Chapter 6. Differentiation in Euclidean Space

6.1 Definition: Let $U \subseteq \mathbb{R}^n$ be open in $\mathbb{R}^n$, let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, and let at $a \in U$, say $a = (a_1, \cdots, a_n)$. We define the k^{th} **partial derivative** of f at a to be

$$\frac{\partial f}{\partial x_k}(a) = {g_k}'(a_k) \text{ , where } g_k(t) = f(a_1, \cdots, a_{k-1}, t, a_{k+1}, \cdots a_n) \,,$$

or equivalently, letting $e_k = (0, 0, \cdots, 0, 1, 0, \cdots, 0)$ be the k^{th} standard basis vector in $\mathbb{R}^n$,

$$\frac{\partial f}{\partial x_k}(a) = {h_k}'(0) \text{ , where } h_k(t) = f(a_1, \cdots, a_{k-1}, a_k + t, a_{k+1}, \cdots a_n) = f(a + t\, e_k) \,,$$

provided that the derivatives exist. Note that g_k and h_k are functions of a single variable.

Sometimes $\frac{\partial f}{\partial x_k}$ is written as f_{x_k} or as f_k. When we write $u = f(x)$, we can also write $\frac{\partial f}{\partial x_k}$ as $\frac{\partial u}{\partial x_k}$, u_{x_k} or u_k. When $n = 3$ and we write x, y and z instead of x_1, x_2 and x_3, the partial derivatives $\frac{\partial f}{\partial x_1}$, $\frac{\partial f}{\partial x_2}$ and $\frac{\partial f}{\partial x_3}$ are written as $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ and $\frac{\partial f}{\partial z}$, or as f_x, f_y and f_z. When $n = 1$ so there is only one variable $x = x_1$ we have $\frac{\partial f}{\partial x}(a) = \frac{df}{dx}(a) = f'(a)$.

6.2 Definition: Let $U \subseteq \mathbb{R}^n$ be open in $\mathbb{R}^n$, let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and let $a \in U$. Write $u = f(x) = \big(f_1(x), f_2(x), \cdots, f_m(x)\big)^T$ with $x = (x_1, x_2, \cdots, x_n)^T$. We define the **derivative matrix**, or the **Jacobian matrix**, of f at a to be the matrix

$$Df(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}$$

and we define the **linearization** of f at a to be the affine map $L : \mathbb{R}^n \to \mathbb{R}^m$ given by

$$L(x) = f(a) + Df(a)(x - a)$$

provided that all the partial derivatives $\frac{\partial f_k}{\partial x_l}(a)$ exist.

6.3 Definition: Let U be open in $\mathbb{R}^n$ and let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$. We say that f is $\mathcal{C}^1$ in U when all the partial derivatives $\frac{\partial f_k}{\partial f_l}$ exist and are continuous in U. The **second order partial derivatives** of f are the functions

$$\frac{\partial^2 f_j}{\partial x_k \partial x_l} = \frac{\partial\left(\frac{\partial f_j}{\partial x_l}\right)}{\partial x_k}.$$

We also write $\frac{\partial^2 f_j}{\partial {x_k}^2} = \frac{\partial^2 f_j}{\partial x_k \partial x_k}$. We say that f is $\mathcal{C}^2$ when all the partial derivatives $\frac{\partial^2 f_j}{\partial x_k \partial x_l}$ exist and are continuous in U. Higher order derivatives can be defined similarly, and we say f is $\mathcal{C}^k$ when all the k^{th} order derivatives $\frac{\partial^k f_j}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}}$ exist and are continuous in U.

6.4 Definition: Let $a \in U$ where U is an open set in $\mathbb{R}$, and let $f : U \subseteq \mathbb{R} \to \mathbb{R}^m$, say $x = f(t) = \big(x_1(t), x_2(t), \cdots, x_m(t)\big)$. Then we write $f'(a) = Df(a)$ and we have

$$f'(a) = Df(a) = \begin{pmatrix} \frac{\partial x_1}{\partial t}(a) \\ \vdots \\ \frac{\partial x_m}{\partial t}(a) \end{pmatrix} = \begin{pmatrix} {x_1}'(a) \\ \vdots \\ {x_m}'(a) \end{pmatrix}.$$

The vector $f'(a)$ is called the **tangent vector** to the curve $x = f(t)$ at the point $f(a)$. In the case that t represents time and $f(t)$ represents the position of a moving point, $f'(a)$ is also called the **velocity** of the moving point at time $t = a$.

6.5 Definition: Let $a \in U$ where U is an open set in $\mathbb{R}^n$ and let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$. We define the **gradient** of f at a to be the vector

$$\nabla f(a) = Df(a)^T = \Big(\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a)\Big)^T = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{pmatrix}.$$

6.6 Note: Recall that for $f : U \subseteq \mathbb{R} \to \mathbb{R}$ and $a \in U$,

$$f \text{ is differentiable at } a \iff \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \text{ exists}$$

$$\iff \exists m \in \mathbb{R}\ \forall \epsilon > 0\ \exists \delta > 0\ \forall x \in U\ \ 0 < |x - a| < \delta \implies \Big|\frac{f(x) - f(a)}{x - a} - m\Big| < \epsilon$$

$$\iff \exists m \in \mathbb{R}\ \forall \epsilon > 0\ \exists \delta > 0\ \forall x \in U\ \ 0 < |x - a| < \delta \implies \big|f(x) - f(a) - m(x - a)\big| < \epsilon\,|x - a|$$

$$\iff \exists m \in \mathbb{R}\ \forall \epsilon > 0\ \exists \delta > 0\ \forall x \in U\ \ |x - a| \leq \delta \implies \big|f(x) - \big(f(a) + m(x - a)\big)\big| \leq \epsilon\,|x - a|.$$

In this case, the number $m \in \mathbb{R}$ is unique, we call it the **derivative** of f at a and denote it by $f'(a)$, and the map $\ell(x) = f(a) + f'(a)(x - a)$ is called the **linearization** of f at a.

6.7 Definition: Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, where U is open. We say f is **differentiable** at $a \in U$ if there is an $m \times n$ matrix A such that

$$\forall \epsilon > 0\ \exists \delta > 0\ \forall x \in U\ \Big(|x - a| \leq \delta \implies \big|f(x) - (f(a) + A(x - a))\big| \leq \epsilon|x - a|\Big).$$

We show below that the matrix A is unique, we call it the **derivative** (matrix) of f at a, and we denote it by $Df(a)$. The affine map $L : \mathbb{R}^n \to \mathbb{R}^m$ given by $L(x) = f(a) + Df(a)(x - a)$, which approximates $f(x)$, is called the **linearization** of f at a. We say f is **differentiable** in U when it is differentiable at every point $a \in U$.

