Math 247 Part II: Differential Calculus

Calculus III (Advanced Version) with Professor Henry Shum

David Duan, 2019 Winter

Contents

1 Directional and Partial Derivatives

1.1 Directional Derivatives

Given a unit vector $\vec{u}$ and a point $\vec{a}$, the directional derivative $D_{\vec{u}}f(\vec{a})$ can be seen as the instantaneous rate of change of the function, moving through $\vec{a}$ in the direction specified by $\vec{u}$:

Definition 1.1.1 Let $A \subseteq \mathbb{R}^n$ be a set with non-empty interior and let $f : A \to \mathbb{R}^m$ be a function. Given $\vec{a} \in \text{int}(A)$ and $\vec{u} \in \mathbb{R}^n$ with $\|\vec{u}\| = 1$, the **directional derivative** of f at $\vec{a}$ in the direction of $\vec{u}$ is defined as

$$D_{\vec{u}}f(\vec{a}) := \lim_{h\to 0} \frac{f(\vec{a}+h\vec{u}) - f(\vec{a})}{h}$$

if the limit exists.

The directional derivative exists at a point for a non-scalar function f iff the directional derivative exists at this point for all its component functions f_i:

Proposition 1.1.2 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$, and $\vec{u} \in \mathbb{R}^n$ be a unit vector. Let $f : A \to \mathbb{R}^m$ be a function with components $f_i : A \to \mathbb{R}$, $i = 1, 2, \dots, m$. Then $D_{\vec{u}}f(\vec{a})$ exists if and only if $D_{\vec{u}}f_i(\vec{a})$ exists for each $i = 1, 2, \dots, m$. Furthermore, if $D_{\vec{u}}f(\vec{a})$ exists, then

$$D_{\vec{u}}f(\vec{a}) = \Big(D_{\vec{u}}f_1(\vec{a}), D_{\vec{u}}f_2(\vec{a}), \dots, D_{\vec{u}}f_m(\vec{a})\Big).$$

We can use Definition 1.1.1 to calculate directional derivatives. Later in Theorem 2.2.1 we will see another approach (using Jacobian matrix).

Example 1.1.3 Let $f(x,y) = xy$ and $\vec{v} = (2,2)$. Calculate $\frac{\partial f}{\partial v}(1,1)$.

Solution. The unit vector $\vec{u}$ in the direction of $\vec{v} = (1,1)$ is $\vec{u} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. Then

$$\begin{aligned} D_{\vec{u}_1}f(\vec{a}) &= \lim_{h\to 0} \frac{f(\vec{a}+h\vec{u}) - f(\vec{a})}{h} \\ D_{\vec{u}_1}f(1,1) &= \lim_{h\to 0} \frac{f((1,1) + h(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})) - f(1,1)}{h} \\ &= \lim_{h\to 0} \frac{(1+\frac{1}{\sqrt{2}}h)^2 - 1}{h} \\ &= \lim_{h\to 0} \frac{\sqrt{2}h + \frac{1}{2}h^2}{h} \\ &= \lim_{h\to 0} \left(\sqrt{2} + \frac{1}{2}h\right) = \sqrt{2}. \end{aligned}$$

□

1.2 Partial Derivatives

The partial derivative can be seen as a special case of directional derivatives, where the unit vector $\vec{u}$ is chosen from the canonical basis of $\mathbb{R}^n$. Intuitively, this is the derivative of f with respect to one of its variables, with the others held constant:

Definition 1.2.1 Let $\{\vec{e}_j : 1 \leq j \leq n\}$ be the standard basis of $\mathbb{R}^n$. Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. Given $\vec{a} \in \text{int}(A)$, the **partial derivative** of f with respect to x_j at $\vec{a}$ is defined as

$$\frac{\partial f}{\partial x_j}(\vec{a}) := D_{\vec{e}_j} f(\vec{a}).$$

Equivalently, this is defined as

$$\frac{\partial f}{\partial x_j}(\vec{a}) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_j + h, \ldots, a_n) - f(a_1, \ldots, a_j, \ldots, a_n)}{h}.$$

As in Proposition 1.1.2, the partial derivative exists at a point for a non-scalar function f iff the partial derivative exists at this point for all its component function f_i:

Proposition 1.2.2 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$, and $f : A \to \mathbb{R}^m$. If $\frac{\partial f}{\partial x_j}(\vec{a})$ exists for some $j \in \{1, 2, \ldots, n\}$, then $\frac{\partial f_i}{\partial x_j}(\vec{a})$ exists for all $i \in \{1, 2, \ldots, m\}$ and

$$\frac{\partial f}{\partial x_j}(\vec{a}) = \left(\frac{\partial f_1}{\partial x_j}(\vec{a}), \frac{\partial f_2}{\partial x_j}(\vec{a}), \ldots, \frac{\partial f_m}{\partial x_j}(\vec{a}) \right).$$

Computing partial derivatives is much easier.

Example 1.2.3 Find the partial derivatives $\frac{\partial f}{\partial x}(x, y)$ and $\frac{\partial f}{\partial y}(x, y)$ for the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = x + \sin(x + y)$.

Solution. Fix one variable and take the derivative as in single-variable case:

$$\frac{\partial f}{\partial x}(x, y) = 1 + \cos(x + y), \quad \frac{\partial f}{\partial y}(x, y) = \cos(x + y). \qquad \square$$

2 Differentiability, Jacobian, Gradient

2.1 Differentiability and Derivative

Recall the following definition from Math 147. The left expression is more familiar but the right one allows us to generalize it easier. In higher dimensions, the derivative f' becomes a linear map T (where both are evaluated at a point a).

Definition 2.1.1 Let $f : \mathbb{R} \to \mathbb{R}$ be a function. If f is **differentiable** at a, then

$$f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h} \iff \lim_{h\to 0}\left(\frac{f(a+h)-f(a)-f'(a)h}{h}\right) = 0.$$

A multivariate function is differentiable at a point $\vec{a}$ if there exists a linear map satisfying the following condition. Note that we restrict $\vec{a} \in \text{int}(A)$ because otherwise the derivative may fail to be unique. This is necessary for all results related to derivatives and differentiability.

