35. Higher derivatives and multivariate extrema

We will now close our discussion of multivariate differentiation by discussing its connection to classification of local extrema of real-valued functions of multiple variables. These topics are frequently mentioned in calculus courses; the point here is to give precise conditions under which the "standard" conclusions are true.

35.1 Higher derivatives.

To motivate the introduction of higher derivatives, we state and prove a multivariate version of Theorem 29.11:

Theorem 35.1 (Multivariate first-derivative test) *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a function and let $x \in \operatorname{int}(\operatorname{Dom}(f))$ be a point where f has a local extremum. Then*

$$\forall v \in \mathbb{R}^n\colon\ \frac{\partial f}{\partial v}(x) \text{ exists } \Rightarrow\ \frac{\partial f}{\partial v}(x) \geqslant 0 \tag{35.1}$$

and

$$\forall i = 1,\dots,n\colon\ \frac{\partial f}{\partial x_i}(x) \text{ exists } \Rightarrow\ \frac{\partial f}{\partial x_i}(x) = 0 \tag{35.2}$$

In particular,

$$f \text{ differentiable at } x\ \Rightarrow\ Df(x) = \nabla f(x) = 0 \tag{35.3}$$

Proof. Suppose that f has a local minimum at x. Then there exists $\delta > 0$ such that $B(x,\delta) \subseteq \operatorname{Dom}(f)$ and

$$\forall z \in B(x,\delta)\colon\ f(z) \geqslant f(x) \tag{35.4}$$

For $v \in \mathbb{R}^n$, this means that, for $h > 0$ with $h\|v\| < \delta$, we have $x + hv \in B(x,\delta)$ and so $f(x+hv) - f(x) \geqslant 0$. Using that we get

$$\frac{\partial f}{\partial v}(x) = \lim_{h\to 0^+} \frac{f(x+hv) - f(x)}{h} \geqslant 0 \tag{35.5}$$

we get the first part of the claim.

For the second part we note that

$$\frac{\partial f}{\partial x_i}(x) = \frac{\partial f}{\partial e_i}(x) = -\frac{\partial f}{\partial(-e_i)}(x) \tag{35.6}$$

By (35.1), the first partial derivative is non-negative and the negative of the second partial derivative is non-positive. This is only possible if all these quantities vanish. The third part of the claim is a consequence of the definition of $\nabla f(x)$. □

The alternative formulations of the first derivative test are introduced to account for all the various possibility that may occur. For instance, for the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x,y) := x^2 + y^2$ the partial derivatives, and the gradient, vanish at the point $(x,y) = (0,0)$ which is also the point where f achieves its (global) minimum. However, the function f defined by $f(x,y) := \sqrt{x^2+y^2}$ is not differentiable at the origin, but in light of $f(0+hv) = h\|v\|$ for $h > 0$ we still have

$$\frac{\partial f}{\partial v}(0,0) = \|v\| \tag{35.7}$$

which is non-negative for all $v \in \mathbb{R}^n$.

As in the single-variable test, the first derivative alone cannot decide the actual "trend" of the function near the point where it vanishes; this can only be decided by second derivatives as done in Theorem 31.3. For this we introduce:

Definition 35.2 (Higher derivatives) *Given* $f\colon \mathbb{R}^n \to \mathbb{R}^m$ *with domain* $\operatorname{Dom}(f)$*, we define functions* $\{D^k f\}_{k\in\mathbb{N}}$ *acting as* $D^k f\colon \mathbb{R}^n \to \mathbb{R}^{m\cdot n^k}$ *by setting* $D^0 f := f$ *with domain* $\operatorname{Dom}(D^0 f) := \operatorname{Dom}(f)$ *and, assuming that* $D^k f$ *has been defined, letting*

$$D^{k+1} f(x) := D(D^k f)(x) \tag{35.8}$$

for all x *in the domain*

$$\operatorname{Dom}(D^{k+1} f) := \bigl\{x \in \operatorname{int}(\operatorname{Dom}(D^k f))\colon D^k f \text{ differentiable at } x\bigr\} \tag{35.9}$$

Throughout, we identify $\mathcal{M}_{n,m}$ *with* $\mathbb{R}^{n\cdot m}$.

