain constant. It proof will readily follow from a more general theorem which we begin to discuss in the following.

In order to understand the situation we consider a function $f\colon X \to \mathbb{R}$, $X \subseteq E \cong \mathbb{R}^n$ and consider two directional derivatives determined by two linearly independent vectors e_1 and e_2, and we assume as usual that a is an interior point of X. We proceed from a by s units to $a + s{\cdot}e_1$ and from there by t units to $a + s{\cdot}e_1 + t{\cdot}e_2$; we focus on the difference $g_t(s) = f(a + s{\cdot}e_1 + t{\cdot}e_2) - f(a + s{\cdot}e_1)$ which obviously also depends on t.

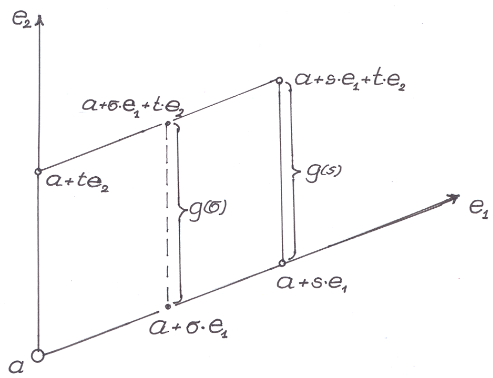

Figure 4.1

Now we are interested in the difference

$$\begin{aligned} G(s) \stackrel{\text{def}}{=} g_t(s) - g_t(0) &= (f(a + s{\cdot}e_1 + t{\cdot}e_2) - f(a + s{\cdot}e_1)) - (f(a + t{\cdot}e_2) - f(a)) \\ &= (f(a + s{\cdot}e_1 + t{\cdot}e_2) - f(a + t{\cdot}e_2)) - (f(a + s{\cdot}e_1) - f(a)). \end{aligned}$$

If we define functions h_t on some neighborhood of 0 in $\mathbb{R}$ by $h_t(s) \stackrel{\text{def}}{=} f(a + s{\cdot}e_1 + t{\cdot}e_2) - f(a + t{\cdot}e_2)$, then we have

$$G(s) = g_t(s) - g_t(0) = h_t(s) - h_0(s).$$

Now let us make the following assumption

(**A.1**) The directional derivatives $p(u) \stackrel{\text{def}}{=} \partial_{u;e_1} f$ exist in an entire neighborhood of a.

Let U be an open ball neighborhood of a such that (A.1) holds for $u \in U$ and assume that s and t are chosen so that $a + s{\cdot}e_1, a + t{\cdot}e_2, a + s{\cdot}e_1 + t{\cdot}e_2 \in U$. Then for each such t the functions $s' \mapsto h_t(s')$ are differentiable for all sufficiently small $0 \le s' \le s$. Thus G is differentiable for all sufficiently small s; and since $h_t'(s) = p(a + s{\cdot}e_1 + t{\cdot}e_2)$ we have $G'(s) = p(a + s{\cdot}e_1 + t{\cdot}e_2) - p(a + s{\cdot}e_1)$. By the Mean Value Theorem 4.29, of Analysis I, there is a number $\sigma = \sigma(s,t)$ between 0 and s such that $G(s) = G'(\sigma)s$. Thus we obtain

$$G(s) = \big(h_t(s) - h_0(s)\big)s = \big(p(a + \sigma{\cdot}e_1 + t{\cdot}e_2) - p(a + \sigma{\cdot}e_1)\big)s, \tag{3}$$

where $\sigma = \sigma(s,t)$ is between 0 and s.

In order to be able to work further on (3) we make a further assumption

(**A.2**) The directional derivatives $\partial_{x;e_2} p$ exist for all x in a neighborhood of a.

We may assume now that U is an open ball neighborhood of a such that (A.1) holds for $u \in U$ and (A.2) holds for $x \in U$. We continue to assume that s and t were chosen so that $a + s{\cdot}e_1, a + t{\cdot}e_2, a + s{\cdot}e_1 + t{\cdot}e_2 \in U$. Now we apply the Mean Value Theorem again and find a number τ between 0 and t such that

$$p(a + \sigma{\cdot}e_1 + t{\cdot}e_2) - p(a + \sigma{\cdot}e_1) = \big(\partial_{a+\sigma\cdot e_1+\tau e_2;e_2} p\big)t.$$

Thus for all sufficiently small s and t we have the statement

$$G(s) = \big(\partial_{(a+\sigma\cdot e_1+\tau\cdot e_2);e_2} p\big)st, \tag{4}$$

with $\sigma = \sigma(s,t)$ between 0 and s and $\tau = \tau(s,t)$ between 0 and t. We observe that $(s,t) \to (0,0)$ implies $(\sigma,\tau) = \big(\sigma(s,t),\tau(s,t)\big) \to (0,0)$.

At this point we make another decisive assumption; for its formulation we abbreviate the expression $\partial_{x;e_2} p$ by $(\partial_2\partial_1 f)(x)$. Note that $\partial_2\partial_1 f\colon U \to \mathbb{R}$ is a well defined function.

If e_1 and e_2 are the first two standard basis vectors of $\mathbb{R}^n$, then this notation is consistent with the one we introduced in Definition 6.48.

(**A.3**) The function $\partial_2\partial_1 f$ is continuous at a.

Then by (3) and by $\lim_{(s,t)\to(0,0)}(\sigma,\tau) = (0,0)$ we get

$$G(s) = (\partial_2\partial_1 f)(a)st + stR(s,t) \tag{4}$$

with a remainder function R satisfying $\lim_{(s,t)\to(0,0)} R(s,t) = 0$.

On the other hand we could produce more information on $g_t(s) = f(a + se_1 + te_2) - f(a + s{\cdot}e_1)$, if we could work with directional derivatives in in the direction of of e_2. Therefore we demand

(**A.4**) The directional derivatives $q(u) \stackrel{\text{def}}{=} \partial_{u;e_2} f$ exist in all points $u = a + se_1$ with sufficiently small s.

Now this means, specifically, that for all sufficiently small s we have remainder functions R_s such that $\lim_{t\to 0} R_s(t) = 0$ and that

$$\begin{aligned} g_t(s) =& f(a + s{\cdot}e_1 + t{\cdot}e_2) - f(a + s{\cdot}e_2) \\ =& \big(\partial_{a+s\cdot e_1;e_2} f\big){\cdot}t + tR_s(t) = q(a + s{\cdot}e_1)t + tR_s(t), \\ g_t(0) =& f(a + t{\cdot}e_2) - f(a) \\ =& \big(\partial_{a;e_2} f\big){\cdot}t + tR_0(t) = q(a)t + tR_0(t). \end{aligned}$$

Thus, expressing $G(s) = g_t(s) - g_t(0)$ from these formulae on the one hand and from (4) on the other, we get the relation

$$q(a + s{\cdot}e_1) - q(a) = (\partial_2\partial_1 f)(a)s + sR(s,t) - R_s(t) + R_0(t) \tag{5}$$

for all sufficiently small s and t.