Definition 2.1.2 Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. We say that f is **differentiable** at $\vec{a} \in \text{int}(A)$ if there exists a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\vec{h}\to\vec{0}} \frac{\|f(\vec{a}+\vec{h}) - f(\vec{a}) - T(\vec{h})\|}{\|\vec{h}\|} = 0.$$

If such a map T exists, then it is called the **derivative of f at $\vec{a}$**. We use the notation $Df(\vec{a})$ for the (unique) derivative of f at $\vec{a}$, i.e., $Df(\vec{a}) = T$.

Similar to Proposition 1.1.2 and 1.2.2, the derivative exists at a point for a non-scalar function f iff the derivative exists at this point for all its component function f_i:

Proposition 2.1.3 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$, and $f : A \to \mathbb{R}^m$. Then, $Df(\vec{a}) = T$ if and only if $Df_i(\vec{a}) = T_i$ for each $i \in \{1, 2, \dots, m\}$. In other words, differentiability of f is equivalent to differentiability of all component functions f_i.

From Math 147, we know that differentiability implies continuity. Later in Section 2.3, we will use its contrapositive: a function is not continuous thus not differentiable at a point.

Theorem 2.1.4 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$, and $f : A \to \mathbb{R}^m$. If f is differentiable at $\vec{a}$, then it is continuous at $\vec{a}$.

Here is an equivalent definition for differentiability (not tested but very helpful). If the "error function" is continuous and evaluates to zero at $\vec{a}$, then the function is differentiable at $\vec{a}$. Note that this theorem gives an easy proof for Theorem 2.1.4, as $f(\vec{x})$ can be expressed as a sum of continuous functions.

Theorem 2.1.5 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$, and $f : A \to \mathbb{R}^m$. The function f is differentiable at $\vec{a}$ if and only if there is a linear map $l : \mathbb{R}^n \to \mathbb{R}^m$ and a function $r : A \to \mathbb{R}^m$ that is continuous at $\vec{a}$ and satisfies $r(\vec{a}) = \vec{0}$, such that $f(\vec{x}) = f(\vec{a}) + l(\vec{x} - \vec{a}) + r(\vec{x})\|\vec{x} - \vec{a}\|$.

2.2 Jacobian Matrix as Derivative

If the function is differentiable at $\vec{a}$, then all of its directional derivatives (and thus all the partial derivatives) exist at $\vec{a}$ and the linear map T is given by the Jacobian matrix.

Theorem 2.2.1 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$, and $f : A \to \mathbb{R}^m$. Suppose f is differentiable at $\vec{a}$ and let $T = Df(\vec{a})$ be the derivative of f at $\vec{a}$. Then:

1. For every unit vector $\vec{u} \in \mathbb{R}^n$, the directional derivative of f at $\vec{a}$ in the direction $\vec{u}$ exists and is $D_{\vec{u}} f(\vec{a}) = T(\vec{u})$.
2. All partial derivatives $\frac{\partial f_i}{\partial x_j}(\vec{a})$, $1 \le i \le m$, $1 \le j \le n$, exists.
3. The $m \times n$ matrix representing T in the standard basis is called the **Jacobian matrix**:

$$J(\vec{a}) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\vec{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\vec{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\vec{a}) & \cdots & \frac{\partial f_m}{\partial x_n}(\vec{a}) \end{pmatrix}.$$

When f is differentiable at $\vec{a}$, the Jacobian matrix defines a linear map $\mathbb{R}^n \to \mathbb{R}^m$, which is the best (pointwise) linear approximation of the function f near $\vec{a}$. This linear map is thus the generalization of the usual notion of derivative, and is called the derivative or the differential of f at $\vec{a}$. It is common to treat $Df(\vec{a})$ as the Jacobian matrix $J(\vec{a})$, rather than the linear transformation T represented by the Jacobian matrix.

$$J = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \implies J_{ij} = \frac{\partial f_i}{\partial x_j}.$$

We now use Theroem 2.2.1 (1) to compute directional derivatives. Note that J for a scalar function is a row vector and the unit vector $\vec{u}$ is a column vector (elements separated by commas).

Example 2.2.2 Let $f(x, y) = xy$ and $\vec{v} = (2, 2)$. Calculate $\frac{\partial f}{\partial v}(1, 1)$.

Solution. We first calculate the Jacobian matrix: $J = \begin{pmatrix} y & x \end{pmatrix} \implies J(1, 1) = \begin{pmatrix} 1 & 1 \end{pmatrix}$. The unit vector $\vec{u}$ in the direction of $\vec{v} = (1, 1)$ is $\vec{u} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. Then by Theroem 2.2.1 (1),

$$D_{\vec{u}} f(1,1) = T(\vec{u}) = J\vec{u} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = 2\frac{1}{\sqrt{2}} = \sqrt{2}. \qquad \square$$

2.3 Sufficient Conditions for Differentiability

The following theorem gives sufficient (but not necessary, i.e., a function not satisfying it can still be differentiable, but a function satisfying it is definitely differentiable) conditions for differentiability: if all partials exist on an open ball centered at $\vec{a}$ and are continuous at $\vec{a}$, then the function must be differentiable.

Theroem 2.3.1 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$, and $f : A \to \mathbb{R}^m$. Let $r > 0$ be such that $B_r(\vec{a}) \subseteq A$. If all partial derivatives $\frac{\partial f_i}{\partial x_j}(\vec{x})$, $1 \leq i \leq m$, $1 \leq i \leq n$, exists on $B_r(\vec{a})$ and are continuous at $\vec{a}$, then f is differentiable at $\vec{a}$.

We now make two very important remarks.

Remark 2.3.2

1. All partial derivatives of f exists at $\vec{a}$ does not guarantee the differentiability of f at $\vec{a}$. For example, consider $f(x, y) = \sqrt{|xy|}$ defined on $\mathbb{R}^2$ is not differentiable at $(0, 0)$, because the function is not continuous at $(0, 0)$.
2. Theorem 2.3.1 gives sufficient but not necessary conditions for differentiability. For example, consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} 0 & (x,y) = (0,0) \\ (x^2 + y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) & \text{otherwise} \end{cases}$$

   The partial derivatives exist but are not continuous at $(0, 0)$. Nevertheless, $Df(0, 0)$ exists and is the zero map.

2.4 Gradient

The gradient is a multi-variable generalization of the derivative of a scalar function.

Definition 2.4.1 The **gradient** of a function $f : A \to \mathbb{R}$ at $\vec{a}$ is the column vector

$$\nabla f(\vec{a}) := \begin{bmatrix} \frac{\partial f}{\partial x_1}(\vec{a}) \\ \frac{\partial f}{\partial x_2}(\vec{a}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\vec{a}) \end{bmatrix}$$

i.e., the vector of partial derivatives.