As for the first derivative, in practice we need to work with a suitable choice of coordinates. This leads to:

Definition 35.3 (Higher partial derivatives) *Given* $f\colon \mathbb{R}^n \to \mathbb{R}$ *with domain* $\operatorname{Dom}(f)$*, a natural* $k \geqslant 1$ *and indices* $i_1, \dots, i_k \in \{1, \dots, n\}$*, the* k*-th partial derivative of* f *with respect to variables* $x_{i_1} \dots x_{i_k}$ *is defined recursively as*

$$\frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}}(x) := \frac{\partial}{\partial x_{i_1}}\left(\frac{\partial^{k-1} f}{\partial x_{i_2} \dots \partial x_{i_k}}\right)(x) \tag{35.10}$$

whenever the $(k-1)$*-st partial derivative on the right exists in an open neighborhood of* x *and is differentiable with respect to* x_{i_1} *at* x.

To give an example, let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x,y) := x^4 + xy^3 \tag{35.11}$$

Then

$$\begin{pmatrix} \dfrac{\partial f}{\partial x}(x,y) \\ \dfrac{\partial f}{\partial y}(x,y) \end{pmatrix} = \begin{pmatrix} 4x^3 + y^3 \\ 3xy^2 \end{pmatrix} \tag{35.12}$$

and for the second derivatives we get

$$\begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(x,y) & \dfrac{\partial^2 f}{\partial y \partial x}(x,y) \\ \dfrac{\partial^2 f}{\partial x \partial y}(x,y) & \dfrac{\partial^2 f}{\partial y^2}(x,y) \end{pmatrix} = \begin{pmatrix} 12x^2 & 3y^2 \\ 3y^2 & 6xy \end{pmatrix} \tag{35.13}$$

where $\frac{\partial^2 f}{\partial x^2}$ is a standard abbreviation for $\frac{\partial^2 f}{\partial x \partial x}$. The fact that we write the result as a matrix is a consequence of the fact that f is scalar-valued and we are talking about the second derivative.

35.2 Equality of mixed partials.

The reader has surely noticed that the mixed partial derivatives in (35.13) are equal. This is not a coincidence due to the following result:

Theorem 35.4 (Clairaut/Schwarz' theorem, proved by H. Schwarz in 1873) *Suppose that $f\colon \mathbb{R}^2 \to \mathbb{R}$ and $x \in \operatorname{int}(\operatorname{Dom}(f))$ are such that, for some $\delta > 0$ with $B(x,\delta) \subseteq \operatorname{Dom}(f)$, the following three conditions hold:*

(1) *f is continuous on $B(x,\delta)$,*
(2) *the partial derivatives $\frac{\partial f}{\partial x_1}$ and $\frac{\partial f}{\partial x_2}$ exist and are continuous on $B(x,\delta)$,*
(3) *the mixed partials $\frac{\partial^2 f}{\partial x_1 \partial x_2}$ and $\frac{\partial^2 f}{\partial x_2 \partial x_1}$ exist on $B(x,\delta)$ and are continuous at x.*

Then

$$\frac{\partial^2 f}{\partial x_1 \partial x_2}(x) = \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) \tag{35.14}$$

Proof. Write e_1 and e_2 for the unit vectors in the coordinate directions in $\mathbb{R}^2$ and let $h\colon \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$h(s,t) := f(x + se_1 + te_2) - f(x + se_1) - f(x + te_2) + f(x) \tag{35.15}$$

with domain $\operatorname{Dom}(h) := \{(s,t) \in \mathbb{R}^2 \colon s^2 + t^2 < \delta^2\}$. The assumptions ensure that h is continuous and continuously differentiable on its domain. Assume for simplicity that $s,t > 0$ and note that $h(0,t) = 0$. The Mean-Value Theorem then gives us $s' \in (0,s)$ such that

$$\begin{aligned} h(s,t) = h(s,t) - h(0,t) &= s\frac{\partial h}{\partial s}(s',t) \\ &= s\left(\frac{\partial f}{\partial x_1}(x + s'e_1 + te_2) - \frac{\partial f}{\partial x_1}(x + s'e_1)\right) \end{aligned} \tag{35.16}$$

