1 Recap

1. Topology of $\mathbb{R}^n$ and invariance of domain.
2. Continuity, differentiability, chain rule.
3. Gradient, first derivative test, global extrema.

2 Review of multivariable calculus

Higher derivatives: It is of course possible that a function is differentiable and the derivative is continuous, and all the second partials exist but it is not twice-differentiable. Here is an even weirder example : $f(x,y) = \frac{xy(x^2-y^2)}{x^2+y^2}$ for $(x,y) \neq (0,0)$ and $f(0,0) = 0$. By properties, f is differentiable on $\mathbb{R}^2 - \{(0,0)\}$. At $(0,0)$ we claim that it is differentiable and the gradient is 0 : $|f(x,y) - 0 - 0| = |xy|\frac{|x^2-y^2|}{x^2+y^2} \leq x^2 + y^2$ (why ?) Therefore, if $\sqrt{x^2+y^2} < \epsilon$, we are done (Why ?) Now $f_x = \frac{y(x^2-y^2)(x^2+y^2)+2x^2y(x^2+y^2)-2x^2y(x^2-y^2)}{(x^2+y^2)^2} = y\frac{x^4-y^4-4x^2y^2}{(x^2+y^2)^2}$, $f_y = -x\frac{y^4-x^4-4x^2y^2}{(x^2+y^2)^2}$ away from $(0,0)$. Thus f_x, f_y are continuous throughout. However, $f_{xy} = \lim_{h\to 0} \frac{f_y(h,0)-f_y(0,0)}{h} = 1$ and $f_{yx} = \lim_{h\to 0} \frac{f_x(0,h)-f_x(0,0)}{h} = -1$. Therefore, $f_{xy} \neq f_{yx}$ in general !
Theorem (Clairaut) : If all the second partials exist and are continuous are a, then the mixed partials are equal.

A function is called C^2 if DF is differentiable and D^2F is continuous (and likewise for C^k). Notation: If $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ is a multi-index, often $\alpha! := \alpha_1!\alpha_2!\ldots$, and $|\alpha| = \alpha_1 + \alpha_2 + \ldots$. $\partial_\alpha f = \frac{\partial^{|\alpha|} f}{\partial(x^1)^{\alpha_1}\partial(x^2)^{\alpha_2}\ldots}$. If a function is C^k for all k, it is said to be C^∞ or smooth.
Proof of Clairaut: We will prove it for two-variable functions. The general case follows from this special case (why?) Let $u(h,k) = f(a+h, b+k) - f(a+h, b)$, $v(h,k) = f(a+h, b+k) - f(a, b+k)$, and $w(h,k) = f(a+h, b+k) - f(a+h, b) - f(a, b+k) - f(a,b)$. Now, $w(h,k) = u(h,k) - u(0,k) = h\partial_x u(c_1, k)$ by MVT for $c_1 \in [0,h]$. This equals $h(\partial_x f(a+c_1, b+k) - \partial_x f(a+c_1, b)) = hk\partial_y\partial_x f(a+c_1, b+c_2)$ where $c_2 \in [0,k]$. Likewise, $w(h,k) = v(h,k) - v(h,0) = hk\partial_x\partial_y f(a+d_1, b+d_2)$ where $d_1 \in [0,h], d_2 \in [0,k]$. Dividing by hk, we see that $\partial_y\partial_x f(a+c_1, b+c_2) = \partial_x\partial_y f(a+d_1, b+d_2)$. By assumption of continuity of the second partials, there exists a $\delta > 0$ such that $|\partial_y\partial_x f(a+c_1, b+c_2) - \partial_y\partial_x f(a,b)| < \frac{\epsilon}{2}$, and $|\partial_x\partial_y f(a+d_1, b+d_2) - \partial_x\partial_y f(a,b)| < \frac{\epsilon}{2}$ whenever $|(h,k)| < \delta$ (why ?). Thus, $|f_{xy}(a,b) - f_{yx}(a,b)| < \epsilon$ for all $\epsilon > 0$. We are done. □
Using the C^2 Clairaut, we can prove the C^k Clairaut for any k.

Theorem 1 (Taylor). *Let $U \subset \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}$ be a C^k function on U. Let $a \in U$ and $|h| < \epsilon$ such that $B_a(\epsilon) \subset U$. Then the polynomial $p_{a,k}(h) = f(a) + \sum_i \frac{\partial f}{\partial x_i}(a)h + \frac{1}{2}\sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(a)h_i h_j + \ldots + \sum_{\|\alpha\|=k} \frac{D^\alpha f(a)}{\alpha!}h^\alpha$ is the unique polynomial of degree $\leq k$ (degree meaning the maximum sum of powers) such that $\lim_{h\to 0} \frac{f(a+h)-p_{k,a}(h)}{|h|^k} = 0$. Moreover, if f is C^{k+1}, then $f(a+h) = p_{k,a}(h) + \sum_{|\alpha|=k+1} \frac{D^\alpha f(\eta)h^\alpha}{\alpha!}$, where η lies in $B_a(h)$.*

