Lecture 31: Extrema of multivariable functions

Setup: $\hat{a} \in \mathbb{R} \xrightarrow{f} \mathbb{R}$ function of $n$ variables $(x_1, \ldots, x_n)$

Definition: $f$ has a (global/absolute) \underline{maximum} at $\hat{a}$ if $f(\hat{a}) \geq f(x)$ for all $x \in \mathbb{R}^n$.

(Then $f(\hat{a})$ is called the \underline{maximum value}.)

- $f$ has a relative/local maximum at $\hat{a}$ if $f\big|_{B(\hat{a}, \varepsilon) \cap \mathbb{R}^n}$ has a maximum at $\hat{a}$ for $\varepsilon > 0$ sufficiently small.

- An extremum (global or local) of $f$ means a maximum or minimum of $f$ (global or local).

Theorem 1: Assume $\hat{a} \in \text{int}(\mathbb{R})$. If $f$ has a local extremum at $\hat{a}$, then $\nabla f(\hat{a}) = \vec{0}$, i.e. $\hat{a}$ is a stationary point of $f$.

Proof: Write $\hat{a} = (a_1, \ldots, a_n)$, and restrict $f$ to the interval $I_j = \{(a_1, \ldots, x_j, \ldots, a_n) | a_j - \varepsilon < x_j < a_j + \varepsilon\}$.

This is a function of a single variable $x_j$, call it $F(x_j) := f(a_1, \ldots, x_j, \ldots, a_n)$, with max/min at $(x_j =) a_j$.
A theorem from Math 203 says that $F'(a_j) = 0$.
But $F'(a_j)$ is $\frac{\partial F}{\partial a_j}(a)$, which is therefore 0 for each $j$. □

Remarks:

(A) At a stationary point, the tangent plane to the graph of $(x) \ z = f(x_1, ..., x_n)$
in $\mathbb{R}^{n+1}$ is $z = f(x)$ ("horizontal").
Why? The normal to $(x)$ is $\nabla (z - f(x)) = (-\nabla f(x), 1) = (\vec{0}, 1)$.

(B) The converse to Theorem 1 is FALSE:
$f(x,y) = x^2 - y^2$ has $\nabla f(0,0) = \vec{0}$, but is concave up in the $\{x\}$ direction.
A saddle-point is a stationary point which is not a local extremum.

Example:

$z = f(x,y) = x^2 + 4xy + y^2$ (what can we say about it?)

- $x^2$-cross-section: $f(x,0) = x^2 \rightarrow$ concave up $f_{xx}(0) = 2 = f_{yy}(0)$
- $y^2$-cross-section: $f(0,y) = y^2 \rightarrow$ also concave up
- Also $D_x f = \nabla f \cdot \vec{u} = u_x f_x + u_y f_y = u_x (2x+4y) + u_y (2y+4x)$
is 0 at $\vec{0}$, so $(0,0)$ is certainly a stationary point of $f$ suggests local minimum at $(0,0)$? Just to be sure, let's differentiate twice in the "SE-direction":

$D^{\mathbb{R}} f = D^{\mathbb{R}} (D_x f) = D^{\mathbb{R}} \left\{ \frac{\partial}{\partial x} (2x+4y-(2y+4x)) \right\} = D^{\mathbb{R}} \{\vec{e}(y-x)\}$

$= \nabla (\vec{e}(y-x)) \cdot \vec{e}(1,-1) = (-\vec{e}, \vec{e}) \cdot \frac{1}{2} (1,-1) = -2$ (!!)
Moral: finding the concavity in the \( x \)- \( y \)-directions may not tell you much about the directions of greatest or least concavity, or what those greatest or least concavities are!

To explore further, let \( \hat{\alpha} \) be a stationary point of a function \( f(x,y) \), and set \( \hat{u} = (\cos \theta, \sin \theta) \). Then

\[
D^2_{\hat{u}} f = D_{\hat{u}} (u_x f_x + u_y f_y) = \nabla (u_x f_x + u_y f_y) \cdot \hat{u} \\
= (u_x f_{xx} + u_y f_{xy}, u_x f_{xy} + u_y f_{yy}) \cdot (u_x, u_y) \\
= (u_x, u_y) \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix} = \hat{u}^T H \hat{u} = \hat{u} \cdot H \hat{u},
\]
where the \( 2 \times 2 \)-matrix-valued function \( H(x,y) \) is called the Hessian of \( f \). Write \( H_0 \) for \( H(\hat{\alpha}) \).