There is an obvious connection between gradient and Jacobian.

Remark 2.4.2

1. Let f be a scalar-valued function. The gradient is a column vector whereas the Jacobian is a row vector. Thus, the gradient is equal to the transpose of the Jacobian:

$$\nabla f(a) = [Df(\vec{a})]^T, \quad Df(\vec{x})(\vec{y}) = \langle \nabla f(\vec{x}), \vec{y} \rangle .$$

2. It is common to extend Definition 2.4.1 to vector-valued functions; the gradient of a function is always the transpose of the Jacobian.

The following theorem (remark: not tested) motivates the famous Gradient Descent algorithm. In words, the gradient points in the direction of steepest ascent. This is the key intuition for the gradient.

Theroem 2.4.3 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$, $f : A \to \mathbb{R}$. If f is differentiable at $\vec{a}$, then:

1. The vector $\vec{v} := (\nabla f(\vec{a}), -1) \in \mathbb{R}^{n+1}$ is orthogonal to the tangent hyperplane of the hypersurface $x_{n+1} = f(\vec{x})$ at the point $(\vec{a}, f(\vec{a}))$.
2. If $\nabla f(\vec{a}) \neq \vec{0}$, then the directional derivative $D_{\vec{u}} f(\vec{a})$ is maximized over all unit vectors $\vec{a} \in \mathbb{R}^n$ at $\vec{u} = \frac{\nabla f(\vec{a})}{\|\nabla f(\vec{a})\|}$ and minimized at $\vec{u} = -\frac{\nabla f(\vec{a})}{\|\nabla f(\vec{a})\|}$.

We sometimes use the differential operator ∇ (often called nabla or del) as a vector.

Remark 2.4.4 We can view the differential operator $\nabla := \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right)$ as a vector and make use of things like dot product (this is used in Section 6, Taylor's Theorem):

1. $[(\vec{h} \cdot \nabla) f](\vec{a}) = \left[\left(h_1 \dfrac{\partial}{\partial x_1} + \cdots + h_n \dfrac{\partial}{\partial x_n}\right) f\right](\vec{a})$
2. For $n = 2$: $[(\vec{h} \cdot \nabla)^2 f](\vec{a}) = \left[\left(h_1 \frac{\partial}{\partial x} + h_2 \frac{\partial}{\partial y}\right)^2 f\right](\vec{a}) = h_1^2 \frac{\partial^2 f}{\partial x^2}(\vec{a}) + 2h_1 h_2 \frac{\partial^2 f}{\partial x \partial y}(\vec{a}) + h_2^2 \frac{\partial^2 f}{\partial y^2}(\vec{a})$.

2.5 Summary: Proving Differentiability

There are three ways of proving a function is differentiable.

1. By Definition 2.1.2, if we could find a linear map (or the Jacobian matrix) satisfying $\lim_{\vec{h} \to \vec{0}} \frac{\|f(\vec{a}+\vec{h}) - f(\vec{a}) - T(\vec{h})\|}{\|\vec{h}\|} = 0$, then the function is differentiable.
2. By Theorem 2.3.1, if we could find an open ball near the given point where all partial derivatives exist and are continuous at the point, then the function is differentiable.

3. By Theorem 3.1.1 and 3.1.2, arbitrary linear combinations and compositions of differentiable functions are differentiable.

3 Rules for Differentiating Functions

As in single-variable case, we have the following rules for differentiating functions. We include the proof for Theorem 3.1.2 as it is testable on the final exam.

Theorem 3.1.1 (Chain Rule) Let $A \subseteq \mathbb{R}^n$, $B \subseteq \mathbb{R}^m$, $f : A \to B$, $g : B \to \mathbb{R}^l$. Suppose $\vec{a} \in \text{int}(A)$ and $\vec{b} := f(\vec{a}) \in \text{int}(B)$. If f is differentiable at $\vec{a}$ and g is differentiable at $f(\vec{a})$, then the composition $h := g \circ f$ defined by $h(\vec{x}) = g(f(\vec{x}))$ is differentiable at $\vec{a}$ and the derivative is
$Dh(\vec{a}) = Dg(f(\vec{a})) \circ Df(\vec{a})$.

Theorem 3.1.2 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$. If $f, g : A \to \mathbb{R}$ are two functions that are differentiable at $\vec{a}$, then:

1. *Addition:* $D(f+g)(\vec{a}) = Df(\vec{a}) + Dg(\vec{a})$,
2. *Scalar Multplication:* $D(\alpha f)(\vec{a}) = \alpha Df(\vec{a})$ for any $\alpha \in \mathbb{R}$,
3. *Product Rule:* $D(fg)(\vec{a}) = g(\vec{a})Df(\vec{a}) + f(\vec{a})Dg(\vec{a})$, and
4. *Quotient Rule:* $D(f/g)(\vec{a}) = \frac{g(\vec{a})Df(\vec{a}) - f(\vec{a})Dg(\vec{a})}{g(\vec{a})^2}$ given $g(\vec{a}) \neq 0$.

Proof. Apply the Chain Rule to the suitable composite functions.

Addition: Let $h : A \to \mathbb{R}^n$ be defined by $h(\vec{x}) := (f(\vec{x}), g(\vec{x}))$. Define $s : \mathbb{R}^2 \to \mathbb{R}$ by $s(x, y) = x + y$. Note that $f + g = s \circ h$. By the Chain Rule, $D(f+g)(\vec{a}) = Ds(h(\vec{a})) \circ Dh(\vec{a})$. Writing the derivatives as Jacobian matrices,

$$J_{f+g}(\vec{a}) = J_s(h(\vec{a}))J_h(\vec{a}) = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} J_f(\vec{a}) \\ J_g(\vec{a}) \end{pmatrix} = J_f(\vec{a}) + J_g(\vec{a})$$

