March 11, 2019

LECTURE 12: EXTREMA.

110.211 HONORS MULTIVARIABLE CALCULUS
PROFESSOR RICHARD BROWN

Synopsis. Local and global extrema are much like their counterparts in single variable calculus. They are just points in the domain of a real-valued function where the function value is locally the lowest or highest. And they occur, if at all, at critical points of the function. If the function is differentiable everywhere, then extrema only occur at places where the derivative (matrix) has zeros in all of its elements. Thus all of the directional derivatives are 0 here also. But since directional derivatives are just derivatives along slices through the function, we can also check the concavity of these slice functions along vector directions in the domain. This leads to a notion of a second directional derivative, and also to one major application of the Hessian matrix of second partials. Relating this to a quadratic form, we construct the Second Derivative Test for a C^2-real-valued function of more than one variable. We then end with the multidimensional counterpart of the Extreme Value Theorem, once we understand what closed and bounded mean for a domain in real n-space.

Helpful Documents. Mathematica: `Extrema`, `MoreExtrema`.

Local extrema. Like in single variable calculus, local extrema are important properties of functions:

Definition 12.1. The function $f : X \subset \mathbb{R}^n \to \mathbb{R}$, for X open, has a *local minimum* at $\mathbf{a} \in X$ if there exists a neighborhood $U(\mathbf{a}) \subset X$ such that $f(\mathbf{x}) \geq f(\mathbf{a})$, for every $\mathbf{x} \in U$. And f has a *local maximum* at $\mathbf{a} \in X$ if there exists a neighborhood $U(\mathbf{a}) \subset X$ such that $f(\mathbf{x}) \leq f(\mathbf{a})$, for every $\mathbf{x} \in U$.

Some Notes:

- A local minimum (maximum) is *global* if $U = X$.
- If $f \in C^1$, then local extrema have a special quality:

  Theorem 12.2. *Given $X \subset \mathbb{R}^n$ open and $f : X \to \mathbb{R}$ a C^1-function, if f has a local extremum at $\mathbf{a} \in X$, then $Df(\mathbf{a})$ is the zero matrix (every entry in the matrix is 0).*

- The proof of this theorem shows that a directional derivative, evaluated at $\mathbf{a}$ in this case, would see a local extremum here in every direction. In particular, in the coordinate directions. And the only matrix $A_{1\times n}$ that takes every possible n-vector to 0 is the 0-matrix.

Definition 12.3. Given $f : X \subset \mathbb{R}^n \to \mathbb{R}$, with X open, a point $\mathbf{a} \in X$ is a *critical point* of f if either

(1) $Df(\mathbf{a}) = 0_{1\times n}$, or
(2) $Df(\mathbf{a})$ is undefined.

Just like in single variable calculus, extrema happen at critical points, but not all critical point need be extrema. Some examples of critical points include:

- sharp mountain ridges or roof peaks, or mountain top points, where the derivative matrix is not defined,
- smoothed over mountain tops where the derivative matrix is the zero matrix,
- saddle points,
- mesas and flood plains.

Recall that the graph of a function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is a subset of $\mathbb{R}^{n+1}$ where $X \subset \mathbb{R}^n$ is pictured as the "floor", and the height above the floor is hte value of the last variable x_{n+1}. At a place $(\mathbf{a}, f(\mathbf{a}))$, where the derivative matrix is the zero matrix, the equation of the tangent space would be

$$x_{n+1} = f(\mathbf{a}) + Df(\mathbf{a})\,(\mathbf{x} - \mathbf{a}) = f(\mathbf{a})$$

and the tangent space is parallel to the floor, or "horizontal".

However, like in Calculua I, extrema do not need to exist at all for particular functions:

Example 12.4. Let $f(x, y) = x^2 + y^2$, on the domain $X = \mathbb{R}^2 - (0, 0)$, the plane without the origin. What would you consider the point in X where f achieves its maximum? How about it's minumum?