6.8 Example: If f is the affine map $f(x) = Ax + b$, then we have $Df(a) = A$ for all a. Indeed given $\epsilon > 0$ we can choose $\delta > 0$ to be anything we like, and then for all x we have

$$\big|f(x) - f(a) - A(x - a)\big| = \big|Ax + b - Aa - b - Ax + Aa\big| = 0 \leq \epsilon|x - a|.$$

6.9 Theorem: *(The Derivative is the Jacobian) Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and let $a \in U$. If f is differentiable at a then the partial derivatives $\frac{\partial f_k}{\partial x_\ell}(a)$ all exist and the matrix A which appears in the definition of the derivative is equal to the Jacobian matrix $Df(a)$.*

Proof: Suppose that f is differentiable at a. Fix indices k and ℓ and let $g(t) = f_k(a + te_\ell)$ so that $\frac{\partial f_k}{\partial x_\ell}(a) = g'(0)$ provided that the derivative $g'(0)$ exists. Let A be a matrix as in the definition of differentiability. Let $\epsilon > 0$. Choose $\delta > 0$ such that for all $x \in U$ with $|x - a| \leq \delta$ we have $\big|f(x) - f(a) - A(x - a)\big| \leq \epsilon\,|x - a|$. Let $t \in \mathbb{R}$ with $|t| \leq \delta$. Let $x = a + t\,e_\ell$. Then we have $|x - a| = |te_\ell| = |t| \leq \delta$ and so $\big|f(x) - f(a) - A(x - a)\big| \leq \epsilon\,|x - a|$. Since for any vector $u \in \mathbb{R}^m$ we have $|u_k| \leq |u|$, we have

$$\begin{aligned} \big|g(t) - g(0) - A_{k,\ell}\,t\big| &= \big|f_k(a + te_\ell) - f_k(a) - \big(A(te_\ell)\big)_k\big| \\ &\leq \big|f(a + te_\ell) - f(a) - A(te_\ell)\big| \\ &= \big|f(x) - f(a) - A(x - a)\big| \\ &\leq \epsilon\,|x - a| = \epsilon\,|t|. \end{aligned}$$

It follows that $A_{k,\ell} = g'(0) = \frac{\partial f_k}{\partial x_\ell}(a)$, as required.

The Matrix Norm

6.10 Definition: Let $A \in M_{m\times n}(\mathbb{R})$ and let $S = \left\{x \in \mathbb{R}^n \,\middle|\, |x| = 1\right\}$. Since S is compact, by the Extreme Value Theorem, the continuous function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = \left|Ax\right|$ attains its maximum value on S. We define the **norm** of the matrix A to be

$$\|A\| = \max\left\{|Ax| \,\middle|\, |x| = 1\right\}.$$

6.11 Lemma: *(Properties of the Matrix Norm) Let $A \in M_{m\times n}(\mathbb{R})$. Then*
(1) $|Ax| \le \|A\|\,|x|$ for all $x \in \mathbb{R}^n$,
(2) if A is invertible then $|Ax| \ge \frac{|x|}{\|A^{-1}\|}$ for all $x \in \mathbb{R}^n$,
(3) $\|A\| \le \sum_{k=1}^{m}\sum_{\ell=1}^{n} |A_{k,\ell}|$, and
(4) $\|A\|$ is equal to the square root of the largest eigenvalue of the matrix A^TA.

Proof: When $x = 0 \in \mathbb{R}^n$ we have $|Ax| = 0 = \|A\|\,|x|$ and when $0 \ne x \in \mathbb{R}^n$ we have

$$|Ax| = \left|\,|x|\,A\tfrac{x}{|x|}\right| = |x|\left|A\tfrac{x}{|x|}\right| \le |x|\,\|A\|.$$

This proves Part 1. To prove Part 2, suppose that A is invertible. Then we can choose $x \in \mathbb{R}^n$ with $|x| = 1$ such that $Ax \ne 0$ so we must have $\|A\| > 0$. Similarly, since A^{-1} is also invertible, we also have $\|A^{-1}\| > 0$. By Part 1, for all $x \in \mathbb{R}^n$ we have $|x| = \left|A^{-1}A\,x\right| \le \|A^{-1}\|\,|Ax|$ so that $|Ax| \ge \frac{|x|}{\|A^{-1}\|}$, as required. To prove Part 3, let $x \in \mathbb{R}^n$ with $|x| = 1$. Then $|x_\ell| \le |x| \le 1$ for all indices ℓ, and so

$$\left|Ax\right| = \left|\sum_{k=1}^{m}(Ax)_k e_k\right| \le \sum_{k=1}^{m}\left|(Ax)_k\right| = \sum_{k=1}^{m}\left|\sum_{\ell=1}^{n} A_{k,\ell}x_\ell\right| \le \sum_{k=1}^{m}\sum_{\ell=1}^{n}|A_{k,\ell}|\,|x_\ell| \le \sum_{k=1}^{m}\sum_{\ell=1}^{n}|A_{k,\ell}|.$$

We omit the proof of Part 4, which we shall not use (it is often proven in a linear algebra course).

6.12 Theorem: *(Differentiability Implies Continuity) Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$. If f is differentiable at $a \in U$, then f is continuous at a.*

Proof: Suppose f is differentiable at a. Note that for all $x \in U$ we have

$$\begin{aligned}|f(x) - f(a)| &= \left|f(x) - f(a) - Df(a)(x-a) + Df(a)(x-a)\right| \\ &\le \left|f(x) - f(a) - Df(a)(x-a)\right| + \left|Df(a)\,(x-a)\right| \\ &\le \left|f(x) - f(a) - Df(a)(x-a)\right| + \|Df(a)\|\,|x-a|\end{aligned}$$

Let $\epsilon > 0$. Since f is differentiable at a we can choose δ with $0 < \delta < \frac{\epsilon}{1+\|Df(a)\|}$ such that

$$|x-a| \le \delta \Longrightarrow \left|f(x) - f(a) - Df(a)(x-a)\right| \le |x-a|$$

and then for $|x-a| \le \delta$ we have

$$\begin{aligned}\left|f(x) - f(a)\right| &\le \left|f(x) - f(a) - Df(a)(x-a)\right| + \|Df(a)\|\,|x-a| \\ &\le |x-a| + \|Df(a)\|\,|x-a| = \big(1 + \|Df(a)\|\big)\,|x-a| \\ &\le \big(1 + \|Df(a)\|\big)\,\delta < \epsilon.\end{aligned}$$

6.13 Theorem: *(Continuous Partial Derivatives Implies Differentiability) Let $U \subseteq \mathbb{R}^n$ be open, let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and let $a \in U$. If the partial derivatives $\frac{\partial f_k}{\partial x_\ell}(x)$ exist in U and are continuous at a then f is differentiable at a.*

Proof: Suppose that the partial derivatives $\frac{\partial f_k}{\partial x_\ell}(x)$ exist in U and are continuous at a. Let $\epsilon > 0$. Choose $\delta > 0$ so that $\overline{B}(a,\delta) \subseteq U$ and so that for all indices k, ℓ and for all $y \in U$ we have $|y-a| \leq \delta \implies \left|\frac{\partial f_k}{\partial x_\ell}(y) - \frac{\partial f_k}{\partial x_\ell}(a)\right| \leq \frac{\epsilon}{nm}$. Let $x \in U$ with $|x-a| \leq \delta$. For $0 \leq \ell \leq n$, let $u_\ell = (x_1, \cdots, x_\ell, a_{\ell+1}, \cdots, a_n)$, with $u_0 = a$ and $u_n = x$, and note that each $u_\ell \in \overline{B}(a,\delta)$. For $1 \leq \ell \leq n$, let $\alpha_\ell(t) = (x_1, \cdots, x_{\ell-1}, t, a_{\ell+1}, \cdots, a_n)$ for t between a_ℓ and x_ℓ, For $1 \leq k \leq m$ and $1 \leq \ell \leq n$, let $g_{k,\ell}(t) = f_k\big(\alpha_\ell(t)\big)$ so that $g_{k,\ell}'(t) = \frac{\partial f_k}{\partial x_\ell}\big(\alpha_\ell(t)\big)$. By the Mean Value Theorem, we can choose $s_{k,\ell}$ between a_ℓ and x_ℓ so that $g_{k,\ell}'(s_{k,\ell})(x_\ell - a_\ell) = g_{k,\ell}(x_\ell) - g_{k,\ell}(a_\ell)$ or, equivalently, so that $\frac{\partial f_k}{\partial x_\ell}\big(\alpha_\ell(s_{k,\ell})\big)(x_\ell - a_\ell) = f_k(u_\ell) - f_k(u_{\ell-1})$. Then

$$f_k(x) - f_k(a) = f_k(u_n) - f_k(u_0) = \sum_{\ell=1}^{n} \big(f_k(u_\ell) - f_k(u_{\ell-1})\big) = \sum_{\ell=1}^{n} \frac{\partial f_k}{\partial x_\ell}\big(\alpha_\ell(s_{k,\ell})\big)(x_\ell - a_\ell).$$