Thus $D(f+g)(\vec{a}) = Df(\vec{a}) + Dg(\vec{a})$. ■

Scalar Multiplication: Fix $\alpha \in \mathbb{R}$. Let $s : \mathbb{R} \to \mathbb{R}$ defined by $s(x) = \alpha x$ and $h : A \to \mathbb{R}$ defined by $h(\vec{a}) = \alpha f(\vec{a})$. Note that $h = s \circ f$. By the Chain Rule, $Dh(\vec{a}) = Ds(f(\vec{a})) \circ Df(\vec{a})$. Since $s : \mathbb{R} \to \mathbb{R}$ is a one-variable scalar function, we know $s(x) = \alpha x \implies s'(x) = \alpha$ (Math147). It follows that
$D(\alpha f)(\vec{a}) = Dh(\vec{a}) = Ds(f(\vec{a})) \circ Df(\vec{a}) = \alpha Df(\vec{a})$ as desired. ■

Product Rule: Let $h : A \to \mathbb{R}^n$ be defined by $h(\vec{x}) := (f(\vec{x}), g(\vec{x}))$. Define $s : \mathbb{R}^2 \to \mathbb{R}$ by $s(x, y) = xy$. Note that $fg = s \circ h$. By the Chain Rule, $D(fg)(\vec{a}) = Ds(h(\vec{a})) \circ Dh(\vec{a})$. Writing the derivatives as Jacobian matrices,

$$J_{fg}(\vec{a}) = J_s(h(\vec{a}))J_h(\vec{a}) = \begin{pmatrix} g(\vec{a}) & f(\vec{a}) \end{pmatrix} \begin{pmatrix} J_f(\vec{a}) \\ J_g(\vec{a}) \end{pmatrix} = g(\vec{a})J_f(\vec{a}) + f(\vec{a})J_g(\vec{a}).$$

Thus, $D(fg)(\vec{a}) = g(\vec{a})Df(\vec{a}) + f(\vec{a})Dg(\vec{a})$. ■

Quotient Rule: Let $h : A \to \mathbb{R}^2$ be defined by $h(\vec{x}) := (f(\vec{x}), g(\vec{x}))$. Define $s : \mathbb{R}^2 \to \mathbb{R}$ by $s(x, y) = x/y$. Note that $fg = s \circ h$. By the Chain Rule, $D(fg)(\vec{a}) = Ds(h(\vec{a})) \circ Dh(\vec{a})$. Suppose $g(\vec{a}) \neq 0$. Writing the derivatives as Jacobian matrices,

$$J_{f/g}(\vec{a}) = J_s(h(\vec{a}))J_h(\vec{a}) = \begin{pmatrix} \frac{1}{g}(\vec{a}) & -\frac{f}{g^2}(\vec{a}) \end{pmatrix} \begin{pmatrix} J_f(\vec{a}) \\ J_g(\vec{a}) \end{pmatrix} = \frac{J_f(\vec{a})}{g(\vec{a})} - \frac{f(\vec{a})J_g(\vec{a})}{g^2(\vec{a})} = \frac{g(\vec{a})J_f(\vec{a}) - f(\vec{a})J_g(\vec{a})}{g^2(\vec{a})}.$$

Thus, $D(f/g)(\vec{a}) = \frac{g(\vec{a})Df(\vec{a}) - f(\vec{a})Dg(\vec{a})}{g(\vec{a})^2}$ if $g(\vec{a}) \neq 0$. $\square$

4 Important Results

4.1 Mean Value Theorem

Recall MVT for single variable from Math 147. MVT states that

1. *There exists a point in the chosen interval where the slope of the tangent is the same as the slope of the straight line joining the end-points of that interval.*
2. *There exists a point in the chosen interval where the instantaneous rate of change of the function is equal to the average rate of change over the entire interval.*
3. *There exists a point in the chosen interval over which the function is continuous and differentiable (but no further information on its position).*

Remark 4.1.1 (MVT in $\mathbb{R}$) Suppose $f(x)$ is a function that satisfies both of the following:

1. $f(x)$ is continuous on the closed interval $[a, b]$.
2. $f(x)$ is differentiable on the open interval (a, b).

Then there exists $c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$.

We now generalize MVT to $\mathbb{R}^n$. There is a nice correspondence between each statement in 4.1.1 vs. 4.1.2: first, we want the path between $\vec{a}$ and $\vec{b}$ to be in the interior of the domain where the derivative is defined; next, we want f to be differentiable on the path and continuous on the closure of the path; finally, we state the existence of $\vec{c}$ on the path.

Theorem 4.1.2 (MVT in $\mathbb{R}^n$) Let $A \subseteq \mathbb{R}^n$, $\vec{a}, \vec{b} \in A$ and define $S := \{\vec{a} + t(\vec{b} - \vec{a}) : 0 < t < 1\}$. Suppose $S \subseteq \text{int}(A)$ and $\overline{S} \subseteq A$. If $f : A \to \mathbb{R}$ is continuous on $\overline{S}$ and differentiable on S, then there exists $\vec{c} \in S$ such that $f(\vec{b}) - f(\vec{a}) = Df(\vec{c})(\vec{b} - \vec{a})$.

Using MVT, we can get the following conclusions:

1. *If the Jacobian is the zero map for all $\vec{a} \in A$, then the function is constant on A.*
2. *If two functions have the same Jacobian matrix, then they differ by only a constant.*

4.2 Linear Approximation

Definition 4.2.1 Recall (from Theorem 2.1.5, alternative definition of differentiability) that if a function $f : A \to \mathbb{R}$ is differentiable at $\vec{a}$, then $f(\vec{x}) = f(\vec{a}) + Df(\vec{a})(\vec{x} - \vec{a}) + r(\vec{x} - \vec{a})$, where the function r is continuous at $\vec{a}$ and satisfies $\lim_{\vec{x} \to \vec{a}} \frac{|r(\vec{x} - \vec{a})|}{\|\vec{x} - \vec{a}\|} = 0$. Then, if $\|\vec{x} - \vec{a}\|$ is small, $|r(\vec{x} - \vec{a})|$ should be small, i.e., $f(\vec{x}) \approx \ell_a^f := f(\vec{a}) + Df(\vec{a})(\vec{x} - \vec{a})$. Thus, we could use ℓ_a^f, called the **linear approximation** to f at $\vec{a}$, to approximate the value of f at $\vec{x}$ near $\vec{a}$.

We do a few examples on how to use this to approximate the value of a function at a given point.

Example 4.2.1 Find the linear approximation to the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) := x^2 \sin(y)$ at the point $\vec{a} = (x, y) = (\sqrt{2}, \pi/4)$.