Now assume that we are given an $\varepsilon > 0$. Then we choose $\delta > 0$ according to (4) in such a fashion that, $|s|, |t| < \delta$ implies $|R(s,t)| < \varepsilon$. For these s and t we have

$$|q(a + s{\cdot}e_2) - q(a) - (\partial_2\partial_1 f)(a)s| \le \varepsilon|s| + |R_s(t)| + |R_0(t)|. \tag{6}$$

We recall $R_s(t)$, $R_0(t) \to 0$ for $t \to 0$ and all s. Thus for each $|s| < \delta$ we let t tend to 0 in (6) and thus finally find that

$$(\forall\varepsilon > 0)(\exists\delta > 0)\ 0 < |s| < \delta \Rightarrow \left|\frac{q(a + s{\cdot}e_1) - q(a)}{s} - (\partial_2\partial_1 f)(a)\right| \le \varepsilon. \tag{7}$$

Notice that the step from (6) to (7) is a bit tricky. A division by a nonzero s in (6) would be still alright, but an attempt to simultaneously letting (s,t) tend to zero would cause failure because of the term $(|R_s(t)| + |R_0(t)|)/s$. The fixing of s and letting t tend to zero first is therefore essential. This strategy is made possible by the fact that t is no longer present in any of the other terms in (6).

Statement (7) means exactly that q has at a a directional derivative in the direction of e_1, equalling $(\partial_2\partial_1 f)(a)$. We therefore have proved the following theorem, that quickly entails Schwarz' Theorem 4.1—with much room to spare.

Theorem 4.2. *For a function $\varphi\colon X \to \mathbb{R}$, $X \subseteq E \cong \mathbb{R}^n$ and for an inner point a of X and two vectors e_1, e_2 of E we write $(\partial_j\varphi)(x) = \partial_{x,;e_j}\varphi$, $j = 1, 2$ in all points x in which these directional derivatives exist. We assume the following hypotheses hold for a $f\colon X \to \mathbb{R}$:*

(i) *The directional derivative $\partial_1 f(x)$ exists for all x in a neighborhood of a.*

(ii) *The directional derivative $(\partial_2\partial_1 f)(x) \stackrel{\text{def}}{=} \partial_2(\partial_1 f)(x)$ exists for all x in a neighborhood U of a and $\partial_2\partial_1 f\colon U \to \mathbb{R}$ is continuous at $x = a$.*

(iii) *The directional derivative $(\partial_2 f)(a + se_1)$ exists in all points $a + se_1$ with sufficiently small s.*

Then $(\partial_1\partial 2f)(a) \stackrel{\text{def}}{=} \partial_1(\partial_2 f)(a)$ exists and equals $(\partial_2\partial_1 f)(a)$. □

Theorem 4.2 and its proof remain true if e_1 and e_2 are linearly dependent, but the information produced in this special case is a triviality.

Let us look at the following example of a function $f\colon \mathbb{R}^2 \to \mathbb{R}$. We observe that

$$\sin t \cos t(\cos^2 t - \sin^2 t)r^2 = \frac{r^2}{2} \sin 2t \cos 2t = \frac{r^2}{4} \sin 4t$$

and define

$$f(w) = \begin{cases} 0 & \text{for } w = 0, \\ \frac{r^2}{4} \sin 4t & \text{for } w = (r\cos t, r\sin t),\ 0 < r \text{ and } 0 \le t < 2\pi. \end{cases}$$

Then, for $(x, y) \neq (0, 0)$ we may write

$$f(x, y) = xy(x^2 - y^2)(x^2 + y^2)^{-1}$$

All second partial derivatives exist everywhere, however,

$$\partial_1\partial_2 f(0) = 1 \quad \text{and} \quad \partial_2\partial_1 f(0) = -1.$$

Thus Theorem 4.2 (and therefore also Schwarz's Theorem 4.1) cannot be improved much by weakening the hypotheses. The partial derivatives $\partial_k\partial_j f(a)$ are also written $\partial^2 f/\partial x_k \partial x_j|_{x=a}$, and the linear map $H(a)$ and its matrix are called the *Hessian*, respectively, *Hesse matrix*. The German expression is *Hesse–Matrix*.

The second degree Taylor expansion

We now return to a twice continuously differentiable function $f\colon X \to \mathbb{R}$ where X is an open set of the Hilbert space $E = \mathbb{R}^n$. Let $a \in X$ and assume that for $\delta > 0$ we have $U_\delta(a) \subseteq X$. We again identify $f''(a)$ with the bilinear map $(v, w) \mapsto (H(x)(v) \mid w)$; under the hypotheses Schwarz' Theorem 4.1 we know that this is a *symmetric* bilinear form, that is, $(H(x)(v) \mid w) = \big(v \mid H(x)(w)\big)$; in particular, $v \mapsto (H(x)(v) \mid v) = \big(v \mid H(x)(v)\big)$ is a *quadratic form.* If B is a symmetric bilinear form, we know that the derivative of the function $f\colon E \to \mathbb{R}$ defined by $f(x) = B(x, x)$ is the linear function given by $f(x)(v) = 2B(v, x)$ by Corollary 3.10. The second derivative is therefore given by $f''(x)(v)(w) = 2B(v, w)$, and the Hesse matrix $H(x)$ of f is determined by $(H(x)(v) \mid w) = f''(x)(v)(w) = B(v, w)$. The the matrix elements $b_{jk} = (H(x)e_j \mid e_k)$ of $H(x)$ are precisely the coefficients $B(e_j, e_k)$ of the bilinear map, where as usual the $e_1, \ldots, e_n$ are the standard basis vectors of $\mathbb{R}^n$.

Theorem 4.3. (Taylor's Theorem of degree 2) *Let $f\colon X \to \mathbb{R}$ be a function on an open set X of the Hilbert space $E = \mathbb{R}^n$ with the standard inner product $(x \mid y)$. Assume that the second partial derivatives exist and are continuous in all $x \in X$. Then we have a symmetric linear map $H(a)\colon E \to E$ with coefficient matrix*

$$\left(\frac{\partial^2 f}{\partial x_j \partial x_k}\bigg|_{x=a} \right)_{j,k=1,\ldots,n}$$

depending continuously on $a \in X$ and defining a quadratic form $v \mapsto (H(x)v \mid v)\colon E \to \mathbb{R}$. The first derivative satisfies $f'(x)(v) = (\operatorname{grad}_x f \mid v)$ and the second derivative $f''(x) \in \operatorname{Hom}\big(E, \operatorname{Hom}(E,\mathbb{R})\big)$ satisfies $f''(x)(v)(w) = (H(x)v \mid w)$. Then for each $a \in X$ there is a function $r_a\colon X \to \mathbb{R}$ which is continuous at a and satisfies $r_a(a) = 0$ such that

$$(8)\quad f(x) = f(a) + (\operatorname{grad}_a f \mid x-a) + \frac{1}{2}\big(H(a)(x-a) \mid x-a\big) + \|x-a\|^2 r_a(x).$$

Proof. Since the second derivatives exist, the first partial derivatives are continuous (cf. 3.5); hence the first derivative $f'(x)$ exists for all x by Theorem 3.14. The second partial derivatives are the first partial derivatives of $f'\colon X \to \operatorname{Hom}(E,\mathbb{R})$; since these are assumed to be continuous, $f''(x)$ exists for all $x \in X$ and $f''\colon X \to \operatorname{Hom}\big(E, \operatorname{Hom}(E,\mathbb{R})\big)$ is continuous.