Let $B \in M_{m\times n}(\mathbb{R})$ be the matrix with entries $B_{k,\ell} = \frac{\partial f}{\partial x_\ell}\big(\alpha_\ell(s_{k,\ell})\big)$. Then (by Parts 1 and 3 of Lemma 6.11) we have

$$\begin{aligned}\Big|f(x) - f(a) - Df(a)(x-a)\Big| &= \Big|\big(B - Df(a)\big)(x-a)\Big| \leq \big\|B - Df(a)\big\|\,|x-a| \\ &\leq \sum_{k,\ell} \left|\tfrac{\partial f_k}{\partial x_\ell}(\alpha_\ell(s_{k,\ell})) - \tfrac{\partial f_k}{\partial x_\ell}(a)\right| |x-a| \leq \epsilon|x-a|.\end{aligned}$$

6.14 Corollary: *If $U \subseteq \mathbb{R}^n$ is open and $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is $\mathcal{C}^1$ then f is differentiable.*

6.15 Corollary: *Every function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, which can be obtained by applying the standard operations (such as multiplication and composition) on functions to basic elementary functions defined on open domains, is differentiable in U.*

6.16 Exercise: For each of the following functions $f : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$, extend the domain of $f(x,y)$ to all of $\mathbb{R}^2$ by defining $f(0,0) = 0$ and then determine whether the partial derivatives of f exist at $(0,0)$ and whether f is differential at $(0,0)$.

(a) $f(x,y) = \frac{xy}{x^2+y^2}$ (b) $f(x,y) = |xy|$ (c) $f(x,y) = \sqrt{|xy|}$

(d) $f(x,y) = \frac{x^3}{x^2+y^2}$ (e) $f(x,y) = \frac{x}{(x^2+y^2)^{1/3}}$ (f) $f(x,y) = \frac{x^3-3xy^2}{x^2+y^2}$

The Chain Rule and the Directional Derivative

6.17 Theorem: *(The Chain Rule) Let $f : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$, let $g : V \subseteq \mathbb{R}^m \to \mathbb{R}^\ell$, and let $h(x) = g(f(x))$. If f is differentiable at a and g is differentiable at $f(a)$ then h is differentiable at a and $Dh(a) = Dg(f(a))Df(a)$.*

Proof: Suppose f is differentiable at a and g is differentiable at $f(a)$. Write $y = f(x)$ and $b = f(a)$. We have

$$\begin{aligned}
\big|h(x) - h(a) - Dg(f(a))Df(a)(x-a)\big| &= \big|g(y) - g(b) - Dg(b)Df(a)(x-a)\big| \\
&= \big|g(y) - g(b) - Dg(b)(y-b) + Dg(b)(y-b) - Dg(b)Df(a)(x-a)\big| \\
&\leq \big|g(y) - g(b) - Dg(b)(y-b)\big| + \|Dg(b)\|\,\big|y - b - Df(a)(x-a)\big| \\
&\leq \big|g(y) - g(b) - Dg(b)(y-b)\big| + \big(1 + \|Dg(b)\|\big)\big|f(x) - f(a) - Df(a)(x-a)\big|
\end{aligned}$$

and

$$\begin{aligned}
|y - b| &= |f(x) - f(a)| \\
&= \big|f(x) - f(a) - Df(a)(x-a) + Df(a)(x-a)\big| \\
&\leq \big|f(x) - f(a) - Df(a)(x-a)\big| + \|Df(a)\|\,|x-a|\,.
\end{aligned}$$

Let $\epsilon > 0$ be given. Since g is differentiable at b we can choose $\delta_0 > 0$ so that

$$|y - b| \leq \delta_0 \Longrightarrow \big|g(y) - g(b) - Dg(b)(y-b)\big| \leq \tfrac{\epsilon}{2(1+\|Df(a)\|)}\,|y-b|\,.$$

Since f is continuous at a we can choose $\delta_1 > 0$ so that

$$|x - a| \leq \delta_1 \Longrightarrow |y - b| = |f(x) - f(a)| \leq \delta_0$$

Since f is differentiable at a we can choose $\delta_2 > 0$ so that

$$|x - a| \leq \delta_2 \Longrightarrow \big|f(x) - f(a) - Df(a)(x-a)\big| \leq |x-a|$$

and we can choose $\delta_3 > 0$ so that

$$|x - a| \leq \delta_3 \Longrightarrow \big|f(x) - f(a) - Df(a)(x-a)\big| \leq \tfrac{\epsilon}{2(1+\|Dg(a)\|)}\,|x-a|\,.$$

Let $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Then for $|x - a| \leq \delta$ we have

$$\begin{aligned}
|y - b| &\leq \big|f(x) - f(a) - Df(a)(x-a)\big| + \big|Df(a)(x-a)\big| \\
&\leq |x-a| + \|Df(a)\|\,|x-a| \\
&= (1 + \|Df(a)\|)\,|x-a|
\end{aligned}$$

so

$$\big|g(y) - g(b) - Dg(b)(y-b)\big| \leq \tfrac{\epsilon}{2(1+\|Df(a)\|)}\,|y-b| \leq \tfrac{\epsilon}{2}\,|x-a|$$

and we have

$$\big(1 + \|Dg(b)\|\big)\big|f(x) - f(a) - Df(a)(x-a)\big| \leq \tfrac{\epsilon}{2}\,|x-a|$$

and so

$$\big|h(x) - h(a) - Dg(f(a))Df(a)(x-a)\big| \leq \tfrac{\epsilon}{2}\,|x-a| + \tfrac{\epsilon}{2}\,|x-a| = \epsilon|x-a|.$$

Thus h is differentiable at a with derivative $Dh(a) = Dg(f(a))Df(a)$, as required.

6.18 Definition: Let $f : U \subseteq \mathbb{R}^n \to R$, let $a \in \mathbb{R}^n$ and let $v \in \mathbb{R}^n$. We define the **directional derivative of f at a with respect to v**, written as $D_v f(a)$, as follows: pick any differentiable function $\alpha : (-\epsilon, \epsilon) \subseteq \mathbb{R} \to U \subseteq \mathbb{R}^n$, where $\epsilon > 0$, such that $\alpha(0) = a$ and $\alpha'(0) = v$ (for example, we could pick $\alpha(t) = a + v\,t$), let $g(t) = f(\alpha(t))$, note that by the Chain Rule we have $g'(t) = Df(\alpha(t))\alpha'(t)$, and then define

$$D_v f(a) = g'(0) = Df(\alpha(0))\,\alpha'(0) = Df(a)\,v = \nabla f(a) \cdot v\,.$$

Notice that the formula for $D_v f(a)$ does not depend on the choice of the function $\alpha(t)$. The **directional derivative of f at a in the direction of** v is defined to be $D_w f(a)$ where w is the unit vector in the direction of v, that is $w = \frac{v}{|v|}$.