Solution. First, we compute the Jacobian: $Df(\vec{a}) = Df(x, y) = \begin{pmatrix} 2x\sin(y) & x^2\cos(y) \end{pmatrix}$. Then

$$\begin{aligned} l^f_{\vec{a}} = f(\vec{a}) + Df(\vec{a})(\vec{x} - \vec{a}) &= f(\sqrt{2}, \pi/4) + Df(\sqrt{2}, \pi/4)((x, y) - (\sqrt{2}, \pi/4)) \\ &= 2\sin(\pi/4) + \begin{pmatrix} 2\sin(\pi/4)/\sqrt{2} & 2\cos(\pi/4) \end{pmatrix} \begin{pmatrix} x - \sqrt{2} \\ y - \pi/4 \end{pmatrix} \\ &= \sqrt{2} + \begin{pmatrix} 2 & \sqrt{2} \end{pmatrix} \begin{pmatrix} x - \sqrt{2} \\ y - \pi/4 \end{pmatrix} \\ &= -\sqrt{2}(1 + \pi/4) + 2x + \sqrt{2}y. \end{aligned}$$

□

Example 4.2.2 Find the linear approximation to the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) := \int_x^y e^{-t^2}\,dt$ at the point $\vec{a} = (x, y) = (2, 2)$.

Solution. By the Fundamental Theorem of Calculus, $\frac{\partial f}{\partial x}(x, y) = -e^{-x^2}$ and $\frac{\partial f}{\partial y}(x, y) = -e^{-y^2}$. Then

$$\begin{aligned} l^f_{\vec{a}} = f(\vec{a}) + Df(\vec{a})(\vec{x} - \vec{a}) &= f(2, 2) + Df(2, 2)((x, y) - (2, 2)) \\ &= 0 + \begin{pmatrix} -e^{-4} & e^{-4} \end{pmatrix} \begin{pmatrix} x - 2 \\ y - 2 \end{pmatrix} \\ &= e^{-4}(y - x). \end{aligned}$$

□

4.3 Implicit Function Theorem

The implicit function theorem gives sufficient conditions on a function ϕ so that the equation $\phi(\vec{x}, \vec{y}) = \vec{0}$ can be solved for $\vec{y}$ in terms of $\vec{x}$ (or vice versa) locally near a base point $(\vec{x}_0, \vec{y}_0)$ that satisfies the same equation $\phi(\vec{x}_0, \vec{y}_0) = \vec{0}$.

Theorem 4.3.1 Let $U \subseteq \mathbb{R}^{n+m}$ be non-empty and open, and let $\phi \in C^1(U, \mathbb{R}^m)$. Suppose there exists $\vec{x}_0 \in \mathbb{R}^n$ and $\vec{y}_0 \in \mathbb{R}^m$ with $(\vec{x}_0, \vec{y}_0) \in U$ satisfying $\phi(\vec{x}_0, \vec{y}_0) = \vec{0}$ and $\det(D_{\vec{y}}\phi(\vec{x}_0, \vec{y}_0)) \neq 0$. Then, there exists $a, b > 0$ such that $B_a(\vec{x}_0) \times B_b(\vec{y}_0) \subseteq U$ and a function $f \in C^1(B_a(\vec{x}_0), B_b(\vec{y}_0))$ such that $\phi(x, f(\vec{x})) = \vec{0}$ for all $\vec{x} \in B_a(\vec{x}_0)$. Moreover, this function f is unique.

5 Higher Order Derivatives

5.1 Higher Order Derivatives

Consider $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}$.

We have already defined the first order partial derivatives $\frac{\partial f}{\partial x_i}(\vec{a})$ for $\vec{a} \in \text{int}(A)$ and $i \in \{1, 2, \ldots, n\}$. The function $\frac{\partial f}{\partial x_i} : \text{int}(A) \to \mathbb{R}$ is called a first order derivative function.

We can define second order partial derivative functions as follows: for $j \in \{1, 2, \ldots, n\}$,

$$\frac{\partial^2 f}{\partial x_j \partial x_i} := \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right),$$

for points where this partial derivative exists.

We can inductively define higher order partial derivatives:

$$\frac{\partial^m f}{\partial_{i_m} \partial_{i_{m-1}} \cdots \partial_{i_2} \partial_{i_1}} := \frac{\partial}{\partial x_{i_m}}\left(\frac{\partial^{m-1} f}{\partial_{i_{m-1}} \cdots \partial_{i_2} \partial_{i_1}}\right),$$

where $i_k \in \{1, 2, \ldots, n\}$ for each $k \in \{1, 2, \ldots, m\}$.

It is common to use the following notation:

1. $\dfrac{\partial f}{\partial x_i} = f_{x_i}$,
2. $\dfrac{\partial f}{\partial x} = f_x$, $\frac{\partial f}{\partial y} = f_y$,
3. $\dfrac{\partial^2 f}{\partial x_j \partial x_i} = f_{x_i x_j}$ (order reversed!),
4. $\dfrac{\partial^2 f}{\partial x \partial y} = f_{yx}$,
5. $\dfrac{\partial^2 f}{\partial x \partial x} = \dfrac{\partial^2 f}{\partial x \partial x} = f_{xx}$.

5.2 Mixed Partial Derivatives

When can we exchange the order of partial derivatives? The following theorem gives us sufficient but not necessary conditions of mixed partial derivatives: if both first-order and second-order partials exist and are continuous at $\vec{a}$, then you can exchange the order of second-order partials.

Theorem 5.2.1 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$, $f : A \to \mathbb{R}$, and $B_\delta(\vec{a}) \subseteq A$ for some $\delta > 0$. Suppose that for some $i, j \in \{1, 2, \ldots, n\}$ that partial derivatives $f_{x_i}, f_{x_j}, f_{x_i x_j}, f_{x_j x_i}$ exists on $B_\delta(\vec{a})$. If $f_{x_i x_j}$ and $f_{x_j x_i}$ are continuous at $\vec{a}$, then $f_{x_i x_j}(\vec{a}) = f_{x_j x_i}(\vec{a})$.