The assertions $f'(x)(v) = (\operatorname{grad}_x f \mid v)$ and $f''(x)(v)(w) = (H(x)v \mid w)$ were proved in the paragraph preceding the theorem.

We now have to establish the existence of r_a such that (8) holds. We simplify matters by considering

$$F(x) = f(x) - f(a) - f'(a)(x-a) - \frac{1}{2}f''(a)(x-a)(x-a).$$

If the assertion is proven for F then it holds for f, but $F(a) = 0$, $F'(a) = 0$, and $F''(a) =$). Thus we may assume without losing generality that the f and its first two derivatives vanish at a. We have to show that $\lim_{x\to a} \|x-a\|^{-2}{\cdot}f(x) = 0$. For this purpose let $\varepsilon > 0$, we have to find a $\delta > 0$ so that $U_\delta(a) \subseteq X$ and such that for $\|x-a\| < \delta$ we have $\|f(x)\| \le \|x-a\|^2\varepsilon$. By hypothesis, the function $f''\colon X \to \operatorname{Hom}(\operatorname{Hom}(\mathbb{R}^n,\mathbb{R}),\mathbb{R})$ is continuous and $f''(a) = 0$. Hence we find a $\delta > 0$ such that $\|f''(x)(e)(e')\| \le \varepsilon$ for all $e, e' \in \mathbb{R}^n$ with $\|e\|, \|e'\| \le 1$. (Another way of saying this is that the operator norm of the Hessian $\|H(x)\|$ is $\le \varepsilon$ for $\|x-a\| < \delta$.) Now assume $0 < \|x-a\| \le \delta$. Set $r = \|x-a\| < \delta$ and let e be the unit vector $r^{-1}{\cdot}\|x-a\|$. We define $\varphi\colon]-\delta,\delta[\to \mathbb{R}$ by $\varphi(t) = f(a+t{\cdot}e)$. Then $\varphi(0) = f(a) = 0$. By the Chain Rule we get $\varphi'(t) = f'(a+t{\cdot}e)(e)$, notably $\varphi'(0) = f'(a)(e) = 0$; and applying the Chain Rule once more we get $\varphi''(t) = f''(a+t.e)(e)(e) = (H(x+t{\cdot}a)(e) \mid e)$, and thus $|\varphi''(t)| \le \varepsilon$ for $|t| < \delta$. Now we apply the second order Taylor Theorem 4.61 to φ and find a $\theta \in [-t,t]$, depending on t such that $\varphi(t) = \varphi(0) + \varphi'(0)t + 1/2\varphi''(\theta)t^2 = t^2/2\varphi''(\theta)$. Therefore $|f(a+t{\cdot}e)| = |\varphi(t)| \le \frac{t^2\varepsilon}{2}$. Setting $t = r$ we obtain $|f(x)| \le \|x-a\|^2{\cdot}\frac{\varepsilon}{2} < \varepsilon\|x-a\|^2$. This had to be shown. □

The second degree Taylor expansion (8) is crucial for a finer investigation of the behavior of a level function f at a critical point a, i.e. a point with $\operatorname{grad}_a f = 0$. Obviously, in a critical point a, the Taylor expansion of degree 2 yields

$$(11)\qquad f(x) - f(a) = \frac{1}{2}(H(a)h \mid h) + \|x-a\|^2 r_a(x); \quad \lim_{x\to a} r_a(x) = 0.$$

If we set $h = x - a$, then

$$(H(a)h \mid h) = \sum_{1 \le j,\, k \le n} (\partial_j \partial_k f) h_j h_k$$

with $(\partial_k \partial_j f) = (\partial_j \partial_k f)$ for all j and k.

In Linear Algebra one deals with the following

Exercise E4.1. Prove:
Proposition. *Let E be a finite dimensional Hilbert space over $K = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. If $\varphi \in \mathrm{Hom}(E, E)$ such that $(\varphi(x)|y) = \big(x|\varphi(y)\big)$ for all $x, y \in E$, then*
(i) *all eigenvalues of φ are real, and*
(ii) *if x is an eigenvector for λ and y an eigenvector for $\kappa \ne \lambda$, then $(x|y) = 0$, that is, x and y are perpendicular.*
(iii) *E has an orthonormal basis $e_1, \dots, e_n$, that is, a basis such that $(e_j|e_k) = \delta_{jk}$ with the Kronecker delta δ_{jk}, consisting of eigenvectors of φ.*
(iv) If $x = \xi_1 \cdot e_1 + \cdots + \xi_n \cdot e_n$ for an orthonormal basis of eigenvectors of φ, then $(\varphi(x)|x) = \sum_{j=1}^n \lambda_j \xi_j^2$.
[Hint. We need recourse to some basic facts on eigenvalues. (i) Since $\mathbb{R}^n \subseteq \mathbb{C}^n$ we may assume $\mathbb{K} = \mathbb{C}$. Let λ be an eigenvalue and let e be a nonzero eigenvector of unit length. Then $\lambda = \lambda(e|e) = (\lambda \cdot e|e) = (\varphi(e)|e) = \big(e|\varphi(e)\big) = (e|\lambda \cdot e) = \overline{\lambda}(e|e) = \overline{\lambda}$. (ii) Let $\varphi(x) = \lambda \cdot x$ and $\varphi(y) = \kappa \cdot y$. Then $\lambda(x|y) = (\varphi(x)|y) = \big(x|\varphi(y)\big) = \kappa(x|y)$, that is, $(\lambda - \kappa)(x|y) = 0$. (iii) Every eigenspace has an orthonormal bases obtained by the *Gram-Schmidt* procedure. The union of these orthonormal bases over all eigenspaces form an orthonormal basis of E. (iv) is now straighforward.]

As a consequence of Exercise E4.1, we obtain the following result.

Proposition 4.4. *Let $f\colon X \to \mathbb{R}$ be a twice continuously differentiable level function on an open subset X of $\mathbb{R}^n$. Assume that $a \in X$ is a critical point. There is an orthonormal basis $e_1, \dots, e_n$ of $\mathbb{R}^n$, real numbers $\lambda_1, \dots, \lambda_n$, and a function $r_a\colon X \to \mathbb{R}$ with $\lim_{x \to a} f(r) = 0$ such that with $x = \sum_{j=1}^n x_j \cdot e_j$ we have*

$$f(x) - f(a) = \frac{1}{2} \sum_{j=1}^{n} \lambda_j (x_j - a_j)^2 + \|x - a\|^2 r(x). \tag{12}$$

□

The quadratic form $h \mapsto (H(a)h, h)$ is positive definite iff $\lambda_j > 0$ for all $j = 1, \dots, n$. If we assume this, then $\|x\|_* \overset{\text{def}}{=} \frac{1}{2}\sqrt{\sum_{j=1}^n \lambda_j x_j^2}$ defines a euclidean norm on E. Since the norm used for the Taylor expansion was arbitrary, we may select the remainder function r in such a fashion that (12) takes the form

$$f(x) - f(a) = \|x - a\|_*^2 + \|x - a\|_*^2 r(x) = \|x - a\|_*^2 \big(1 + r(x)\big). \tag{12'}$$