6.19 Remark: Some books only define the directional derivative in the case that vector is a unit vector.

6.20 Theorem: *Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be differentiable at $a \in U$. Say $f(a) = b$. The gradient $\nabla f(a)$ is perpendicular to the level set $f(x) = b$, it is in the direction in which f increases most rapidly, and its length is the rate of increase of f in that direction.*

Proof: Let $\alpha(t)$ be any curve in the level set $f(x) = b$, with $\alpha(0) = a$. We wish to show that $\nabla f(a) \perp \alpha'(0)$. Since $\alpha(t)$ lies in the level set $f(x) = b$, we have $f(\alpha(t)) = b$ for all t. Take the derivative of both sides to get $Df(\alpha(t))\alpha'(t) = 0$. Put in $t = 0$ to get $Df(a)\,\alpha'(0) = 0$, that is $\nabla f(a) \cdot \alpha'(0) = 0$. Thus $\nabla f(a)$ is perpendicular to the level set $f(x) = b$.

Next, let u be a unit vector. Then $D_u f(a) = \nabla f(a) \cdot u = |\nabla f(a)| \cos\theta$, where θ is the angle between u and $\nabla f(a)$. So the maximum possible value of $D_u f(a)$ is $|\nabla f(a)|$, and this occurs when $\cos\theta = 1$, that is when $\theta = 0$, which happens when u is in the direction of $\nabla f(a)$.

The Geometry of the Linearization

6.21 Note: There are several geometric objects (curves and surfaces, and higher dimensional analogues) that we can associate with a given function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$. The **graph** of f is the set $\text{Graph}(f) = \big\{(x, f(x)) \,\big|\, x \in U\big\} \subseteq \mathbb{R}^{n+m}$. We say that the graph of f is given **explicitlty** by the equation $y = f(x)$. The **null set** of f is the set $\text{Null}(f) = f^{-1}(0) = \big\{x \in U \,\big|\, f(x) = 0\big\} \subseteq \mathbb{R}^n$, and more generally, when $a \in U$ and $f(a) = b$, the **inverse image** of b under f, also called the **level set** $f^{-1}(b)$, is given by $f^{-1}(b) = \big\{x \in U \,\big|\, f(x) = b\big\} \subseteq \mathbb{R}^n$. We say the level set $f^{-1}(b)$ is given **implicitly** by the equation $f(x) = b$. The **range** of f is the set $\text{Range}(f) = \big\{f(t) \,\big|\, t \in U\big\} \subseteq \mathbb{R}^m$. We say that the range of f is given **parametrically** by the equation $x = f(t)$.

When f is differentiable at $a \in U$, it is approximated by its linearization near $x = a$, that is when $x \cong a$ we have

$$f(x) \cong L(x) = f(a) + Df(a)(x - a)\,.$$

The geometric objects $\text{Graph}(f)$, $\text{Null}(f)$, $f^{-1}(b)$ and $\text{Range}(f)$ are approximated by the affine spaces $\text{Graph}(L)$, $\text{Null}(L)$, $L^{-1}(b)$ and $\text{Range}(L)$. Each of these affine spaces is called the (affine) **tangent space** of its corresponding geometric object: the space $\text{Graph}(L)$ is called the (affine) tangent space of the set $\text{Graph}(f)$ at the point $\big(a, f(a)\big)$; when $f(a) = b$ the space $L^{-1}(b)$ is called the (affine) tangent space to $f^{-1}(b)$ at the point a; and the space $\text{Range}(L)$ is called the (affine) tangent space of the set $\text{Range}(f)$ at the point $f(a)$. When a tangent space is 1-dimensional we call it a **tangent line** and when a tangent space is 2-dimensional we call it a **tangent plane**.

The Mean Value Theorem

6.22 Definition: For $a, b \in \mathbb{R}^n$, we define the **line segment** from a to b to be the set

$$[a,b] = \big\{a + t(b-a)\big|0 \leq t \leq 1\big\}.$$

For $A \subseteq \mathbb{R}^n$ we say the A is **convex** when for all $a, b \in A$ we have $[a,b] \subseteq A$.

6.23 Exercise: Show, using the triangle inequality, that $B(a,r)$ is convex for all $a \in \mathbb{R}^n$ and $r > 0$.

6.24 Theorem: *(The Mean Value Theorem) Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with U open in $\mathbb{R}^n$. Suppose that f is differentiable in U. Let $u \in \mathbb{R}^m$ and let $a, b \in U$ with $[a,b] \subseteq U$. Then there exists $c \in [a,b]$ such that*

$$Df(c)(b-a) \cdot u = \big(f(b) - f(a)\big) \cdot u.$$

Proof: Let $\alpha(t) = a + t(b-a)$ and define $g : [0,1] \to \mathbb{R}$ by $g(t) = f\big(\alpha(t)\big) \cdot u$. By the Chain Rule, we have $g'(t) = \big(Df(\alpha(t))\alpha'(t)\big) \cdot u = \big(Df(\alpha(t))(b-a)\big) \cdot u$. By the Mean Value Theorem (for a real-valued function of a single variable) we can choose $s \in [0,1]$ such that $g'(s) = g(1) - g(0)$, that is $\big(Df(\alpha(s))(b-a)\big) \cdot u = f(b) \cdot u - f(a) \cdot u = \big(f(b) - f(a)\big) \cdot u$. Thus we can take $c = \alpha(s) \in [a,b]$ to get $Df(c)(b-a) \cdot u = \big(f(b) - f(a)\big) \cdot u$.

6.25 Corollary: *(Vanishing Derivative) Let $U \subseteq \mathbb{R}^n$ be open and connected and let $f : U \to \mathbb{R}^m$ be differentiable with $Df(x) = O$ for all $x \in U$. Then f is constant in U.*

Proof: Let $a \in U$ and let $A = \big\{x \in U\big|f(x) = f(a)\big\}$. We claim that A is open (both in $\mathbb{R}^n$ and in U). Let $b \in A$, that is let $b \in U$ with $f(b) = f(a)$. Since U is open we can choose $r > 0$ so that $B(b,r) \subseteq U$. Let $c \in B(b,r)$. Since $B(b,r)$ is convex we have $[b,c] \subseteq B(b,r) \subseteq U$. Let $u = f(c) - f(b)$ and choose $d \in [b,c]$, as in the Mean Value Theorem, so that $\big(Df(d)(c-b)\big) \cdot u = \big(f(c) - f(b)\big) \cdot u$. Then we have

$$\big|f(c) - f(b)\big|^2 = \big(f(c) - f(b)\big) \cdot u = \big(Df(d)(c-b)\big) \cdot u = 0$$

since $Df(d) = O$. Since $\big|f(c) - f(b)\big| = 0$ we have $f(c) = f(b) = f(a)$, and so $c \in A$. Thus $B(b,r) \subseteq A$ and so A is open, as claimed. A similar argument shows that if $b \in U \setminus A$ and we chose $r > 0$ so that $B(b,r) \subseteq U$ then we have $f(c) = f(b)$ for all $c \in B(b,r)$ hence $B(b,r) \subseteq U \setminus A$ and hence $U \setminus A$ is also open. Note that A is non-empty since $a \in A$. If $U \setminus A$ was also non-empty then U would be the union of the two non-empty open sets A and $U \setminus A$, and this is not possible since U is connected. Thus $U \setminus A = \emptyset$ so $U = A$. Since $U = A = \big\{x \in U\big|f(x) = f(a)\big\}$ we have $f(x) = f(a)$ for all $x \in U$, so f is constant in U.