So \( \hat{u} \cdot H \hat{u} \) is the concavity at \( \hat{\alpha} \) in the direction \( \hat{u} \), which depends on \( \theta \). We want the directions of extreme concavity:

\[
0 = \frac{d}{d\theta} \hat{u} \cdot H_0 \hat{u} = \hat{u}' \cdot H_0 \hat{u} + \hat{u} \cdot H_0 \hat{u}' = 2 \hat{u} \cdot H_0 \hat{u}' \\
\Rightarrow \hat{u}' \perp H_0 \hat{u} \quad \Rightarrow \hat{u} \parallel H_0 \hat{u} \quad \Rightarrow \hat{u} \text{ is an eigenvector of } H_0.
\]

If \( \hat{u} = \hat{v} \) is eigenvector with eigenvalue \( \lambda \), then

\[
D^2_{\hat{u}} f = \hat{v} \cdot H_0 \hat{v} = \hat{v} \cdot \lambda \hat{v} = \lambda \| \hat{v} \|^2 = \lambda \text{ is the concavity!}
\]

\[
\hat{u}' = \frac{d}{d\theta} (\cos \theta, \sin \theta) = (-\sin \theta, \cos \theta) \quad \text{is} \quad \perp \hat{u} = (\sin \theta, \cos \theta)
\]
Writing $\lambda_1 \leq \lambda_2$ for the 2 eigenvalues, these are the minimum of maximum concavities of $f$ at $\bar{a}$.

**FACT 1** \[ \Delta := \det \begin{pmatrix} f_{xx}(\bar{a}) & f_{xy}(\bar{a}) \\ f_{yx}(\bar{a}) & f_{yy}(\bar{a}) \end{pmatrix} \] is the product of max + min concavities at $\bar{a}$.

But if $\Delta > 0$, this tells me only that $\lambda_1, \lambda_2 > 0$ or $\lambda_1, \lambda_2 < 0$. Which is it?

Consider $f_{xx}(\bar{a})$, the concavity in the $x$-direction. As such, clearly in between the max/min concavities:

$\lambda_1 \leq f_{xx}(\bar{a}) \leq \lambda_2$.

So if $f_{xx}(\bar{a}) > 0$, then $\lambda_2 > 0$; and since we assumed $\Delta = \lambda_1 \lambda_2 > 0$, $\lambda_1 > 0$ too! \[ \Rightarrow \]

**FACT 2**

<table>
<thead>
<tr>
<th>$\Delta$</th>
<th>$f_{xx}(\bar{a})$</th>
<th>Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td>local min</td>
</tr>
<tr>
<td>&gt; 0</td>
<td>&lt; 0</td>
<td>local max</td>
</tr>
<tr>
<td>&lt; 0</td>
<td></td>
<td>saddle = stationary point which is not a local extremum</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>inconclusive</td>
</tr>
</tbody>
</table>

(2nd derivative test)

To recap: the 2 facts are for use in the case where $f$ is twice continuously differentiable in a neighborhood at $\bar{a}$, with $\nabla f(\bar{a}) = \bar{a}$. 1/2/3 correspond to the quadratic form $Q(\bar{a}) = \bar{a}^T H(\bar{a}) \bar{a}$ being 1/positive definite /2/negative definite /3/ indefinite.
Now we generalize to $n$ variables and examine
$f(x)$ on the gray area in the picture:

$$g(u) := f(\alpha + uh) \quad \text{for } u \in [-1,1].$$

**Chain Rule**

$$g'(u) = \nabla f(\alpha + uh) \cdot h = \sum_{j=1}^{n} h_j f_{x_j}(\alpha + uh)$$

**Chain Rule**

$$g''(u) = \nabla (\sum_{j=1}^{n} h_j f_{x_j}(\alpha + uh)) \cdot h = \sum_{i,j} h_i h_j f_{x_i x_j}(\alpha + uh)$$

$$= h^T H(\alpha + uh) h, \quad \text{where } H = \left[ \frac{\partial^2 f}{\partial x_i \partial x_j} \right] = \text{Hessian of } f$$

Taylor remainder:

$$f(\alpha + h) - f(\alpha) = g(u) - g(0) = g'(0) + \frac{1}{2} g''(c)$$

$$= \nabla f(\alpha) \cdot h + \frac{1}{2} h^T H(\alpha + ch) h$$

$$= \nabla f(\alpha) \cdot h + \frac{1}{2} h^T H(\alpha) h + \|h\|^2 E_2(\alpha ; h) \quad \text{(t)}$$