Since $\lim_{x \to a} r(x) = 0$ we find a $\delta > 0$ such that $\|x - a\|_* < \delta$ implies $x \in X$ and $|r(x)| < 1$. Thus (12′) shows that $f(x) - f(a) > 0$ for $0 < \|x - a\|_* < \delta$. Hence f attains a local minimum at a. Therefore we have

Corollary 4.5. *If, under the hypotheses of* Proposition 4.4, *the quadratic form $H(a)$ is positive (respectively, negative) definite, then f attains in a a local minimum, respectively, maximum.* □

The simple examples of the quadratic function $f(x, y) = x^2 + y^2$:

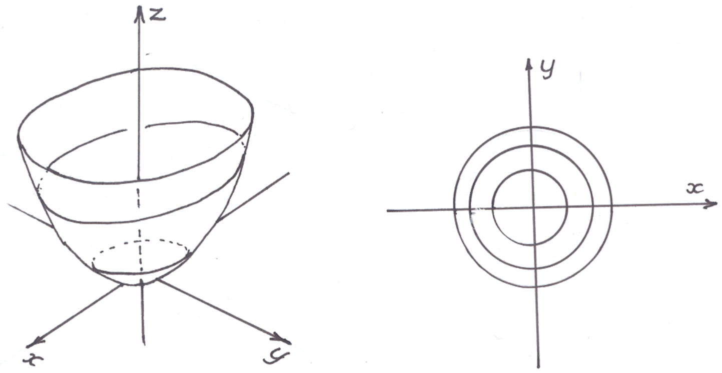

Figure 4.2

or the quadratic function f, $f(x, y) = x^2 - y^2$:

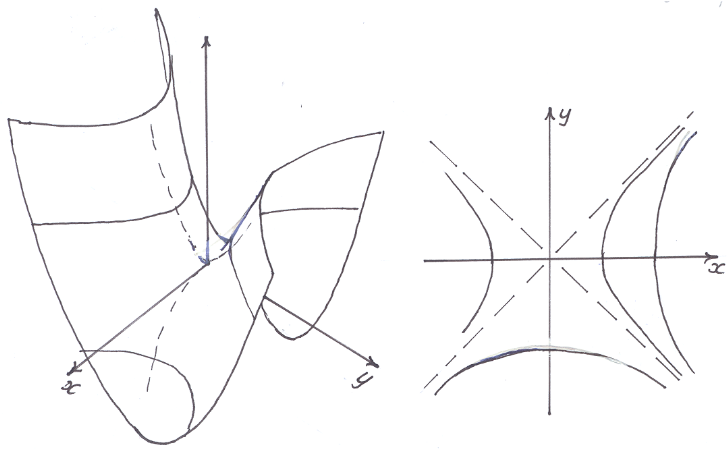

Figure 4.3

or the degenerate quadratic function$f$, $f(x, y) = x^2$:

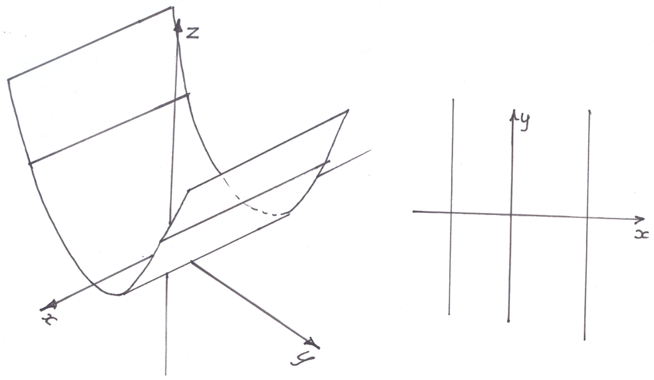

Figure 4.4

Illustrate what happens in the presence and in the absence of definiteness of $f''(0)$ in the critical point 0.

It is a good exercise to draw the corresponding pictures of $f(x,y) = x^2 + y^3$.

Exercise E4.2. Sketch the graph of $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^2 + y^3$ and draw a picture of the level lines.

Higher Derivatives and Theorem of Taylor

The purpose of this section is to treat Taylor's Theorem for level functions in full generality.

We consider an open subset X of a finite dimensional normed space E and a function $f\colon X \to \mathbb{R}$ whose higher derivatives we wish to consider sucessively. We shall assume that successive derivatives exist and are continuous as far as we shall consider them.

First derivative. $f'\colon X \to \mathrm{Hom}(E,\mathbb{R})$.
Second derivative. $f''\colon X \to \mathrm{Hom}\,\big(E, \mathrm{Hom}(E,\mathbb{R})\big)$.
Third derivative. $f^{(3)}\colon X \to \mathrm{Hom}\,\big(E, \mathrm{Hom}\,\big(E, \mathrm{Hom}(E,\mathbb{R})\big)\big)$.
m-th derivative. $f^{(m)}\colon X \to \underbrace{\mathrm{Hom}\,\big(E, \mathrm{Hom}(E \cdots \mathrm{Hom}(E,}_{m\text{ times}}\mathbb{R}\underbrace{))\cdots\big)}_{m\text{ times}}$.

Obviously we have to deal with the iterated Hom-vectorspaces such as they occur as range spaces of the higher derivatives of a level function. Therefore we have to discuss some *multilinear algebra.* We first illustration what we are doing in the case of replacing the n^2-dimensional vector space $\mathrm{Hom}\,\big(E, \mathrm{Hom}(E,\mathbb{R})\big)$ by a more managable one.

An interlude on multilinear algebra

Let $\varphi \in \mathrm{Hom}\,\big(E, \mathrm{Hom}(E,\mathbb{R})\big)$. Thus φ is a linear map $E \to \mathrm{Hom}(E,\mathbb{R})$. That is, for an element $v \in E$, the image $\varphi(v)$ is itself a linear form $\varphi(v)\colon E \to \mathbb{R}$. Specifically, $\varphi(v)(w) \in \mathbb{R}$ for all $w \in E$. Since $\varphi\colon E \to \mathrm{Hom}(E,\mathbb{R})$ is linear, we have $\varphi(t{\cdot}v_1 + v_2) = t{\cdot}\varphi(v_1) + \varphi(v_2)$ fo all $t \in \mathbb{R}$ and $v_1, v_{\in} E$. By the definition of pointwise scalar multiplication and addition of functions this means tha

$$(13)\qquad (\forall t \in \mathbb{R},\, v_1,\, v_2,\, w \in E)\, \varphi(t{\cdot}v_1 + v_2)(w) = t{\cdot}\varphi(v_1)(w) + \varphi(v_2)(w).$$

Since $\varphi(v)\colon E \to \mathbb{R}$ is linear for all v, similarly we have

$$(14)\qquad (\forall t \in \mathbb{R},\, v,\, w_1,\, w_2 \in E)\, \varphi(v)(t{\cdot}w_1 + w_2) = t{\cdot}\varphi(v)(w_1) + \varphi(v)(w_2).$$

Let us define a function $\widetilde{\varphi}\colon E \times E \to \mathbb{R}$ by

$$(15)\qquad \widetilde{\varphi}\colon E \times E \to \mathbb{R}, \quad \widetilde{\varphi}(v, w) = \varphi(v)(w).$$