So how does one detect an extremum, given a critical point? Really, it is all about the structure. Some ideas:

(1) Look for extreme bahavior by simply testing functions values for points "near" the critical point. In single variable calculus, we sometimes call this the 0th Derivative Test. This method is sometimes employed when the derivative matrix is not defined at a critical point.
(2) If f is differentiable at a critical point $\mathbf{a}$, then the derivative matrix is the zero matrix there. Again, this means that every directional derivative will also be 0 at $\mathbf{a}$. Recall that directional derivatives are defined via vertical slices through the graph of the function f along directions through $\mathbf{a}$ in the domain. If one follows the curve where $\mathbf{graph}(f)$ intersects the slice, adn sees the derivative go from negative (before $\mathbf{a}$), to 0 (at $\mathbf{a}$), to positive (after $\mathbf{a}$), and this happens in every direction, then you have a local minimum at $\mathbf{a}$. Of course, one can generalize this stipulation and say that in some or all directions, the derivative can stay 0 near $\mathbf{a}$, and we would still have a local min there. One can characterize this as a form of First Derivative Test for a critical point.
(3) Or if, within each slice, f, restricted to the slice is concave up or down and stays that way for all of the slices, then f is locally extreme at $\mathbf{a}$. (Again, all that is really necessary is that the concavity is not mixed along the slices.) If locally extreme is every direction, then locally extreme.

Example 12.5. For the parabolic bowl, $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x^2 + y^2$, we know that $f \in C^1$, since it is a polynomial, and

$$Df(\mathbf{x}) = \begin{bmatrix} 2x & 2y \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

only at the origin $x = y = 0$. So we have

$$D_{\mathbf{v}}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = Df(\mathbf{0})\mathbf{v} = \begin{bmatrix} 0 & 0 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0, \quad \forall \mathbf{v} \in \mathbb{R}^2.$$

So choose a direction $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2 \subset \mathbb{R}^3$ in the floor of $\mathbb{R}^3$, considered the domain of f. Then the vertical plane containing $\mathbf{v}$ is defined by the vector orthogonal to both $\mathbf{v}$ and the vector $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ (why is this true?), and is

$$\begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & 0 \\ 0 & 0 & 1 \end{vmatrix} = \begin{bmatrix} v_2 \\ -v_1 \\ 0 \end{bmatrix}, \quad \text{or} \quad v_2 x - v_1 y = 0.$$

Notice that, in the domain, this is just the line $y = \frac{v_2}{v_1}x$, at least when $v_1 \neq 0$.

Now the graph of this line is the set of points

$$\left(x, \frac{v_2}{v_1}, f\left(x, \frac{v_2}{v_1}\right)\right) = \left(x, \frac{v_2}{v_1}, x^2\left(1 + \frac{v_2^2}{v_2^2}\right)\right) \in \mathbb{R}^3.$$

So let $f_{\mathbf{v}} = f\Big|_{y=\frac{v_2}{v_1}x} : \mathbb{R} \to \mathbb{R}$, where $f_{\mathbf{v}}(x) = x^2\left(1 + \frac{v_2^2}{v_1^2}\right)$. Here, of course, we have the data

$$f'_{\mathbf{v}}(0) = 2x\left(1 + \frac{v_2^2}{v_1^2}\right)\Bigg|_{x=0} = 0, \quad \text{and} \quad f''_{\mathbf{v}}(0) = 2\left(1 + \frac{v_2^2}{v_1^2}\right)\Bigg|_{x=0} > 0.$$

Hence, within the slice formed by $\mathbf{v}$, $f_{\mathbf{v}}$ is concave up, and, at least in this direction, according to the Second Derivative Test for an extremum from Calculus I, the point $\mathbf{0}$ corresponds to a local min of $f\Big|_{\mathbf{v}}$. And since this will be true of all choices of $\mathbf{v}$ (we did ntoot specify values of the entries v_1 and v_2), we can safely conclude that $\mathbf{a} = \mathbf{0}$ is a local minimum for f. Yes, I know, it is a global minimum on any domain that contains the origin

Now for $f : X \subset \mathbb{R}^n \to \mathbb{R}$ a C^2-function, we already have access to all of its second derivative information in the form of the Hessian of f,

$$HF(\mathbf{a}) = \begin{bmatrix} f_{x_1x_1}(\mathbf{a}) & \cdots & f_{x_1x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ f_{x_nx_1}(\mathbf{a}) & \cdots & f_{x_nx_n}(\mathbf{a}) \end{bmatrix}.$$

Recall that $D_{\mathbf{v}}f(\mathbf{a}) = Df(\mathbf{a})\mathbf{v}$. One can also show that $D^2_{\mathbf{v}}f(\mathbf{a}) = \mathbf{v}^T Hf(\mathbf{a})\mathbf{v}$ is the second directional derivative of f in the direction of $\mathbf{v}$. It directly measures the concavity of the curve which is the intersection of $\mathbf{graph}(f)$ with the slice determined by $\mathbf{v}$ at $\mathbf{a}$.

Exercise 1. Show that the second directional derivative of f in the direction of $\mathbf{v}$ is given by $D^2_{\mathbf{v}}f(\mathbf{a}) = \mathbf{v}^T Hf(\mathbf{a})\mathbf{v}$.