6 Taylor's Theorem

6.1 Single Variable Taylor's Theorem

We can use Taylor series to approximate the value of a function f evaluated at x near a:

Theorem 6.1.1 Let $I \subseteq \mathbb{R}$ be an interval and let $a \in I$. If the function $f : I \to \mathbb{R}$ is $(p+1)$-times differentiable on I for some integer $p \geq 0$, then for any $x \in I$ with $x \neq a$, there exists ξ between x and a such that

$$\begin{aligned} f(x) &= \sum_{k=0}^{p} \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(p+1)}(\xi)}{(p+1)!}(x-a)^{p+1} \\ &= \underbrace{f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(p)}(a)}{p!}(x-a)^p}_{p\text{th order Taylor polynomial}} + \underbrace{\frac{f^{(p+1)}(\xi)}{(p+1)!}(x-a)^{p+1}}_{\text{Lagrange Remainder}}. \end{aligned}$$

6.2 Continuously Differentiable Functions

Let $C(A, \mathbb{R})$ denotes the set of continuous functions from $A \subseteq \mathbb{R}^n$ to $\mathbb{R}$.

If $A \subseteq \mathbb{R}^n$ is open, then $C^k(A, \mathbb{R})$ denotes the set of functions from A to $\mathbb{R}$ with all partial derivatives of order up to and including k being continuous on A.

If $A \subseteq \mathbb{R}^n$ is not open and has non-empty interior, then $C^k(A, \mathbb{R})$ denotes the set of functions from A to $\mathbb{R}$ with all partial derivatives of order up to and including k being continuous on $\text{int}(A)$ and extendable to continuous functions on A.

We sometimes shorten the notation to $f \in C^k(A)$ and say the function $f \in C^k(A, \mathbb{R})$ is **of class C^k** (on A). Note that, if $A \subseteq \mathbb{R}$, then $C^k(A, \mathbb{R})$ is the k-times continuously differentiable real-valeud functions.

For continuously differentiable functions, we get this nice corollary from Theorem 5.2.1, which states that we can freely switch the order of partials:

Corollary 6.2.1 Let $A \subseteq \mathbb{R}^n$ be nonempty and open. Suppose $f \in C^k(A, \mathbb{R})$ for some integer $k \geq 2$. For any partial derivative of order $I \in \{2, 3, \ldots, k\}$, the order in which partial derivatives are token does not matter.

6.3 Multivariate Taylor's Theorem

We generalize Taylor's theorem to higher dimensions. Read Remark 2.4.4 for notations.

Theorem 6.2.2 Let U be an open and convex set and let $f \in C^{p+1}(U, \mathbb{R})$ for some integer $p \geq 0$. Suppose $\vec{a}, \vec{x} \in U$ and define $\vec{h} := \vec{x} - \vec{a}$. Then, there exists $\xi \in (0,1)$ such that

$$f(\vec{x}) = f(\vec{a}+\vec{h}) = f(\vec{a}) + \sum_{k=1}^{p} \frac{1}{k!}\left[\left(\vec{h}\cdot\nabla\right)^{k} f\right](\vec{a}) + R_p(\xi),$$

where

$$R_p(t) := \frac{1}{(p+1)!}\left[\left(\vec{h}\cdot\nabla\right)^{(p+1)} f\right](\vec{a}+t\vec{h}).$$

We can write Taylor's Theorem in a different way using the notation:

$$(D^{(k)}f)_{\vec{a}}(\vec{h}) = \begin{cases} \displaystyle\sum_{j_1,\dots,j_k=1}^{n} \frac{\partial^k f}{\partial x_{j_1},\dots,\partial x_{j_k}}(\vec{a}) h_{j_1}\cdots h_{j_k} & k \geq 1, \\ f(\vec{a}) & k = 0. \end{cases}$$

The result of Taylor's Theorem becomes (very similar to the single variable case):

$$f(\vec{a}+\vec{h}) = \sum_{k=0}^{p} \frac{(D^{(k)}f)_{\vec{a}}(\vec{h})}{k!} + \frac{1}{(p+1)!}(D^{(p+1)}f)_{\vec{a}+\xi\vec{h}}(\vec{h}).$$

Explicit example for $p = 1$:

$$f(\vec{a}+\vec{h}) = f(\vec{a}) + \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\vec{a}) h_k + \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_k}(\vec{a}+\xi\vec{h}) h_j h_k.$$

Here is an alternative approach to Taylor's Theorem with less hypotheses. Theorem 6.2.2 requires convexity as we are looking for a path, but here we only need an open ball and continuous differentiability.

Definition 6.2.3 Let $U \subseteq \mathbb{R}^n$ be open (but not necessarily convex) and let $f \in C^p(U, \mathbb{R})$ for some integer $p \geq 0$. Suppose $\vec{a}, \vec{x} \in U$. We define the **Taylor polynomial** of order p for f at $\vec{a}$ as

$$P^f_{p,\vec{a}}(\vec{x}) := f(\vec{a}) + \sum_{k=1}^{p} \frac{1}{k!}\left[\{(\vec{x}-\vec{a})\cdot\nabla\}^k f\right](\vec{a}) = \sum_{k=0}^{p} \frac{1}{k!}(D^{(k)}f)_{\vec{a}}(\vec{x}-\vec{a})$$

and the **Taylor remainder** of order p for f at $\vec{a}$ as

$$R^f_{p,\vec{a}}(\vec{x}) := f(\vec{x}) - P^f_{p,\vec{a}}(\vec{x}).$$

Theorem 6.2.4 Let $U \subseteq \mathbb{R}^n$ be open and $f \in C^p(U, \mathbb{R})$ for some integer $p \geq 0$. For any $\vec{a} \in U$, the Taylor remainder of order p satisfies

$$\lim_{\vec{x}\to\vec{a}} \frac{R^f_{p,\vec{a}}(\vec{x})}{\|\vec{x}-\vec{a}\|^p} = 0.$$

We now do an example of first and second order Taylor polynomial.

Example 6.2.5 Compute the first and second order Taylor polynomial for the function $f(x,y) = xy + x^2 + y^2$ about the point $(1,1)$.