By (13) and (14), $\widetilde{\varphi}$ is a bilinear function or form. Let us denote by $\mathrm{Hom}^m(E;\mathbb{R})$ the set of all multilinear forms $E \underbrace{\times \cdots \times}_{m\,\text{times}} E \to \mathbb{R}$, that is, maps which are linear in each argument separately if the other arguments are fixed. Then we have $\widetilde{\varphi} \in \mathrm{Hom}^2(E;\mathbb{R})$. As an example, if $v_1, \ldots, v_n \in \mathbb{R}^n$, let $\det(v_1, \ldots, v_n)$ denote the determinant of the matrix whose rows are $v_1, \dot{,} v_n$ in that order, then $\det\colon E^n \to \mathbb{R}$ is an example of a multilinear form. Notice that $\mathrm{Hom}^m(E;\mathbb{R} \subseteq \mathbb{R}^{E^m}$ is closed under pointwise scalar multiplication and under pointwise addition in the vector space $\mathbb{R}^{E^m}$ of *all* functions $E^m \to \mathbb{R}$ and is, therefore, a vector space. Thus (15) defines a function

$$\varphi \mapsto \widetilde{\varphi} : \mathrm{Hom}\,\big(\mathrm{Hom}(E,\mathbb{R})\big) \to \mathrm{Hom}^2(E;\mathbb{R}).$$

Exercise E4.3. Show that $\varphi \mapsto \widetilde{\varphi}$ is linear.
[Hint. Prove for instance that $(\varphi + \psi)^{\sim} = \widetilde{\varphi} + \widetilde{\psi}$.]

Conversely, if $\beta\colon E \times E \to \mathbb{R}$ is a bilinear form, then $\beta(v, \cdot)\colon E \to \mathbb{R}$ is linear and $v \mapsto \beta(v, \dot{})\colon E \to \mathrm{Hom}(E,\mathbb{R})$ is linear as well. We define

$$(16)\qquad \beta^*\colon E \to \mathrm{Hom}(E,\mathbb{R}), \quad \beta^*(v)(w) = \beta(v, w).$$

Exercise E4.4. Show that the function $\beta \mapsto \beta^*$ is linear and that it is an inverse of the function $\varphi \mapsto \varphi^*$.
[Hint. Prove for instance that $(\alpha + \beta)^* = \alpha^* + \beta$. Moreover, show that $\widetilde{\varphi})* = \varphi$ and $\widetilde{\beta*} = \beta$.]

After this exercise we know that $\varphi \mapsto \widetilde{\varphi}\colon \mathrm{Hom}\,\big(\mathrm{Hom}(E,\mathbb{R})\big) \to \mathrm{Hom}^2(E;\mathbb{R})$ is an isomorphism of vector spaces, and is defined quite naturally.

Proposition 4.6. *For each natural number m and each*

$$\varphi \in \underbrace{\mathrm{Hom}\,\big(E, \mathrm{Hom}(E \cdots \mathrm{Hom}(E, \mathbb{R}}_{m\,\text{times}}\, \underbrace{))\cdots)}_{m\,\text{times}}$$

we define $\widetilde{\varphi}(v_1,\dots,v_m) \stackrel{\text{def}}{=} \varphi(v_1)(v_2)\cdots(v_m)$ *for* $v_1,\dots,v_m \in E$*. Then* $\widetilde{\varphi}\colon E^m \to \mathbb{R}$ *is a multilinear form and*

$$\varphi \mapsto \widetilde{\varphi} : \underbrace{\operatorname{Hom}\big(E, \operatorname{Hom}(E \cdots \operatorname{Hom}(E,}_{m\,\text{times}} \mathbb{R} \underbrace{))\cdots)}_{m\,\text{times}} \to \operatorname{Hom}^m(E;\mathbb{R}) \tag{17}$$

is an isomorphism of vector spaces.

Proof . Exercise. □

Exercise E4.5. Prove Proposition 4.6.
[Hint. Either induction, or arguments applying to m arguments completely analogous to those which we went through above for 2 arguments, show that $\widetilde{\varphi}$ is multilinear and that $\varphi \mapsto \widetilde{\varphi}$ is linear. For a multilinear form $\beta\colon E^m \to \mathbb{R}$ define β^* in the domain of $\varphi \mapsto \widetilde{\varphi}$ exactly as it was done for bilinear maps in (16) and show that $\widetilde{\varphi})* = \varphi$ and $\widetilde{\beta^*} = \beta$.]

As a consequence of this interlude "*we may "identify" the two isomorphic vector spaces in* (17) *and therefore consider the m-th derivative* $f^{(m)}(x)$ *of a function* $f\colon X \to \mathbb{R}$ *with* X *open in* E *as a multilinear form with* m *arguments, writing*

$$f^{(m)}(x)(v_1,\dots,v_m) \quad \textit{instead of} \quad f^{(m)}(x)(v_1)(v_2)\cdots(v_m).$$

Remark 4.7. Assume $E = \mathbb{R}^n$. Let $\beta \in \operatorname{Hom}^m(E;\mathbb{R})$, and let $e_1,\dots,e_n$ be the standard basis vectors of $\mathbb{R}^n$,

$$e_1 = (1,0,\dots,0), \quad e_2 = (0,1,0,\dots,0), \dots, e_n = (0,\dots,0,1).$$

Define

$$a_{j_1\cdots j_m} \stackrel{\text{def}}{=} \beta(e_{j_1},\dots,e_{j_m}), \quad 1 \le j_k \le n, \quad k = 1, dots, m.$$

Now take

$$v_1 = v_1^{(1)},\dots,v_n^{(1)},\dots,v_m = (v_1^{(m)},\dots,v_n^{(m)}),$$

then

$$\beta(v_1,\dots,v_m) = \sum_{1 \le j_k \le n} a_{j_1\cdots j_m} v_{j_1}^{(1)} \cdots v_{j_m}^{(m)}. \tag{18}$$

Proof . This is a straightforward exercise. □

Exercise E4.6. Prove Remark 4.7.
[Hint. Write $v_1 = \sum_{j_1=1}^n v_{j_1}^{(1)} e_{j_1}, \dots$ and use multilinearlity.]

This remark shows how multinear maps are handled in a computational fashion. The numbers $a_{j_1\cdots j_m}$ are called the *coefficients of the multilinear form* β. The case $m = 2$ is familiar from linear algebra; the coefficients $a_{j_1 j_2}$ simply form an $n \times n$ matrix.

Since $f^{(m)}$ is exactly such a multilinear form as β in the preceding remark, the question arises what the coefficients $a_{j_1\cdots j_m}$ are in this case.