Using the Mean-Value Theorem in turn gives us $t' \in (0,t)$ such that

$$h(s,t) = st\frac{\partial^2 f}{\partial x_2 \partial x_1}(x + s'e_1 + t'e_2) \tag{35.17}$$

Performing the previous steps in the opposite order in turn gives us $t'' \in (0,t)$ and $s'' \in (0,s)$ such that

$$h(s,t) = st\frac{\partial^2 f}{\partial x_1 \partial x_2}(x + s''e_1 + t''e_2) \tag{35.18}$$

Dividing by st we conclude that the mixed partials in (35.17) and (35.18) are equal. Taking $s \to 0^+$ and $t \to 0^+$, the assumed continuity implies that the mixed partials converge to those in (35.14). □

The conditions (1-3) are not all independent. Indeed, by Lemma 32.5 the continuity of the first partial derivatives implies differentiability and Lemma 32.4 then gives continuity of the function itself. That being said, the continuity of the mixed partials at x is essential for the proof and also for the conclusion. This is seen in the standard example of $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) := \begin{cases} xy\dfrac{x^2 - y^2}{x^2 + y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases} \tag{35.19}$$

This f has second partial derivatives throughout $\mathbb{R}^2$ yet

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1 \quad \text{and} \quad \frac{\partial^2 f}{\partial y \partial x}(0,0) = -1 \tag{35.20}$$

While f is even twice differentiable on $\mathbb{R}^2$ and so the second partial derivatives exist throughout $\mathbb{R}^2$ as well, quite apparently, neither of these two conditions guarantees equality of mixed partials. The mixed partials (which are equal everywhere on $\mathbb{R}^2 \smallsetminus \{(0,0)\}$) of this f are of course not continuous at $(0,0)$.

We remark that the problem of equality of mixed partials has attracted attention of many mathematicians going back to Euler in mid 18th centry. However, the early published proofs were incomplete and the result received a proper and correct treatment only in the second half of the 19th century. Alternative versions exist differing mainly in point (3) of the assumptions, but the above is perhaps the one most natural.

35.3 Role of the second derivative.

We will now elaborate further on the role of the second derivative. Just as for the single-variable case, the second derivative captures the trend of the first derivative and thus expresses the way the graph of the function "curves" near the point. We start with:

Definition 35.5 *Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}$ admits the second partial derivatives at the point $x \in \mathrm{Dom}(f)$. The matrix*

$$\mathrm{Hess}[f](x) := \left\{ \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right\}_{i,j=1}^{n} \tag{35.21}$$

is called the Hessian of f *at x.*

We now observe:

Lemma 35.6 *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be defined and twice continuously differentiable in an open ball centered at x. Then*

$$f \text{ has a local minimum at } x \ \Rightarrow\ \forall v \in \mathbb{R}^n\colon v \cdot \mathrm{Hess}[f](x)v \geqslant 0 \tag{35.22}$$

The conclusion means that the Hessian at x is positive semi-definite.

Proof. Let $h\colon \mathbb{R} \to \mathbb{R}$ be defined by

$$h(t) := f(x+tv) - f(x) \tag{35.23}$$

which assuming that $\nabla f(x) = 0$, as happens at local extrema by the first derivative test, can be written as

$$h(t) = f(x+tv) - f(x) - tv \cdot \nabla f(x) \tag{35.24}$$

The function is continuous with $h(0) = 0$ and so l'Hospital's Rule tells us

$$\begin{aligned} \lim_{t\to 0} \frac{h(t)}{t^2} &= \lim_{t\to 0} \frac{h'(t)}{2t} = \lim_{t\to 0} \frac{v \cdot \nabla f(x+tv) - v \cdot \nabla f(x)}{2t} \\ &= \sum_{i=1}^{n} \lim_{t\to 0} \frac{\frac{\partial f}{\partial x_i}(x+tv) - \frac{\partial f}{\partial x_i}(x)}{2t} v_i = \frac{1}{2} \sum_{i=1}^{n} \lim_{t\to 0} \frac{\partial f}{\partial x_j \partial x_i}(x+tv) v_i v_j \end{aligned} \tag{35.25}$$