Solution. Here are the explicit formula for first and second order Taylor polynomial.

$$P^f_{1,\vec{a}} \approx f(\vec{a}) + \sum_{i=0}^{n} \frac{\partial f}{\partial x_i}(\vec{a})(\vec{x}_i - \vec{a}_i),$$
$$P^f_{2,\vec{a}} \approx f(\vec{a}) + \sum_{i=0}^{n} \frac{\partial f}{\partial x_i}(\vec{a})(\vec{x}_i - \vec{a}_i) + \frac{1}{2}\sum_{i=0}^{n}\sum_{j=0}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a})(\vec{x}_i - \vec{a}_i)(\vec{x}_j - \vec{a}_j).$$

First, $f(1,1) = 3$. Next, we compute all first and second order partial derivatives:

$$\left.\frac{\partial f}{\partial x}\right|_{(1,1)} = 3, \quad \left.\frac{\partial f}{\partial y}\right|_{(1,1)} = 3, \quad \left.\frac{\partial^2 f}{\partial x^2}\right|_{(1,1)} = 2, \quad \left.\frac{\partial^2 f}{\partial y^2}\right|_{(1,1)} = 2, \quad \frac{\partial^2 f}{\partial x \partial y} = \left.\frac{\partial^2 f}{\partial y \partial x}\right|_{(1,1)} = 1.$$

Then, the first order Taylor polynomial is

$$P^f_{1,(1,1)} \approx 3 + 3(x-1) + 3(y-1) = 3x + 3y - 3.$$

The second order Taylor polynomial is

$$\begin{aligned} P^f_{2,(1,1)} &\approx 3 + 3(x-1) + 3(y-1) + \frac{1}{2}\big(2(x-1)^2 + 2(y-1)^2 + (x-1)(y-1) + (x-1)(y-1)\big) \\ &= 3x + 3y - 3 + (x-1)^2 + (y-1)^2 + (x-1)(y-1). \end{aligned}$$

□

Equivalently, we could use gradient and Hessian to compute Taylor polynomials (see remark above Theorem 7.4.2 on quadratic forms):

$$P^f_{1,\vec{a}} = f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a}) + R^f_{1,\vec{a}},$$
$$P^f_{2,\vec{a}} = f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a}) + \frac{1}{2}(\vec{x} - \vec{a})^T Hf(\vec{a})(\vec{x} - \vec{a}) + R^f_{2,\vec{a}}.$$

7 Second Derivative Test

7.1 Single Variable Second Derivative Test

Recall the following from Math 147:

Theroem 7.1.1 If the function f is twice differentiable at $x = c$, then the graph of f is concave upward at $(c, f(c))$ if $f''(c) > 0$ and concave downward if $f''(c) < 0$.

Definition 7.1.2 Points on the graph of f where the concavity changes from up-to-down or down-to-up are called the **inflection points** of the graph.

Theorem 7.1.3 If $f'(c)$ exists and $f''(c)$ changes sign at $x = c$, then the point $(c, f(c))$ is an **inflection point** of the graph of f. If $f''(c)$ exists at the inflection point, then $f''(c) = 0$.

The Second Derivative Test relates the concepts of critical points, extreme values, and concavity to give a very useful tool for determining whether a critical point on the graph of a function is a local minimum or maximum.

Theorem 7.1.4 Suppose that c is a critical point at which $f'(c) = 0$, that $f'(x)$ exists in a neighborhood of c, and that $f''(c)$ exists. Then f has a local minimum at c if $f''(c) < 0$ and a local minimum at c if $f''(c) > 0$. If $f''(c) = 0$, the test is no informative.

7.2 Critical Points

We now define relevant terms in higher dimensions.

Definition 7.2.1 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$, $f : A \to \mathbb{R}$. We say that $\vec{a}$ is a:

1. **critical point** (or **stationary point**) of f if $\nabla f(\vec{a}) = \vec{0}$.
2. **local maximum** of f if there exists $\delta > 0$ such that $f(\vec{x}) \leq f(\vec{a})$ for all $\vec{x} \in B_\delta(\vec{a})$.
3. **local minimum** of f if there exists $\delta > 0$ such that $f(\vec{x}) \geq f(\vec{a})$ for all $\vec{x} \in B_\delta(\vec{a})$.
4. **saddle point** of f is $\vec{a}$ is a critical point of f and for any $\delta > 0$, there exists points $\vec{x}, \vec{y} \in B_\delta(\vec{a})$ such that $f(\vec{x}) < f(\vec{a}) < f(\vec{y})$.

If $\vec{a}$ is a local extremum and the gradient exists, then the gradient is zero at $\vec{a}$.

Theorem 7.2.2 Let $A \subseteq \mathbb{R}^n$, $\vec{a} \in \text{int}(A)$, $f : A \to \mathbb{R}$. If $\vec{a}$ is a local minimum or local maximum of f and $\nabla f(\vec{a})$ exists, then $\vec{a}$ is a critical point of f.

Proof. Since $\vec{a} \in \text{int}(A)$ and ∇f exists, we must have

$$\lim_{h\to 0} \frac{f(\vec{a} + h\vec{e}_j) - f(\vec{a})}{h} = \lim_{h\to 0^+} \frac{f(\vec{a} + h\vec{e}_j) - f(\vec{a})}{h} = \lim_{h\to 0^-} \frac{f(\vec{a} + h\vec{e}_j) - f(\vec{a})}{h}.$$

for each component $j \in \{1, 2, \dots, n\}$. Suppose $\vec{a}$ is a local maximum. Then there exists $\delta > 0$ such that $f(\vec{x}) \le f(\vec{a})$ for all $\vec{x} \in B_\delta(\vec{a})$. Fix $i \in \{1, 2, \dots, n\}$. For any h satisfying $0 < h < \delta$,

$$f(\vec{a} + h\vec{e}_i) - f(\vec{a}) \le 0 \implies \lim_{h \to 0^+} \frac{f(\vec{a} + h\vec{e}_i) - f(\vec{a})}{h} \le 0.$$

Similarly, for h satisfying $-\delta < h < 0$,

$$f(\vec{a} + h\vec{e}_i) - f(\vec{a}) \le 0 \implies \lim_{h \to 0^-} \frac{f(\vec{a} + h\vec{e}_i) - f(\vec{a})}{h} \ge 0.$$

Here, $\frac{\partial f}{\partial x_i}(\vec{a}) = 0$ for each $i \in \{1, 2, \dots, n\}$. A similar argument shows that $\nabla f(\vec{a}) = \vec{0}$ if $\vec{a}$ is a local minimum. □

7.3 Hessian Matrix and Quadratic Form

Just as we recorded all first-order partial derivatives in the Jacobian matrix, we here record all second-order partial derivatives in the Hessian matrix.

Definition 7.3.1 We define the **Hessian matrix** $H \in \mathbb{R}^{n \times n}$ of f at $\vec{a}$ by $H = [H_{ij}]$ where

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a}).$$

The Hessian matrix is also denoted by $D^2 f(\vec{a})$.