Proposition 4.8. *Let $f\colon X \to \mathbb{R}$, X open in $\mathbb{R}^m$ be an m times differentiable level function. Then for each $x \in X$, the coefficients $a_{j_1\cdots j_k}$ of the multilinear form $f^{(k)}$, $k = 1, 2, \ldots, m$ are*

$$a_{j_1\cdots j_k} = (\partial_{j_1} \cdots \partial_{j_k} f)(x). \tag{19}$$

Proof . We prove by induction that

$$f^{(k)}(x)(v_1, \ldots, v_k) = \sum_{\substack{1\le j_p\le n\\ 1\le p\le k}} (\partial_{j_1} \cdots \partial_{j_k} f)(x) v^{(1)}_{j_1} \cdots v^{(k)}_{j_k} \tag{20_k}$$

holds for $k = 1, 2, \ldots, m$. For $k = 1$ we know from 7.12 that

$$f'(x)(v) = (\mathrm{grad}_x\, f|v) = (\partial_1 f)(v_1) + \cdots + (\partial_n f)(v_n).$$

Thus (20_1) is true. Assume that the assertion has been proved for $1, 2, \ldots, k < m$. Then $f^{(k)}\colon X \to F \overset{\text{def}}{=} \mathrm{Hom}^k(E, \mathbb{R})$ and for $v_1, \ldots, v_k \in E$ we set $F(x) = f^{(k)}(x)(v_1, \ldots, v_k)$ for $x \in X$. Then the function $F\colon X \to \mathbb{R}$ is differentiable by hypothesis and with an identification of $\mathrm{Hom}\big(\mathrm{Hom}^k(E, \mathbb{R})\big)$ with $\mathrm{Hom}^{k+1}(E; \mathbb{R})$ we write the linear form $F'(x)\colon E \to \mathbb{R}$ as

$$F'(x)(v_{k+1}) = f^{(k+1)}(x)(v_1, \ldots v_k, v_{k+1})$$

Now by 7.12 once more we have

$$F'(x)(v_{k+1}) = (\partial_1 F)(x) v^{(k+1)}_1 + \cdots + (\partial_n F)(x) v^{(k+1)}_n. \tag{21}$$

Applying the induction hypothesis to F we know that

$$F(x)(v_1, \ldots, v_k) = \sum_{\substack{1\le j_p\le n\\ 1\le p\le k}} (\partial_{j_1} \cdots \partial_{j_k} f)(x) v^{(1)}_{j_1} \cdots v^{(k)}_{j_k}. \tag{20_k}$$

Taking (20_k) and (21) together we obtain (20_{k+1}), and this completes the induction. □

Definition 4.9. A multilinear form $\beta \in \mathrm{Hom}^k(E; \mathbb{R})$ is called *symmetric* if for each $j = 1, \ldots, k-1$ we have

$$\beta(v_1, \ldots, v_j, v_{j+1}, \ldots, v_k) = \beta(v_1, \ldots, v_{j+1}, v_j, \ldots, v_k). \tag{22}$$

Proposition 4.10. *For a multilinear form $\beta \in \mathrm{Hom}^k(E; \mathbb{R})$ the following statements are equivalent:*

(i) *β is symmetric.*

(ii) *For each permutation (that is, bijection) $\sigma\colon \{1, \ldots, k\} \to \{1, \ldots, k\}$ we have*

$$\beta(v_1, \ldots, v_k) = \beta(v_{\sigma(1)}, \ldots, v_{\sigma(n)}). \tag{23}$$

Proof . In the elementary theory of permutation groups one shows that the full group of all permutations of the set $\{1, \ldots, n\}$ is generated by permutations of two adjacent elements. This proves (i)⇒(ii). The reverse implication is trivial. □

Exercise E4.7. Prove that every permutation

$$f = \begin{pmatrix} 1 & 2 & \cdots & n \\ f(1) & f(2) & \cdots & f(n) \end{pmatrix}$$

of $\{1, 2, \ldots, n\}$ can be written as a composition of permutations which are compositions of "transpositions" of "adjacent elements"

$$t_j = \begin{pmatrix} n_1 & n_2 & \cdots & n_j & n_{j+1} & \cdots & n_k \\ n_1 & n_2 & \cdots & n_{j+1} & n_j & \cdots & n_k \end{pmatrix}.$$

[Hint. Step 1: Show that every permutation is a composition of cyclic permutations of suitable disjoint subsets $S = \{n_1, \ldots, n_k\}$, $n_1 < \cdots < n_k$, say

$$c_S = \begin{pmatrix} n_1 & n_2 & \cdots & n_{k-1} & n_k \\ n_2 & n_3 & \cdots & n_k & n_1 \end{pmatrix}.$$

A convenient notation of c_S is $(n_1\, n_2\, \cdots\, n_k)$. Step 2: Show

$$(n_1\, n_2\, \cdots\, n_k) = (n_1\, n_2) \circ (n_2\, n_3) \circ \cdots \circ (n_{k-2}\, n_{k-1}) \circ (n_{k-1}\, n_k)]$$

It is convenient to agree to a piece of notation which is primarily applicable to *symmetric* multilinear forms.

Definition 4.11. For a multilinear form $\beta \in \mathrm{Hom}^k(E; \mathbb{R})$ and x, $v \in E$ we write

$$\begin{aligned} \beta\star x^k &\stackrel{\text{def}}{=} \beta(x, \ldots, x). \\ (\beta\star x^{k-1})(v) &\stackrel{\text{def}}{=} \beta(x \ldots, x, v). \end{aligned} \tag{24}$$ □

Notice that we are not forming here a k-th power or a $k-1$-st power, but that $\beta\star x^k$ is a number, that is, and element of $\mathbb{R}$, and that $\beta\star x^{k-1}: E \to \mathbb{R}$ is a linear form, that is, an element of $\mathrm{Hom}(E, \mathbb{R})$. If $\delta: E \to E^k$ is the *diagonal map* defined by $\delta(v) = (v, \ldots, v)$ then

$$\beta \star x^k = (\beta \circ \delta)(x). \tag{25}$$

We shall now give an estimate for the value of a multilinear form. (Cf. E7.14 preceding 7.23.)

Lemma 4.12. *Assume that* $\|.\|$ *is a norm on* E *and that* $\beta \in \mathrm{Hom}^k(E; \mathbb{R})$ *is a multilinear form. Then*

$$\|\beta\| \stackrel{\text{def}}{=} \sup\{|\beta(v_1, \ldots, v_k)| : v_1, \ldots, v_k \in E;\ \|v_1\|, \ldots, \|v_k\| \leq 1\} \tag{26}$$

is well defined and

$$(\forall v_1, \ldots, v_n)\, |\beta(v_1, \ldots, v_n)| \leq \|\beta\|{\cdot}\|v_1\| \cdots \|v_k\|. \tag{27}$$

Proof . First we have to argue that the set $\{|\beta(v_1,\dots,v_n)\| : \|v_1\|,\dots,\|v_k\|\} \subseteq \mathbb{R}$ is bounded. Let $B \stackrel{\text{def}}{=} \{v \in E : \|v\| \le 1\}$ denote the unit ball in E. Then B^k is the unit ball in the normed space E^k equipped with the norm given by $\|(v_1,\dots,v_k)\| = \max\{\|v_1\|,\dots,\|v_k\|\}$. Then B^k is compact by 6.29. Since

$$(v_1,\dots,v_k) \mapsto |\beta(v_1,\dots,v_k)| : B^k \to \mathbb{R}$$

is continuous, it attains its maximum $\|\beta\|$ by the Theorem of the Minimum and Maximum 3.52. This shows that $\|\beta\|$ well defined.