The Inverse and the Implicit Function Theorems

6.26 Theorem: *(The Inverse Function Theorem) Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ where $U \subseteq \mathbb{R}^n$ is open with $a \in U$. Suppose that f is $\mathcal{C}^1$ in U and that $Df(a)$ is invertible. Then there exists an open set $U_0 \subseteq U$ with $a \in U_0$ such that the set $V_0 = f(U_0)$ is open in $\mathbb{R}^n$ and the restriction $f : U_0 \to V_0$ is bijective, and its inverse $g = f^{-1} : V_0 \to U_0$ is $\mathcal{C}^1$ in V_0. In this case we have $Dg\big(f(a)\big) = Df(a)^{-1}$.*

Proof: Let $A = Df(a)$ and note that A is invertible. Since U is open and f is $\mathcal{C}^1$, we can choose $r > 0$ so that $B(a,r) \subseteq U$ and so that $\left|\frac{\partial f_k}{\partial x_\ell}(x) - \frac{\partial f_k}{\partial f_\ell}(a)\right| \leq \frac{1}{2n^2\|A^{-1}\|}$ for all k, ℓ. Let $U_0 = B(a,r)$ and note that for all $x \in U_0$ we have $\big\|Df(x) - A\big\| \leq \frac{1}{2\|A^{-1}\|}$.

Claim 1: for all $x \in U_0$, the matrix $Df(x)$ is invertible.
Let $x \in U_0$ and suppose, for a contradiction, that $Df(x)$ is not invertible. Then we can choose $u \in \mathbb{R}^n$ with $|u| = 1$ such that $Df(x)u = 0$. But then we have

$$\big\|Df(x) - A\big\| \geq \big|(Df(a) - A)u\big| = \big|Au\big| \geq \tfrac{|u|}{\|A^{-1}\|} = \tfrac{1}{\|A^{-1}\|}$$

which contradicts the fact that since $x \in U_0$ we have $\big\|Df(x) - A\big\| \leq \frac{1}{2\|A^{-1}\|}$.

Claim 2: for all $b, c \in U_0$ we have $\big|f(c) - f(b) - A(c-b)\big| \leq \frac{\|c-b|}{2\|A^{-1}\|}$.
Let $b, c \in U_0$. Let $\alpha(t) = b + t(c-b)$ and note that $\alpha(t) \in U_0$ for all $t \in [0,1]$. Let $\phi(t) = f\big(\alpha(t)\big) - L\big(\alpha(t)\big)$ where L is the linearization of f at a given by $L(a) = f(a) + Df(a)(x-a)$, and note that $\phi(1) - \phi(0) = \big(f(c) - L(c)\big) - \big(f(b) - L(b)\big) = f(c) - f(b) - A(c-b)$. By the Chain Rule, we have $\phi'(t) = Df\big(\alpha(t)\big)\alpha'(t) - DL\big(\alpha(t)\big)\alpha'(t) = \Big(Df\big(\alpha(t)\big) - A\Big)(c-b)$ and so

$$\big|\phi'(t)\big| \leq \Big\|Df\big(\alpha(t)\big) - A\Big\|\,|c-b| \leq \tfrac{|c-b|}{2\|A^{-1}\|}.$$

By the Mean Value Theorem, using $u = \phi(1) - \phi(0)$, we choose $t \in [0,1]$ such that

$$\begin{aligned}\big|\phi(1) - \phi(0)\big|^2 &= (\phi(1) - \phi(0)) \cdot u = (D\phi(t)(1-0)) \cdot u = \phi'(t) \cdot u \\ &= \big|\phi'(t) \cdot (\phi(1) - \phi(0))\big| \leq \big|\phi'(t)\big|\,\big|\phi(1) - \phi(0)\big|\end{aligned}$$

by the Cauchy Schwarz Inequality, and hence $|\phi(1) - \phi(0)| \leq |\phi'(t)| \leq \frac{|c-b|}{2\|A^{-1}\|}$, that is

$$\big|f(c) - f(b) - A(c-b)\big| \leq \tfrac{|c-b|}{2\|A^{-1}\|}.$$

Claim 3: for all $b, c \in U_0$ we have $\big|f(c) - f(b)\big| \geq \frac{|c-b|}{2\|A^{-1}\|}$.
Let $b, c \in U_0$. By the Triangle Inequality we have

$$\big|f(c) - f(b) - A(c-b)\big| \geq \big|A(c-b)\big| - \big|f(c) - f(b)\big| \geq \tfrac{|c-b|}{\|A^{-1}\|} - \big|f(c) - f(b)\big|$$

and so, by Claim 3, we have

$$\big|f(c) - f(b)\big| \geq \tfrac{|c-b|}{\|A^{-1}\|} - \big|f(c) - f(b) - A(c-b)\big| \geq \tfrac{|c-b|}{\|A^{-1}\|} - \tfrac{|c-b|}{2\|A^{-1}\|} = \tfrac{|c-b|}{2\|A^{-1}\|}.$$

It follows that when $b \neq c$ we have $f(b) \neq f(c)$, so the restriction of f to U_0 is injective.

Claim 4: the restriction of f to U_0 is injective, hence $f : U_0 \to V_0 = f(U_0)$ is bijective.
By Claim 3, when $b, c \in U_0$ with $b \neq c$ we have $\big|f(c) - f(b)\big| \geq \frac{|c-b|}{2\|A^{-1}\|} > 0$ so that $f(b) \neq f(c)$. Thus the restriction of f to U_0 is injective, as claimed.

Claim 5: the set V_0 is open in $\mathbb{R}^n$.
Let $p \in V_0$. Let $b = g(p)$ so that $p = f(b)$. Choose $s > 0$ so that $\overline{B}(b,s) \subseteq U_0$. We shall show that $B\big(p, \frac{s}{4\|A^{-1}\|}\big) \subseteq V_0$. Let $q \in B\big(b, \frac{s}{4\|A^{-1}\|}\big)$. We need to show that $q \in V_0 = f(U_0)$ and in fact we shall show that $q \in f\big(B(b,s)\big)$. To do this, define $\psi : U \to \mathbb{R}$ by $\psi(x) = \big|f(x) - q\big|$. Since ψ is continuous, it attains its minimum value on the compact set $\overline{B}(b,s)$, say at $c \in \overline{B}(b,s)$. We shall show that $c \in B(b,s)$ and that $f(c) = q$ so we have $q \in f\big(B(b,s)\big)$, hence $q \in f(U_0) = V_0$, hence $B\big(b, \frac{s}{4\|A^{-1}\|}\big) \subseteq V_0$, and hence V_0 is open.

