These notes cover the last part of the course that was not lectured in detail (only definitions and statements of results were given). This part is therefore not examinable and is intended for those interested. I will be using (and continuing) the numbering of results in lectures.

Local extrema

We are given an open set $U \subset \mathbb{R}^m$, a function $f\colon U \to \mathbb{R}$ and a point $a \in U$.

We say f has a *local maximum at* a if there is an $r > 0$ such that $D_r(a) \subset U$ and $f(x) \leqslant f(a)$ for all $x \in D_r(a)$.

We say f has a *local minimum at* a if there is an $r > 0$ such that $D_r(a) \subset U$ and $f(x) \geqslant f(a)$ for all $x \in D_r(a)$.

We say f has a *local extremum at* a if f has a local maximum or a local minimum at a.

Proposition 13. *We are given an open set $U \subset \mathbb{R}^m$, a function $f\colon U \to \mathbb{R}$ and a point $a \in U$. If f is differentiable at a and f has a local extremum at a, then $f'(a) = 0$.*

Before the proof, we make the following definition arising naturally from the above result. We say a is a *stationary point of* f if f is differentiable at a and $f'(a) = 0$.

Proof. Replacing f with $-f$, we may assume that f has a local maximum at a. Now assume that $f'(a) \neq 0$. Then there exists $u \in \mathbb{R}^m$ such that $f'(a)(u) \neq 0$. By rescaling, we may assume that $f'(a)(u) > 0$ and $\|u\| = 1$. Now, by the definition of differentiability we have

$$f(a+h) = f(a) + f'(a)(h) + \|h\|\varepsilon(h)$$

where $\varepsilon(0) = 0$ and ε is continuous at 0. Choose $\delta > 0$ such that if $\|h\| \leqslant \delta$, then $|\varepsilon(h)| < f'(a)(u)$ and $f(a+h) \leqslant f(a)$. Putting $h = \delta u$, we have

$$0 \geqslant f(a+\delta u) - f(a) = \delta\big(f'(a)(u) + \varepsilon(\delta u)\big) > 0$$

which is a contradiction. □

Note. *The converse of Proposition 13 is false in general. E.g.,* 0 *is a stationary point of $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$, but f has no local extremum at* 0.

Second-order Taylor expansion

Here we prove a special case of Taylor's theorem in higher dimension that will be used to prove a converse of the previous result under additional assumptions.

Lemma 14. *We are given an open set $U \subset \mathbb{R}^m$, a function $f\colon U \to \mathbb{R}^n$ and a point $a \in U$. If f is twice differentiable at a, then*

$$f(a+h) = f(a) + f'(a)(h) + \frac{1}{2}f''(a)(h,h) + \mathrm{o}(\|h\|^2)$$

Proof. By considering components of f, we may assume that $n = 1$. By definition of the second derivative, we have

$$f'(a+h) = f'(a) + f''(a)(h) + \|h\|\varepsilon(h)$$

where $\varepsilon(0) = 0$ and ε is continuous at 0. We next define

$$g(h) = f(a+h) - f(a) - f'(a)(h) - \frac{1}{2}f''(a)(h,h)$$

which is defined on some open neighbourhood of 0. We need to show that $g(h) = \mathrm{o}(\|h\|^2)$. Fix h and define $\varphi\colon [0,1] \to \mathbb{R}$ by $\varphi(t) = g(th)$. Note that

$$\varphi(t) = f(a+th) - f(a) - tf'(a)(h) - \frac{t^2}{2}f''(a)(h,h)$$

So φ is continuous on $[0,1]$ and differentiable on $(0,1)$ with

$$\begin{aligned}\varphi'(t) &= f'(a+th)(h) - f'(a)(h) - tf''(a)(h,h)\\ &= f'(a+th)(h) - f'(a)(h) - f''(a)(th,h)\\ &= \big[f'(a+th) - f'(a) - f''(a)(th)\big](h)\\ &= \|th\|\varepsilon(th)(h)\end{aligned}$$

By the MVT, there exists $t = t(h) \in (0,1)$ such that

$$g(h) = \varphi(1) - \varphi(0) = \varphi'(t) = \|th\|\varepsilon(th)(h)$$

It follows by Lemma 1 in the lectures that $|\varepsilon(th)(h)| \leqslant \|\varepsilon(th)\|\|h\|$, and hence $|g(h)| \leqslant \|h\|^2\|\varepsilon(th)\|$, from which the result follows. □

Classification of extreme points and the Hessian

Recall from Linear Algebra that a symmetric bilinear map $T\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is positive definite if $T(x,x) > 0$ for all $x \in \mathbb{R}^m \setminus \{0\}$ and negative definite if $T(x,x) < 0$ for all $x \in \mathbb{R}^m \setminus \{0\}$.

Theorem 15. *We are given an open set $U \subset \mathbb{R}^m$, a function $f\colon U \to \mathbb{R}$ and a point $a \in U$. Assume that f is twice differentiable on U and f'' is continuous at a.*

If a is a stationary point of f and $f''(a)$ is positive definite, then f has a local minimum at a.

If a is a stationary point of f and $f''(a)$ is negative definite, then f has a local maximum at a.

Note By Theorem 12, if f'' exists on a neighbourhood of a and is continuous at a, then $f''(a)$ is a symmetric bilinear map. The corresponding $m \times m$ matrix H defined by

$$H_{i,j} = f''(a)(e_i, e_j) = D_iD_jf(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a)$$

is thus symmetric. It is called the *Hessian of f at a*. Recall that $f''(a)$, or equivalently the matrix H, is positive definite (respectively, negative definite) if and only if all eigenvalues of H are positive (respectively, negative). This will be used in the proof below.

Proof of Theorem 15. Assume that $f'(a) = 0$ and $f''(a)$ is positive definite. Let $u_1, \dots, u_m$ be an orthonormal basis of $\mathbb{R}^m$ such that $f''(a)(u_i, u_j) = 0$ if $i \neq j$. Note that in this case $f''(a)(u_i,u_i)$, $1 \leqslant i \leqslant m$, are the eigenvalues of the Hessian with multiplicities. Set

$$\mu = \min\{f''(a)(u_i,u_i) : 1 \leqslant i \leqslant m\}$$

Then $\mu > 0$ and for $h = \sum_{i=1}^m h_iu_i \in \mathbb{R}^m$, we have

$$f''(a)(h,h) = \sum_{i,j=1}^m h_ih_jf''(a)(u_i,u_j) = \sum_{i=1}^m h_i^2 f''(a)(u_i,u_i) \geqslant \mu\sum_{i=1}^m h_i^2 = \mu\|h\|^2$$

Now, by Lemma 14, we have

$$f(a+h) = f(a) + \frac{1}{2}f''(a)(h,h) + \|h\|^2\varepsilon(h)$$

where $\varepsilon(0) = 0$ and ε is continuous at 0. Choose $\delta > 0$ such that $|\varepsilon(h)| < \mu/2$ whenever $\|h\| < \delta$. Then

$$f(a+h) - f(a) \geqslant \|h\|^2\big(\mu/2 + \varepsilon(h)\big) \geqslant 0$$

whenever $h \in D_\delta(0)$. Thus, f has a local minimum at a. The proof of the second statement is similar. □