Now we prove (27). If any of the v_j is zero, then (27) is trivially true. Now assume that $\|v_j\| > 0$ for $j = 1,\dots,k$. Then $\frac{1}{\|v_j\|}{\cdot}v_j \in B$, and by the definition of $\|\beta\|$ in (80) we have

$$\left|\beta\left(\frac{1}{\|v_1\|}{\cdot}v_1,\dots,\frac{1}{\|v_k\|}{\cdot}v_k\right)\right| \le \|\beta\|.$$

The multilinearity of β allows us to multiply this inequality with $\|v_1\|\cdots\|v_k\|$ and to obtain (27). □

Notice that this generalizes, at least as far as forms are concerned, the definition of the operator norm (see 6.33). It should be clear that Lemma 4.12 generalizes to arbitrary multilinear maps $E_1 \times \ldots \times E_k \to F$ where the E_j and F are finite dimensional normed vector spaces over $\mathbb{K} = \mathbb{R}, \mathbb{C}$.

Exercise E4.8. Prove that Lemma 4.12 defines a norm $\|\cdot\|{:}\,\mathrm{Hom}^k(E;\mathbb{R})$.

Now let us differentiate a multilinear form (cf. 7.23):

Lemma 4.13. Let $\beta \in \mathrm{Hom}^k(E,\mathbb{R})$. Then

$$\begin{aligned} &\beta'(x_1,\dots,x_k)(v_1,\dots,v_k)\\ =&\beta(v_1,x_2,\dots,x_k) + \beta(x_1,v_2,x_3,\dots,x_k) + \cdots \beta(x_1,\dots,x_{k-1},v_k). \end{aligned} \tag{28}$$

In particular,

$$(\beta\circ\delta)'(x)(v) = \beta(v,x,\dots,x) + \beta(x,v,x,\dots,x) + \cdots \beta(x,\dots,x,v). \tag{29}$$

If β is symmetric, then

$$(\forall x,\, v \in E)\, (\beta\circ\delta)'(x)(v) = n{\cdot}(\beta{\star}x^{k-1})(v), \tag{30}$$

$$(\forall x, \in E)\, (\beta\circ\delta)'(x) = n{\cdot}(\beta{\star}x^{k-1}), \tag{31}$$

$$(\forall x, \in E)\, (\beta{\star}x^{k})' = k{\cdot}(\beta{\star}x^{k-1}). \tag{32}$$

In (32) we have defined $(\beta{\star}x^n)' \stackrel{\text{def}}{=} (\beta\circ\delta)'(x)$.

Proof . We compute, using multilinearity, $\beta(x_1 + v_1, x_2 + v_2,\dots,x_k + v_k) = \beta(x_1,\dots,x_k) + \sum_{1\le j\le k}\beta(x_1,\dots,v_j,\dots,x_k) + r(v_1,\dots,v_2)$, where $r(v_1,\dots,v_k) = \sum_{1\le j_1<j_2\le\le k}\beta(x_1,\dots,v_{j_1},\dots,v_{j_2},\dots,x_k) + \cdots$. Set $v \stackrel{\text{def}}{=} (v_1,\dots,v_k)$ and define

$\|v\| \stackrel{\text{def}}{=} \max\{\|v_1\|,\dots,\|v_k\|\}$. Then

$$\begin{aligned}\|r(v)\| \leq & \sum_{1\leq j_1<j_2<\cdots\leq k} \|\beta\|\cdot\|x_1\|\cdots\|v_{j_1}\|\cdots\|v_{j_2}\|\cdots\|x_k\| + \cdots \\ \leq & \|v\|^2\cdot C_1 + \|v\|^3\mathbb{C}_2 + \cdots \|v\|^k\cdot\|\beta\|\end{aligned}$$

with numbers C_j which depend on $X = (x_1,\dots,x_k)$ only. It follows that

$$\lim_{\substack{v\to 0\\ v\neq 0}} \frac{|r(v)|}{\|v\|} = 0.$$

By the definition of the derivative in 7.2 and the uniqueness statement in 7.3 we may conclude (28). Now (29) is an immediate consequence. If β is symmetric, then $\beta(x,\dots,v,\dots,x) = \beta(x,\dots,x,v) = (\beta\star x^{k-1})(v)$. Now (30), (31) and (32) follow successivley from this and the definitions. $\square$

In a warning note following 4.20 we pointed out that instructors of elementary calculus like a notation of the type $(x^n)' = nx^{n-1}$ which is conceptually problematic, because the prime $'$ operates on *functions*, associating with a function f again a function f'. However, x^n is *not* a function; $x \mapsto x^n$ is a function p_n. Likewise it requires a lot of mind reading to recognize $(x^n)'$ as the function p_n'

A similar warning is in order concerning the notation $(\beta\star x^n)'$ used in (31). We chose it so as to make it evident that the formula (32) generalizes the formula $p_n' = n\cdot p_{n-1}$.

The Taylor formula

We now finally consider an m-times continuously differentiable level function $f\colon X \to \mathbb{R}$, X open in a finite d.imensional normed space E such as e.g. $\mathbb{R}^n$. Then $f^{(k)}(a) \in \mathrm{Hom}^k(E;\mathbb{R})$, $k = 1,\dots,m$, $a \in X$. We write $f^{(0)} = f$.

Lemma 4.14. *Define* $P\colon E \to \mathbb{R}$ *by*

$$\begin{aligned}P(x) =& f(a) + \frac{1}{1!}\cdot f'(a)\star(x-a) + \frac{1}{2!}\cdot f''(a)\star(x-a)^2 + \cdots \\ &+ \frac{1}{m!}\cdot f^{(m)}(a)\star x^m.\end{aligned} \tag{33}$$

Then $P^{(k)}(0) = f^{(k)}(a)$ *for* $k = 0,\dots,m$

Proof . We claim that

$$\begin{aligned}P^{(k)}(x) =& f^{(k)}(a) + \frac{1}{1!} f^{(k+1)}(a)\star(x-a) + \frac{1}{2!}\cdot f^{(k+2)}(a)\star(x-a)^2 + \cdots \\ &+ \frac{1}{(m-k)!}\cdot f^{(m-k)}(a)\star x^{m-k},\end{aligned}$$

$k = 0,1,\dots,m$. This follows by induction from (32). Putting $x = a$ we obtain the assertion. $\square$

We call P the *Taylor polynomial* for f of degree m. Now we are ready for the following result. The basic idea of its proof was introduced in the special case of 4.3.