Claim 5(a): we have $c \in B(b,s)$.
Suppose, for a contradiction, that $c \notin B(b,s)$ so we have $|c - b| = s$. Then

$$\begin{aligned} \psi(b) &= \big|f(b) - q\big| = |p - q| < \tfrac{s}{4\|A^{-1}\|} \text{ and, using Claim 3,} \\ \psi(c) &= \big|f(c) - q\big| \geq \big|f(c) - f(b)\big| - \big|f(b) - q\big| \geq \tfrac{|c-b|}{2\|A^{-1}\|} - |p - q| \\ &= \tfrac{s}{2\|A^{-1}\|} - |p - q| > \tfrac{s}{2\|A^{-1}\|} - \tfrac{s}{4\|A^{-1}\|} = \tfrac{s}{4\|A^{-1}\|} \end{aligned}$$

so that $\psi(b) < \psi(c)$. But this contradicts the fact that $\psi(c)$ is the minimum value of $\psi(x)$ in $\overline{B}(b,s)$, so we have $c \in B(b,s)$, as claimed.

Claim 5(b): we have $f(c) = q$.
Suppose, for a contradiction, that $f(c) \neq q$ so we have $\psi(c) > 0$. Let $v = q - f(c)$ so that $|v| = \psi(c) > 0$. Let $u = A^{-1}v$ so that $v = Au$. Then for $0 \leq t \leq 1$, using Claim 2, we have

$$\begin{aligned} \psi(c + tu) &= \big|f(c + tu) - q\big| \leq \big|f(c + tu) - f(c) - Atu\big| + \big|f(c) + Atu - q\big| \\ &\leq \tfrac{|tu|}{2\|A^{-1}\|} + |tv - v| = \tfrac{t|A^{-1}v|}{2\|A^{-1}\|} + (1 - t)|v| \leq \tfrac{t}{2}\,|v| + (1 - t)|v| = \big(1 - \tfrac{t}{2}\big)|v|. \end{aligned}$$

Since $|v| > 0$ we have $\psi(c + tu) \leq \big(1 - \frac{t}{2}\big)|v| < |v| = \psi(c)$. But this again contradicts the fact that $\psi(x)$ attains its minimum value at c, and so we have $f(c) = q$, as claimed.

Claim 6: the function g is differentiable in V_0 with $Dg\big(f(b)\big) = Df(b)^{-1}$ for all $b \in U_0$.
Let $p \in V_0$ and let $b = g(p)$ so that $f(b) = p$. Let $B = Df(b)$. Note that B is invertible by Claim 1. Let $C = B^{-1}$. Let $y \in V_0$ and let $x = g(y) \in U_0$ so that $y = f(x)$. Then we have

$$\begin{aligned} \big|g(y) - g(p) - C(y - p)\big| &= \big|x - b - C(f(x) - f(b))\big| = \big|CB\big(x - b - C(f(x) - f(b))\big)\big| \\ &= \big|C\big(Bx - Bb - (f(x) - f(b))\big)\big| \leq \|C\|\big|f(x) - f(b) - B(x - b)\big| \end{aligned}.$$

Also, as shown above, we have $|y - p| = \big|f(x) - f(b)\big| \geq \frac{|x-b|}{2\|A^{-1}\|}$ so that

$$|x - b| \leq 2\|A^{-1}\|\,|y - p|.$$

It follows that g is differentiable at p with $Dg(p) = C = Df(b)^{-1}$, as claimed. Indeed, given $\epsilon > 0$, since f is differentiable at b with $Df(b) = B$ we can choose $\delta_1 > 0$ so that when $|x - a| < \delta_1$ we have $\big|f(x) - f(b) - B(x - b)\big| \leq \frac{\epsilon}{2\|A^{-1}\|\,\|C\|}|x - b|$, and since g is continuous at b we can choose $\delta > 0$ so that when $|y - p| < \delta$ we have $|x - b| = |g(y) - g(b)| < \delta_1$. When $|y - p| < \delta$, the above inequalities give $\big|g(y) - g(b) - C(y - p)\big| \leq \epsilon\,|y - p|$.

Claim 7: the function g is $\mathcal{C}^1$ in V_0.
By the cofactor formula for the inverse of a matrix, for all $y \in V_0$ and all indices k, ℓ,

$$\frac{\partial g_k}{\partial y_\ell}(y) = \big(Dg(y)\big)_{k,\ell} = \big(Df(g(y))^{-1}\big)_{k,\ell} = \frac{(-1)^{k+\ell}}{\det Df(g(y))}\det E$$

where is E is the matrix obtained from $Df(g(y))$ by removing the k^{th} column and the ℓ^{th} row. Thus $\frac{\partial g_k}{\partial y_\ell}(y)$ is a continuous function of y, as claimed.

6.27 Corollary: *(The Parametric Function Theorem) Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^{n+k}$ be $\mathcal{C}^1$. Let $a \in U$ and suppose that $Df(a)$ has rank n. Then* Range(f) *is locally equal to the graph of a $\mathcal{C}^1$ function.*

Proof: Since $Df(a)$ has maximal rank n, it follows that some $n \times n$ submatrix of $Df(a)$ is invertible. By reordering the variables in $\mathbb{R}^{n+k}$, if necessary, suppose that the top n rows of $Df(a)$ form an invertible $n \times n$ submatrix. Write $f(t) = \big(x(t), y(t)\big)$, where $x(t) = \big(x_1(t), \cdots, x_n(t)\big)$ and $y(t) = \big(y_1(t), \cdots, y_k(t)\big)$, so that we have

$$Df(t) = \begin{pmatrix} Dx(t) \\ Dy(t) \end{pmatrix}$$

with $Dx(a)$ invertible. By the Inverse function Theorem, the function $x(t)$ is locally invertible. Write the inverse function as $t = t(x)$ and let $g(x) = y\big(t(x)\big)$. Then, locally, we have $\text{Range}(f) = \text{Graph}(g)$ because if $(x,y) \in \text{Graph}(g)$ and we choose $t = t(x)$ then we have $(x,y) = \big(x, g(x)\big) = \big(x(t), g(x(t))\big) = \big(x(t), y(t)\big) \in \text{Range}(f)$ and, on the other hand, if $(x,y) \in \text{Range}(f)$, say $(x,y) = \big(x(t), y(t)\big)$ then we must have $t = t(x)$ so that $y(t) = y\big(t(x)\big) = g(x)$ so that $(x,y) = \big(x(t), y(t)\big) = \big(x, g(x)\big) \in \text{Graph}(g)$.

6.28 Corollary: *(The Implicit Function Theorem) Let $f : U \subseteq \mathbb{R}^{n+k} \to \mathbb{R}^k$ be $\mathcal{C}^1$. Let $p \in U$, suppose that $Df(p)$ has rank k and let $c = f(p)$. Then the level set $f^{-1}(c)$ is locally the graph of a $\mathcal{C}^1$ function.*

Proof: Since $Df(p)$ has rank k, it follows that some $k \times k$ submatrix of f is invertible. By reordering the variables in $\mathbb{R}^{n+k}$, if necessary, suppose that the last k columns of $Df(p)$ form an invertible $k \times k$ matrix. Write $p = (a,b)$ with $a = (p_1, \cdots, p_n) \in \mathbb{R}^n$ and $b = (p_{n+1}, \cdots, p_{n+k}) \in \mathbb{R}^k$ and write $z = f(x,y)$ with $x \in \mathbb{R}^n$, $y \in \mathbb{R}^k$ and $z \in \mathbb{R}^k$, and write

$$Df(x,y) = \big(\tfrac{\partial z}{\partial x}(x,y), \tfrac{\partial z}{\partial y}(x,y)\big)$$

with $\frac{\partial z}{\partial y}(a,b)$ invertible. Define $F : U \subseteq \mathbb{R}^{n+k} \to \mathbb{R}^{n+k}$ by $F(x,y) = \big(x, f(x,y)\big) = (w,z)$. Then we have

$$DF = \begin{pmatrix} I & O \\ \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{pmatrix}$$

with $DF(a,b)$ invertible. By the Inverse Function Theorem, $F = F(x,y)$ is locally invertible. Write the inverse function as $(x,y) = G(w,z) = \big(w, g(w,z)\big)$ and let $h(x) = g(x,c)$. Then, locally, we have $f^{-1}(c) = \text{Graph}(h)$ because

$$\begin{aligned} f(x,y) = c &\iff F(x,y) = (x,c) \iff (x,y) = G(x,c) \\ &\iff (x,y) = \big(x, g(x,c)\big) \iff (x,y) \in \text{Graph}(h). \end{aligned}$$