provided the limit on the right exists. This is ensured by the assumption of continuity of the second derivative and so we get

$$\lim_{t\to 0}\frac{h(t)}{t^2}=\frac{1}{2}v\cdot\text{Hess}[f](x)v \tag{35.26}$$

The claim follows from the fact that, if x is the point of a local minimum, then $h(t)\geqslant 0$ for t sufficiently small. □

The above lemma gives a necessary condition for f having a local minimum. Just a small strengthening of this condition is also necessary:

Theorem 35.7 (Multivariate second derivative test) *Let $f\colon\mathbb{R}^n\to\mathbb{R}$ be defined and twice continuously differentiable in $B(x,\delta)$, for some $\delta>0$. Then the conditions*

$$\nabla f(x)=0 \tag{35.27}$$

and

$$\forall z\in B(x,\delta)\,\forall v\in\mathbb{R}^n\colon\, v\cdot\text{Hess}[f](z)v\geqslant 0 \tag{35.28}$$

imply that f has a local minimum at x.

Proof. Given $z\in B(x,\delta)$, let $v:=z-x$ and note that $\|v\|<\delta$. Let$h$ be defined as in the previous proof. Since $h(0)=0$ and $h'(0)=0$, Taylor's theorem gives us existence of $t\in(0,1)$ such that

$$h(1)=\frac{1}{2}h''(t')=\frac{1}{2}v\cdot\text{Hess}[f](x+t'v)v \tag{35.29}$$

Since $x+t'v=(1-t')x+t'z\in B(x,\delta)$, the right-hand side is non-negative. But this means that $h(1)=f(z)-f(x)\geqslant 0$, proving that f has a local minimum at x. □

We note that the assumption that the Hessian of f at x is positive semi-definite alone does not suffice for the conclusion. (An example of this appeared in the homework.) The positive semi-definiteness of the Hessian can be checked using *Sylvester's criterion*: An $n\times n$-matrix is positive semi-definite if and only if all of its principal minors are non-negative. (A principal minor is the determinant of a sub-matrix obtained by removing rows of some indices and columns of the same indices.) As for the single-variable case, the positive semi-definiteness of the Hessian implies convexity.

To demonstrate the power of the Second derivative test on an example, consider the function $f\colon\mathbb{R}^2\to\mathbb{R}$ given by

$$f(x,y):=x^2+y^2-xy \tag{35.30}$$

This function has the Hessian

$$\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix} \tag{35.31}$$

which is positive definite and, since $\nabla f(0,0)=0$, the function has a local minimum at $(0,0)$. Another scenario is represented by function $-f$ whose Hessian is not positive semi-definite, but this is merely due to the reversal of the sign. What matters is that all eigenvalues of the Hessian are of the same sign, either all non-negative, in which case we get a local minimum, or all non-positive, in which case we get a local maximum.

A key difference to the single-variable case is that, in higher dimension, there is also an intermediate scenario: Indeed, the function

$$f(x, y) := x^2 + y^2 - 8xy \tag{35.32}$$

has the Hessian

$$\begin{pmatrix} 2 & -8 \\ -8 & 2 \end{pmatrix} \tag{35.33}$$

whose diagonal entries are non-negative but the determinant is $4 - 64 = -60 < 0$ and so the matrix is not positive semidefinite. The graph of f exhibits a "saddle point" at $(0, 0)$, with one direction curving-up and the other direction curving-down. These directions are identified by diagonalizing the matrix in a basis of eigenvectors: the positive eigenvalue corresponds to the "convex" direction, the negative eigenvalue corresponds to the "concave" direction.