Proof. Uniqueness is easy and left as an exercise. Let $h \neq 0$ (if it is equal to 0, we are done). Consider the one-variable function $q(t) = f(a + t\frac{h}{\|h\|})$ on $|t| < \epsilon$. This function

is C^k (because it is a composition of C^k functions). Thus we can apply the one-variable Taylor theorem to it to conclude that $q(\|h\|) = q(0) + q'(0)\|h\| + \ldots$.
Now we claim inductively that $\frac{q^{(m)}(t)\|h\|^m}{m!} = \sum_{|\alpha|=m} \frac{D^\alpha f(a+t\frac{h}{\|h\|})h^\alpha}{\alpha!}$:
Indeed, for $m = 1$ we are done by the Chain rule. Assume the truth of this statement for $1, 2\ldots, m-1$. We apply the induction hypothesis to $q^{(m-1)}(t)$ to conclude that $\frac{q^{(m)}(t)\|h\|^m}{m!} = \frac{\|h\|}{m} \sum_{|\alpha|=m-1} \frac{d}{dt} \frac{D^\alpha f(a+t\frac{h}{\|h\|})h^\alpha}{\alpha!} = \sum_{|\alpha|=m-1} \sum_i \frac{\partial_{x_i} D^\alpha f(a+t\frac{h}{\|h\|})h_i h^\alpha}{\alpha! m} = \sum_i \sum_{|\alpha|=m-1} \frac{\partial_{x_i} D^\alpha f(a+t\frac{h}{\|h\|})h_i h^\alpha}{\alpha! m}$. We want to compare the last expression to $\sum_{|\beta|=m} \frac{D^\beta f(a+t\frac{h}{\|h\|})h^\beta}{\beta!}$.

Given i such that $\beta_i \geq 1$, every multi-index vector β can be written uniquely as $\beta = \alpha + e_i$ where $|\alpha| = m-1$. However, this can be done for each such i. Hence, if we fix β, then $\frac{1}{\alpha! m} = \frac{\alpha_i + 1 = \beta_i}{\beta! m}$ and if we sum over all i giving rise to the same β, then we get $\sum_i \frac{\beta_i}{\beta! m} = \frac{1}{\beta!}$. Hence these two expressions are the same, and we are done.

Now the one-variable Taylor theorem, i.e., the following theorem, completes the proof.

Theorem 2. *Let $U \subset \mathbb{R}$ be an open set and $f : U \to \mathbb{R}$ a C^k function on U. Let $a \in U$ and $|h| < \epsilon$ such that $(a-\epsilon, a+\epsilon) \in U$. Then the polynomial $p_{k,a}(h) = f(a) + f'(a)h + \frac{f''(a)h^2}{2!} + \ldots + \frac{f^{(k)}(a)h^k}{k!}$ is the unique polynomial of degree $\leq k$ such that $\lim_{h\to 0} \frac{f(a+h)-p_{k,a}(h)}{h^k} = 0$. Moreover, if f is C^{k+1}, then $f(a+h) = p_{k,a}(h) + \frac{f^{(k+1)}(\eta)h^{k+1}}{(k+1)!}$, where η lies between a and $a+h$.*

□

The second derivative test: Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}$ be a C^1 function. Suppose f is C^2 at $a \in U$ and $Df_a = 0$.
Theorem: If $\sum_{i,j} f_{ij}(a)v^i v^j > 0$ for all $v \neq 0$, then a is a local minimum. If $\sum_{i,j} f_{ij}(a)v^i v^j < 0$ for all $v \neq 0$, then a is a local maximum. Conversely, if a is a local minimum $\sum_{i,j} f_{ij}(a)v^i v^j \geq 0$ for all v (and likewise for a local max).
Proof: Taylor's theorem (with the Peano form of the remainder) implies that if $\sum_{i,j} f_{ij}(a)v^i v^j > 0$, $f(a+v) - f(a) > 0$ for all small v. Thus a is a local minimum. Likewise for local maxima. Conversely, if $\sum_{i,j} f_{ij}(a)v^i v^j < 0$ for some v, then $f(a+v) - f(a) < 0$ and hence a cannot be a local minimum. (Likewise for local maxima.) □

The matrix f_{ij} is called the 'Hessian' matrix of f. For future purposes a twice differentiable function whose second partials are continuous is said to be convex if $\sum_{i,j} v_i f_{ij}(a) v_j \geq 0 \ \forall\ v \neq 0$. Often, optimization studies convex functions. In general, given a symmetric $n \times n$ real matrix A, i.e., $A^T = A$, i.e., $A_{ij} = A_{ji}$, it is said to be *positive-definite* if $\sum_{i,j} A_{ij} v_i v_j > 0$ for all $v \neq 0$, i.e., $v^T A v > 0$ for all $v \neq 0$. This condition is somewhat subtle. For 2×2 matrices, $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$, supposing $v^T = (x, y)$, we see that $v^T A v = ax^2 + dy^2 + 2bxy$. This expression is positive for all $(x, y) \neq (0, 0)$ if and only $a > 0$ and $ad - b^2 = \det(A) > 0$. (Why ?) Similar but more complicated conditions exist for $n \times n$ matrices. (Positive-definiteness is the same as having only positive eigenvalues by the way.)

Example: Find local extrema of $f(x,y) = x^2 - y^2$ on $\mathbb{R}^2$: $\nabla f = (2x, -2y) = (0,0)$ only at the origin. The Hessian is $\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$. So the origin is neither a local max nor a local min (despite the second derivative being non-zero ! (and none of the eigenvalues being zero)). There is a direction in which f increases and a direction in which it decreases. Such points are called “Saddle points". More generally, the eigenvectors and eigenvalues of the Hessian tell us about these “principal" directions.