6.29 Remark: We can also find a formula for Dh where h is the function in the above proof. Since $G(w,z) = \big(w, g(w,z)\big)$ we have $DG(w,z) = \begin{pmatrix} I & O \\ \frac{\partial g}{\partial w} & \frac{\partial g}{\partial z} \end{pmatrix}$ and we also have $DG(w,z) = DF(x,y)^{-1} = \begin{pmatrix} I & O \\ -\big(\frac{\partial z}{\partial y}\big)^{-1} \frac{\partial z}{\partial x} & \big(\frac{\partial z}{\partial y}\big)^{-1} \end{pmatrix}$ so, since $h(x) = g(x,c)$, we have

$$Dh(x) = \tfrac{\partial g}{\partial w}(x,c) = -\big(\tfrac{\partial z}{\partial y}\big)^{-1} \tfrac{\partial z}{\partial x}(x,y).$$

6.30 Lemma: *(Iterated Limits) Let I and J be open intervals in $\mathbb{R}$ with $a \in I$ and $b \in J$, let $U = (I \times J) \setminus \{(a,b)\}$, and let $f : U \to \mathbb{R}$. Suppose that $\lim\limits_{y \to b} f(x,y)$ exists for every $x \in I$ and that $\lim\limits_{(x,y)\to(a,b)} f(x,y) = u \in \mathbb{R}$. Then $\lim\limits_{x \to a} \lim\limits_{t \to b} f(x,y) = u$.*

Proof: Define $g : I \to \mathbb{R}$ by $g(x) = \lim\limits_{y \to b} f(x,y)$. Let $\epsilon > 0$. Since $\lim\limits_{(x,y)\to(a,b)} f(x,y) = u$ we can choose $\delta > 0$ such that for all $(x,y) \in U$ with $0 < \big|(x,y) - (a,b)\big| \leq 2\delta$ we have $\big|f(x,y) - u\big| \leq \epsilon$. Let $x \in I$ with $0 < |x - a| \leq \delta$. For all $y \in J$ with $0 < |y - b| \leq \delta$ we have $0 < \big|(x,y) - (a,b)\big| \leq |x-a| + |y-b| \leq 2\delta$ and so $\big|f(x,y) - u\big| \leq \epsilon$ and hence

$$|g(x) - u| \leq \big|g(x) - f(x,y)\big| + \big|f(x,y) - u\big| \leq \big|g(x) - f(x,y)\big| + \epsilon.$$

Take the limit as $y \to b$ on both sides to get $|g(x) - u| \leq \epsilon$. Thus $\lim\limits_{x \to a} g(x) = u$, as required.

6.31 Theorem: *(Mixed Partials Commute) Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ where U is open in $\mathbb{R}^n$ with $a \in U$, and let $k, \ell \in \{1, \cdots, n\}$. Suppose $\frac{\partial^2 f}{\partial x_k \partial x_\ell}(x)$ exists in U and is continuous at a, $\frac{\partial f}{\partial x_k}(x)$ exists and is continuous in U, and $\frac{\partial^2 f}{\partial x_\ell \partial x_k}(a)$ exists. Then $\frac{\partial^2 f}{\partial x_\ell \partial x_k}(a) = \frac{\partial^2 f}{\partial x_k \partial x_\ell}(a)$.*

Proof: When $k = \ell$ there is nothing to prove, so suppose that $k \neq \ell$. Choose $r > 0$ so that $B(a, 2r) \subseteq U$. For $|x| < r$ and $|y| < r$ note that the points a, $a + xe_k$, $a + ye_\ell$ and $a + xe_k + ye_\ell$ all lie in $B(a, 2r)$. For $|X| < r$ and $|y| < r$, define

$$g(x,y) = f(a + xe_k + ye_\ell) - f(a + xe_k) - f(a + ye_\ell) + f(a).$$

By the Mean Value Theorem, applied to the function $f(a + xe_k + ye_\ell) - f(a + ye_\ell)$ as a function of y, we can choose t between 0 and y such that

$$y\Big(\tfrac{\partial f}{\partial x_\ell}(a + xe_k + te_\ell) - \tfrac{\partial f}{\partial x_\ell}(a + te_\ell)\Big) = g(x,y).$$

By the Mean Value Theorem, applied to the function $\frac{\partial f}{\partial x_\ell}(a + xe_k + te_\ell)$ as a function of x, we can choose s between 0 and x such that

$$x\, \tfrac{\partial^2 f}{\partial x_k \partial x_\ell}(a + se_k + te_\ell) = \tfrac{\partial f}{\partial x_\ell}(a + xe_k + te_\ell) - \tfrac{\partial f}{\partial x_\ell}(a + te_\ell).$$

Also by the Mean Value Theorem, applied to the function $f(a + xe_k + ye_\ell) - f(a + xe_k)$ as a function of x, we can choose r between 0 and x such that

$$x\Big(\tfrac{\partial f}{\partial x_k}(a + re_k + ye_\ell) - \tfrac{\partial f}{\partial x_k}(a + re_\ell)\Big) = g(x,y).$$

Then for $|x| < r$ and $0 < |y| < r$ we have

$$\frac{\frac{\partial f}{\partial x_k}(a + re_k + ye_\ell) - \frac{\partial f}{\partial x_k}(a + re_k)}{y} = \frac{\partial^2 f}{\partial x_k \partial x_\ell}(a + se_k + te_\ell).$$

Since $\frac{\partial^2 f}{\partial x_k \partial x_\ell}$ is continuous, the limit on the right as $(x,y) \to (0,0)$ is equal to $\frac{\partial^2 f}{\partial x_k \partial x_\ell}(a)$, and since $\frac{\partial f}{\partial x_k}$ is continuous, the limit as $y \to 0$ of the limit as $x \to 0$ on the left is equal to $\frac{\partial^2 f}{\partial x_\ell \partial x_k}(a)$, so the desired result follows from the above lemma.

6.32 Corollary: *If $U \subseteq \mathbb{R}^n$ is open and $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^2$ in U then we have $\frac{\partial^2 f}{\partial x_\ell \partial x_k}(x) = \frac{\partial^2 f}{\partial x_k \partial x_\ell}(x)$ for all $x \in U$ and for all k, ℓ.*

6.33 Exercise: Verify that for $f(x,y) = \frac{x^2}{x^2 + y^2}$ we have $\lim\limits_{x \to 0} \lim\limits_{y \to 0} f(x,y) \neq \lim\limits_{y \to 0} \lim\limits_{x \to 0} f(x,y)$.

6.34 Exercise: Let $f(x,y)=\left\{\begin{array}{cl}\dfrac{xy(x^2-y^2)}{x^2+y^2} & \text{, if } (x,y)\neq(0,0)\\ 0 & \text{, if } (x,y)=(0,0)\end{array}\right\}$. Verify that the mixed partial derivatives $\frac{\partial^2 f}{\partial x\partial y}(0,0)$ and $\frac{\partial^2 f}{\partial y\partial x}(0,0)$ both exist, but they are not equal.