Taylor's Theorem

Theorem 4.15. *Assume that X is an open subset of a finite dimensional normed space E such as $\mathbb{R}^n$, and that $f: X \to \mathbb{R}$ is an m-times continuously differentiable level function. Then for each $a \in X$, there is a function $r: X \to \mathbb{R}$ such that $\lim_{x\to a} r(x) = 0$ and that*

$$(34)\quad f(x) = f(a) + \frac{1}{1!}\cdot f'(a)\star(x-a) + \cdots + \frac{1}{m!}\cdot f^{(m)}(a)\star(x-a)^m + \|x-a\|^m\cdot r(x).$$

Proof. Let $F = f - P$ where P is the Taylor polynomial for f of degree m. Then for each $a \in X$ we define

$$r(x) = \begin{cases} 0 & \text{if } x = a, \\ \|x-a\|^{-m}\cdot F(x) & \text{if } x \neq a. \end{cases}$$

Then (34) holds and we have to show that $r(x) \to 0$ for $x \to a$. Since X is open, there is a positive number ρ such that the open ball $U_\rho(a)$ of radius ρ around a is entirely contained in X. Assume that $\varepsilon > 0$ is given. We must show that $r(x) < \varepsilon$ for all x which are sufficiently close to a. Since $f^{(m)}$ and then also $F^{(m)}$ are continuous and $F^{(m)}(a) = 0$, there is a $\delta > 0$, $\delta \le \rho$ such that $\|x-a\| \le \delta$ implies $\|F^{(m)}(x)\| < \varepsilon\cdot m!$, where $\|F^{(m)}(x)\|$ is defined as in 4.12. Now let $x \in U_\delta(a)$, $x \neq a$. We set $e \overset{\text{def}}{=} \frac{1}{\|x-a\|}\cdot(x-a)$. Then e is a unit vector and we define a function $\varphi: [0,\rho[\ \to \mathbb{R}$ by $\varphi(t) \overset{\text{def}}{=} F(a + t\cdot e)$. By the Chain Rule we can successively differentiate φ at least m times as follows

1) $\varphi'(t) = F'(a+t\cdot e)(e)$,
2) $\varphi''(t) = \big(F''(a+t\cdot e)(e)\big)(e) = F''(a+t\cdot e)\star e^2$,

$\vdots$

k) $\varphi^{(k)}(t) = F^{(k)}(a+t\cdot e)\star e^k$, $k = 1, 2, \dots, m$.

Then Lemma 4.14 and the definition of F imply that $\varphi^{(k)}(0) = 0$ for $k = 0, \dots, m$. Now we apply Lemma 4.60 to φ and find a number $u(t) \in [0, t[$, $0 \le t < \rho$, such that

$$(35)\qquad F(a+t\cdot e) = \varphi(t) = \frac{1}{m!}\cdot\varphi^{(m)}\big(u(t)\big)t^m, \quad 0 \le t < \rho.$$

Now $\varphi^{(m)}\big(u(t)\big) = F^{(m)}(a+u(t)\cdot e)\star e^m$. We set $\tau = \|x-a\| > 0$; then $x = a+\tau\cdot e$ and note that from (89) and $0 < u(\tau) < \tau = \|x-a\| < \delta$ we estimate $\|r(x)\| = \|x-a\|^{-m}\cdot\|F(x)\| = \frac{1}{m!\tau^m}\cdot\|F^{(m)}(a+u(\tau)\cdot e)\star e^m\cdot\tau^m\| \le \frac{1}{m!}\cdot\|F^{(m)}(x)\|\cdot\|e\|^m < \varepsilon$. This completes the proof. $\square$

From Proposition 4.8 we know the coefficients of the multilinear form $f^{(k)}(a) \in \mathrm{Hom}^k(E;\mathbb{R})$. If $h = x - a$ then the number $f^{(k)}(a)\star h^k$ is given by

$$\sum_{1\le j_1,\dots,j_k\le n} (\partial_{j_1}\cdots\partial_{j_k} f)(a_1,\dots,a_n)h_{j_1}\cdots h_{j_k}, \quad (n^k \text{ summands}). \tag{36}$$

However, this is not the last word, since several of these n^k summands agree after Schwarz' Theorem. Indeed, if in the k-tuple $(j_1,\dots,j_k)$ the number $j \in \{1,\dots,n\}$ occurs p_j-times, then $p_1 + \ldots + p_n = k$ und $h_{j_1}\cdots h_{j_k} = h_1^{p_1}\cdots h_n^{p_n}$. We write

$$\begin{aligned}
\mathbf{p} &= (p_1,\dots,p_n),\\
|\mathbf{p}| &= p_1 + \cdots + p_n,\\
\mathbf{h}^{\mathbf{p}} &= h_1^{p_1}\cdots h_n^{p_n},\\
\partial^{\mathbf{p}} f &= \partial_1^{p_1}\cdots\partial_n^{p_n} f, \text{ and}\\
\binom{k}{\mathbf{p}} &= \frac{k!}{p_1!\cdots p_n!}.
\end{aligned}$$

The multiplicities in the terms $\mathbf{h}^{\mathbf{p}}$ occuring in the sum (36) are known from the expansion $(h_1 + \cdots + h_n)^k = \sum_{|\mathbf{p}|=k}\binom{k}{\mathbf{p}}\mathbf{h}^{\mathbf{p}}$. Then (36) can also be written in the form of $\sum_{|\mathbf{p}|=k}\binom{k}{\mathbf{p}}(\partial^{\mathbf{p}} f)(a)\mathbf{h}^{\mathbf{p}}$. If we now abbreviate $p_1!\cdots p_n!$ by $\mathbf{p}!$, then we can write the Taylor formula (34) in following fashion

$$f(a+h) = \sum_{|\mathbf{p}|\le k} \frac{1}{\mathbf{p}!}(\partial^p f)(a)\mathbf{h}^{\mathbf{p}} + \|h\|^k r(h), \quad \mathbf{p} = (p_1,\dots,p_n). \tag{37}$$

In the form of (37) the Taylor polynomial is accessible to computation as the partial derivatives $\partial^p f$ can be computed directly via successive partial derivation.

Postscript

The idea of higher derivatives is more complicated in the calculus of several variable than it was in the case of one variable: In Analysis I passing to higher derivatives was just "more of the same." The one major result that arises from the existence of higher derivatives is Taylor's Theorem. This remains true in the case of several variables, but the technical complications are substantially higher here.

It is perhaps a relief in the direction of most applications that the essential applications work with a Taylor formula of degree 2, and here the complications are moderate. This is why we treat this case separately, leaving the instructor a choice to skip across the general degree n-version of Taylor's Theorem.

Moreover, the most crucial theorem in this area arises when we consider second derivatives: The Theorem of H. A. Schwarz greatly simplifies the information contained in the second derivative of a level function $f\colon X \to \mathbb{R}$ where X is usually an open subset of $\mathbb{R}^n$. The "second derivative" $f''(x)$ is a bilinear map whose coefficients form a matrix, the Hesse matrix $H(x)$, and according to Schwarz'

Theorem this matrix is symmetric, and the second derivative is a quadratic form. All of this works under the hypothesis that that second partial derivatives are continuous. This allows us to apply all the information provided by linear algebra on real quadratic forms, and this in turn says that with a very good approximation, in the vicinity of a point a level function behaves like a quadratic functions which is of particular interest in a critical point at which the gradient vanishes.

Deep down Schwarz' Theorem has what in topology one would call a "homological" flavor: Our proof shows that it amounts to tracking the values of a function along the boundary of a very small rectangle (or paralellogram) with the result "zero"; actually we went around half-way along one half and then around the other, equalizing the two.

The higher derivatives of a level function turn out to be multilinear forms. If one handles the formalism efficiently, the notation of the Taylor Theorem is so close to that of one variable calculus that it can be easily remembered. However, in all of this streamlining one should not forget that the simplicity of the notation conceals considerable technical complications.