Recall from linear algebra, we could use associated quadratic forms to classify matrices:

Definition 7.3.2 A function $Q : \mathbb{R}^n \to \mathbb{R}$ is a **quadratic form** if there exists a real symmetric $n \times n$ matrix A such that $Q(\vec{u}) = \vec{u}^T A \vec{u}$ for all $\vec{u} \in \mathbb{R}^n$.

Let $Q : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form. Then Q is:

1. *positive definite* if $Q(\vec{u}) > 0$ for all $\vec{u} \neq \vec{0}$,
2. *negative definite* if $Q(\vec{u}) < 0$ for all $\vec{u} \neq \vec{0}$,
3. *positive semidefinite* if $Q(\vec{u}) \ge 0$ for all $\vec{u}$,
4. *negative semidefinite* if $Q(\vec{u}) \le 0$ for all $\vec{u}$,
5. *indefinite* if $Q(\vec{u}) > 0$ for some $\vec{u}$ and $Q(\vec{u}) < 0$ for some $\vec{u}$.

We can also write equivalent definitions applied to the matrix A. For example, A is *positive definite* if $\vec{u}^T A \vec{u} > 0$ for all $\vec{u} \neq \vec{0}$.

Here is a key result from linear algebra: the Spectral Theorem of Real Symmetric Matrices says symmetric real matrices have real eigenvalues and are diagonalizable.

Theorem 7.3.3 If $A \in \mathbb{R}^{n\times n}$ is symmetric, then the eigenvalues $\{\lambda_i : 1 \le i \le n\}$ are all real and there is an orthonormal matrix $P \in \mathbb{R}^{n\times n}$ such that $P^T AP = D$, where $D := \mathsf{diag}(\lambda_1, \ldots, \lambda_n)$.

Define $\vec{v} := P^T\vec{u}$, we have $Q(\vec{u}) = (P\vec{v})^T A(P\vec{v}) = \vec{v}^T D\vec{v} = \sum_{i=1}^n \lambda_i v_i^2$. Hence, we can classify quadratics based on the eigenvalues of A.

Proposition 7.3.4 Let $Q : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form with associated matrix $A \in \mathbb{R}^{n\times n}$. Then, Q is

1. *positive definite* if and only if the eigenvalues of A are all positive,
2. *negative definite* if and only if the eigenvalues of A are all negative,
3. *indefinite* if some of the eigenvalues of A are positive and some are negative,
4. *positive semidefinite* if and only if all eigenvalues of A are non-negative,
5. *negative semidefinite* if and only if all eigenvalues of A are non-positive.

7.4 Second Derivative Test

The following lemma is not essential but still helpful to understand why Second Derivative Test works; it gives a funny bound (before, we are generally bounding positive values from above and negative values from below, but here we are finding a lower bound for a positive Q and an upper bound for a negative Q) to the possible values of Q.

Lemma 7.4.1 Let $Q : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form.

1. If Q is positive definite, then there exists $M > 0$ such that $Q(\vec{u}) \ge M\|\vec{u}\|^2$ for all $\vec{u} \in \mathbb{R}^n$.
2. If Q is negative definite, then there exists $M > 0$ such that $Q(\vec{u}) \le -M\|\vec{u}\|^2$ for all $\vec{u} \in \mathbb{R}^n$.

Here is the intuition for second derivative test. Let $U \subseteq \mathbb{R}^n$ be an open set and $f \in C^2(U, \mathbb{R})$. Suppose $\vec{a} \in U$ is a critical point of f. By Alterative Taylor's Theorem,

$$\begin{aligned} f(\vec{a} + \vec{h}) &= f(\vec{a}) + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a})h_i h_j + R^f_{2,\vec{a}}(\vec{a} + \vec{h}) \\ &= f(\vec{a}) + \frac{1}{2}Q(\vec{h}) + R^f_{2,\vec{a}}(\vec{a} + \vec{h}), \end{aligned}$$

where the remainder goes to zero as $\vec{x} \to \vec{a}$ and $Q(\vec{h})$ is bounded from above/below. Thus, inside an open ball $B_\delta(\vec{a})$ with sufficiently small radius, $f(\vec{a} + \vec{h})$ is always greater/less than $f(\vec{a})$. Thus, we could use the quadratic form to determine if $\vec{a}$ is a local minimum, local maximum, or saddle.

Theorem 7.4.2 Let $U \subseteq \mathbb{R}^n$ be open and let $f \in C^2(U, \mathbb{R})$. Suppose $\vec{a} \in U$ is a critical point of f and let $Q : \mathbb{R}^n \to \mathbb{R}$ be the quadratic form associated with the Hessian matrix of f at $\vec{a}$. Then,

1. $\vec{a}$ is a local maximum of f if Q is negative definite,

2. $\vec{a}$ is a local minimum of f if Q is positive definite,
3. $\vec{a}$ is a saddle point of f if Q is indefinite.

To use the second derivative test to determine whether a critical point is a local maximum, local minimum, or saddle point, first compute the Hessian matrix at the given point, then find all eigenvalues and apply the theorem above.

Example 7.4.3 The function $f(x, y) := x^2 + y^2$ on the domain $A := \{(x, y) : x^2 + y^2 \leq 1\}$ has one critical point $(0, 0)$. The Hessian matrix at $(0, 0)$ is $H = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. The eigenvalues are 2 (repeated), all positive. By the second derivative test, $(0, 0)$ is a local minimum. □

Here is a quick recap on eigenvalues.

Remark 7.4.4 Given a matrix A, we want to find the values of λ which satisfy the **characteristic equation** of the matrix A, namely the values for which $\det(A - \lambda I) = 0$, where I is the appropriate identity matrix. First, form the matrix $A - \lambda I$, which is equal to A with λ subtracted from each entry on the main diagonal. Then, calculate $\det(A - \lambda I)$ and find all of its roots. Crucially, the Hessian matrix is symmetric (following from Theorem 5.2.1, mixed partial derivatives), so by Theorem 7.3.3 (Spectral Theorem for Real Symmetric Matrices), all of its eigenvalues are guaranteed to be positive. Thus, the roots to the characteristic equation are all real.

7.5 Summary: Optimization

Lagrange Multiplier Theorem is not tested, so all we have is the second derivative test, which can be used to find local extrema of a given function on an open domain. By Definition 7.2.1, critical points occur when the gradient is zero. To classify the critical point, we apply the second derivative test (see Example 7.4.3). Finally, we can use EVT to justify existence of global extrema when needed.