(t) where

$$E_2(\alpha ; h) = \begin{cases} \frac{1}{2\|h\|^2} h^T H(\alpha + ch) - H(\alpha) h^2, & h \neq 0 \\ 0, & h = 0 \end{cases}$$

has

$$|E_2(\alpha ; h)| = \frac{1}{2\|h\|^2} \left| \sum_{i,j} h_i h_j \left( \frac{\partial^2 f}{\partial x_i \partial x_j}(\alpha + ch) - \frac{\partial^2 f}{\partial x_i \partial x_j}(\alpha) \right) \right|$$

$$\leq \frac{1}{2\|h\|^2} \left| \sum_{i,j} \|h\| |h_j| \left( \frac{\partial^2 f}{\partial x_i \partial x_j}(\alpha + ch) - \frac{\partial^2 f}{\partial x_i \partial x_j}(\alpha) \right) \right|$$

$$\leq \frac{1}{2\|h\|^2} \left| \sum_{i,j} \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(\alpha + ch) - \frac{\partial^2 f}{\partial x_i \partial x_j}(\alpha) \right| \right|$$

as $h \to 0$ by continuity of 2nd partials.
where $E_2(\delta; \bar{h})$ is $o(1)$ ($\rightarrow 0$ as $\delta \to \delta_0$) assuming $f$ is twice continuously differentiable on $B(\bar{x}; \epsilon)$.

Now recall from Lecture 15 the Lemma on Quadratic Forms: Let $Q(\bar{h}) = \bar{h}^T A \bar{h}$, $A$ symmetric $n \times n$ with eigenvalues $\lambda_1 \leq \ldots \leq \lambda_n$. Then

(i) $Q(\bar{h}) > 0 \ \forall \bar{h} \neq 0 \iff \text{all } \lambda_j > 0 \iff$ A positive-definite

(ii) $Q(\bar{h}) < 0 \ \forall \bar{h} \neq 0 \iff \text{all } \lambda_j < 0 \iff$ A negative-definite

(iii) $Q(\bar{h})$ takes $\pm$ value $\iff \lambda_i < 0 < \lambda_n \iff$ A indefinite.

Theorem 2: If $f$ is twice continuously differentiable, with a stationary point at $\bar{x} = \bar{a}$, then:

(i) $H(\bar{a})$ pos. definite $\implies f$ has rel. minimum at $\bar{a}$.

(ii) $H(\bar{a})$ neg. definite $\implies f$ has rel. maximum at $\bar{a}$.

(iii) $H(\bar{a})$ indefinite $\implies f$ has saddle point at $\bar{a}$.

Proof: Since $\nabla f(\bar{a}) = \bar{0}$ at a stationary point, $(\star)$ becomes

$$f(\bar{a} + \bar{h}) - f(\bar{a}) = \bar{h}^T H(\bar{a}) \bar{h} + \frac{1}{2} \bar{h}^T H(\bar{a}) \bar{h} = E_2(\delta; \bar{h}).$$

The main idea is that this term doesn't matter, and so we are done by the lemma.

Here is how this works for (i):
• Let $0 \leq \lambda_1 \leq \cdots \leq \lambda_n$ be eigenvalues of $H(\xi)$. Pick $\epsilon > 0$ s.t. $|E_\xi(\xi;h)| < \frac{1}{q} \lambda_i$ for $0 < \|h\| < \epsilon$.

• Let $u \in (0,\lambda_1)$, so that the eigenvalues $\lambda_j - u$ of $H(\xi) - uI_n$ are all positive. Lemma: $\xi^T [H(\xi) - uI_n] \xi > 0 \forall \xi \neq 0$.

• So $\xi^T H(\xi) \xi > u \|\xi\|^2 \forall \xi \neq 0$, and taking $u = \frac{1}{2} \lambda_1$ gives $\frac{1}{2} \xi^T H(\xi) \xi > \frac{1}{q} \lambda_1 \|\xi\|^2 > \frac{1}{q} \|\xi\|^2 |E_\xi(\xi;h)| \geq 0$.

• $f(\xi + h) - f(\xi) = \frac{1}{2} \xi^T H(\xi) \xi + \|\xi\|^2 |E_\xi(\xi;h)| \geq \frac{1}{2} \xi^T H(\xi) \xi - \|\xi\|^2 |E_\xi(\xi;h)| > 0$ for $h \neq 0$, and $f$ has a relative minimum at $\xi$. \hfill \Box