Hence if $D^2_{\mathbf{v}}f(\mathbf{a}) = \mathbf{v}^T Hf(\mathbf{a})\mathbf{v} > 0$, for every direction $\mathbf{v}$ at $\mathbf{a} \in X$, then we can be assured that there is a local minimum of f at $\mathbf{a}$.

Some notes:

- For every $n \times n$ matrix A, one can construct a *quadratic form*: A real-valued function $Q : \mathbb{R}^n \to \mathbb{R}$ defined by
$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \sum_{i=1}^{n} a_{ij} x_i x_j.$$
In dimension-1, any quadratic form will look like $Q(x) = ax^2$, and in dimension-2, with $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we have $Q(x,y) = ax^2 + (b+c)xy + dx^2$. In general, a quadratic form will be a polynomial in the variables given, with each monomial corresponding either to the product of two variables or the square of one of them, with total exponent 2 (sum the exponents of each of the factors).
- Quadratic forms are invariant under conjugations of the matrix defining the form. Hence we can always take A to be symmetric (which means $a_{ij} = a_{ji}$ throughout A, or $A^T = A$.
- Hessians are always symmetric! (why?)
- A quadratic form $Q(\mathbf{x})$ is called *positive definite* if $Q(\mathbf{x}) > 0$, for every $\mathbf{x} \neq \mathbf{0}$. (And *negative definite* if $Q(\mathbf{x}) < 0$, for every $\mathbf{x} \neq \mathbf{0}$.)

Theorem 12.6. *For $X \subset \mathbb{R}^n$ open, let $f : X \to \mathbb{R}$ be C^2 with a critical point $\mathbf{a} \in X$.*

(1) if $Hf(\mathbf{a})$ is positive definite, then f has a local minimum at $\mathbf{a}$.
(2) if $Hf(\mathbf{a})$ is negative definite, then f has a local maximum at $\mathbf{a}$.
*(3) If **det** $Hf(\mathbf{a}) \neq 0$, and neither positive nor negative definite, then $\mathbf{a}$ is a* saddle.

There is a mechanical process for determining when a matrix if positive or negative definite and it is all linear algebra. In essence, it involves testing the *leading principal minors* of $Hf(\mathbf{a})$ to see if $Hf(\mathbf{a})$ is positive definite, negative definite, or neither.

Indeed, let $Q(\mathbf{a}) = \mathbf{x}^T a_{n\times n}\mathbf{x}$ be a quadratic form. Define the *kth leading principal minor* of A to be the determinant of
$$A_k = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}.$$
This A_k is a $k \times k$ submatrix of A consisting of entries that are both in the first k rows and the first k columns of A. Of course, A has n of these:
$$A_1 = a_{11}. \quad A_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, A_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \ldots, \quad A_n = A.$$
So what can we say?

- If all of these leading principal minors are positive, so if $\mathbf{det}\, A_k > 0$, for $k = 1, \ldots, n$, then A, and hence $Q(\mathbf{x})$, is positive definite.
- A, and hence $Q(\mathbf{x})$, is negative definite if $\mathbf{det}\ A < 0$ for k-odd and $\mathbf{det}\ A > 0$ for k-even.
- A is called *indefinite* if neither of the two cases above holds but all of the leading principal minors are non-zero.

- $Q(\mathbf{x})$ is called *degenerate*, as is A, if $\mathbf{det}\ A = 0$, and nondegenerate otherwise. Note here that it is certainly possible that $Q(\mathbf{x})$ is nondegenerate but at least one of the leading principal minors is 0. Just take any nonsingular matrix with $a_{11} = 0$.

Lastly, the Extreme Value Theorem from single variable calculus has a counterpart in vector calculus. Recall that a set $X \subset \mathbb{R}^n$ is *closed* if it contains all of its boundary points. A set $X \subset \mathbb{R}^n$ is called *bounded* if there exists a real number $M > 0$ such that

$$||\mathbf{x}|| < M, \quad \forall \mathbf{x} \in X.$$

And a set $X \subset \mathbb{R}^n$ is called *compact* if it is both closed and bounded in $\mathbb{R}^n$.

Theorem 12.7 (The Extreme Value Theorem)**.** *If $X \subset \mathbb{R}^n$ is compact and $f : X \to \mathbb{R}$ is continuous, then f has a global maximum and a global minimum on X.*

Just like in single variable calculus, it is certainly possible for a function f on a possibly nonclosed or unbounded (or both)X to have global extrema. But it is only guaranteed to have each when X is compact and f is continuous.