6.35 Definition: for $f:U\subseteq\mathbb{R}^n\to\mathbb{R}$, where U is open in $\mathbb{R}^n$ with $a\in U$, we define $D^0f(a)=f(a)$ and for $\ell\in\mathbb{Z}^+$ we define the ℓ^{th} **total differential** of f at a to be the map $D^\ell f(a):\mathbb{R}^n\to\mathbb{R}$ given by

$$D^\ell f(a)(u)=\sum_{k_1=1}^n\sum_{k_2=1}^n\cdots\sum_{k_\ell=1}^n\frac{\partial^\ell f}{\partial x_{k_1}\partial x_{k_2}\cdots\partial x_{k_\ell}}(a)\,u_{k_1}u_{k_2}\cdots u_{k_\ell}$$

provided that all of the ℓ^{th} order partial derivatives exist at a.

6.36 Example: When $f:U\subseteq\mathbb{R}^2\to\mathbb{R}$ is $\mathcal{C}^2$ (so the mixed partial derivatives commute) we have

$$\begin{aligned}D^0f(u,v)&=f(a,b)\\ D^1f(a,b)(u,v)&=\tfrac{\partial f}{\partial x}(a,b)\,u+\tfrac{\partial f}{\partial y}(a,b)\,v\\ D^2f(a,b)(u,v)&=\tfrac{\partial f}{\partial x^2}(a,b)\,u^2+2\tfrac{\partial f}{\partial x\partial y}(a,b)\,uv+\tfrac{\partial f}{\partial y^2}(a,b)\,v^2.\end{aligned}$$

6.37 Theorem: *(Taylor's Theorem) Let $f:U\subseteq\mathbb{R}^n\to\mathbb{R}$ where U is open in $\mathbb{R}^n$. Suppose that the m^{th} oder partial derivatives of f all exist in U. Then for all $a,x\in U$ such that $[a,x]\subseteq U$ there exists $c\in[a,x]$ such that*

$$f(x)=\sum_{\ell=0}^{m-1}\tfrac{1}{\ell!}\,D^\ell f(a)(x-a)\;+\;\tfrac{1}{m!}\,D^mf(c)(x-a).$$

Proof: Let $a,x\in U$ with $[a,x]\subseteq U$. Let $\alpha(t)=a+t(x-a)$ for all $t\in\mathbb{R}$ and note that $\alpha(t)\in U$ for $0\le t\le 1$. Since U is open and α is continuous, we can choose $\delta>0$ so that $\alpha(t)\in U$ for all $t\in I=(-\delta,1+\delta)$. Define $g:I\to\mathbb{R}$ by $g(t)=f(\alpha(t))$. By the Chain Rule, we have

$$g'(t)=Df\big(\alpha(t)\big)\alpha'(t)=Df\big(\alpha(t)\big)(x-a)=\sum_{i=1}^n\tfrac{\partial f}{\partial x_i}\big(\alpha(t)\big)(x_i-a_i)=D^1f\big(\alpha(t)\big)(x-a).$$

By the Chain Rule again, we have

$$g''(t)=\sum_{i=1}^n\Big(\sum_{j=1}^n\tfrac{\partial^2 f}{\partial x_j\partial x_i}\big(\alpha(t)\big)(x_j-a_j)\Big)(x_i-a_i)=D^2f\big(\alpha(t)\big)(x-a).$$

An induction argument shows that

$$g^{(\ell)}(t)=D^\ell f\big(\alpha(t)\big)(x-a).$$

By Taylor's Theorem, applied to the function $g(t)$ on the interval $[0,1]$, we can choose $s\in[0,1]$ such that $g(1)=\sum\limits_{\ell=0}^{m-1}\tfrac{1}{\ell!}g^{(\ell)}(0)\;+\;\tfrac{1}{m!}g^{(m)}(s)$, that is

$$f(x)=\sum_{\ell=0}^{m-1}\tfrac{1}{\ell!}D^\ell f(a)(x-a)\;+\;\tfrac{1}{m!}D^mf\big(\alpha(s)\big)(x-a).$$

Thus we can choose $c=\alpha(s)\in[a,x]$.

Positive Definiteness and the Second Derivative Test

6.38 Definition: For $f: U \subseteq \mathbb{R}^n \to \mathbb{R}$, where U is open in $\mathbb{R}^n$ with $a \in U$, we define the m^{th} **Taylor polynomial** of f at a to be the polynomial

$$T^m f(a)(x) = \sum_{\ell=0}^{m} \tfrac{1}{\ell!} D^\ell f(a)(x-a)$$

provided that all the m^{th} order partial derivatives exist at a. When f is $\mathcal{C}^2$ in U (so that the mixed partial derivatives commute) we have

$$T^2 f(a)(x) = f(a) + Df(a)(x-a) + \tfrac{1}{2}(x-a)^T Hf(a)\,(x-a)$$

where $Hf(a) \in M_{n\times n}(\mathbb{R})$ is the symmetric matrix with entries $Hf(a)_{k,\ell} = \frac{\partial^2 f}{\partial x_k \partial x_\ell}(a)$. The matrix $Hf(a)$ is called the **Hessian matrix** of f at a.

6.39 Definition: Let $A \in M_n(\mathbb{R})$ be a symmetric matrix. We say that
(1) A is **positive-definite** when $u^T A u > 0$ for all $0 \neq u \in \mathbb{R}^n$,
(2) A is **negative-definite** when $u^T A u < 0$ for all $0 \neq u \in \mathbb{R}^n$, and
(3) A is **indefinite** when there exist $0 \neq u, v \in \mathbb{R}^n$ with $u^T A u > 0$ and $v^T A v < 0$.

6.40 Theorem: *(Characterization of Positive-Definiteness by Eigenvalues) Let $A \in M_n(\mathbb{R})$ be symmetric. Then*
(1) A is positive-definite if and only if all of the eigenvalues of A are positive,
(2) A is negative-definite if and only if all of the eigenvalues of A are negative, and
(3) A is indefinite if and only if A has a positive eigenvalue and a negative eigenvalue.

Proof: Suppose that A is positive definite. Let λ be an eigenvalue of A and let u be a unit eigenvector for λ. Then $\lambda = \lambda |u|^2 = \lambda(u \cdot u) = \lambda u \cdot u = Au \cdot u = u^T A u > 0$. Conversely, suppose that all of the eigenvalues of A are positive. Since A is symmetric, we can orthogonally diagonalize A. Choose a matrix $P \in M_n(\mathbb{R})$ with $P^T = P$ so that $P^T A P = D = \text{diag}(\lambda_1, \cdots, \lambda_n)$. Given $0 \neq u \in \mathbb{R}^n$, let $v = P^T u$. Note that $v \neq 0$ since P^T is invertible. Thus $u^T A u = u^T P D P^T u = v^T D v = \sum_{i=1}^{n} \lambda_i {v_i}^2 > 0$ since every $\lambda_i > 0$ and some $v_i \neq 0$. This proves Part (1). The proofs of Parts (2) and (3) are fairly similar.

6.41 Theorem: *(Characterization of Positive-Definiteness by Determinant) Let $A \in M_n(\mathbb{R})$ be symmetric. For each k with $1 \leq k \leq n$, let $A^{(k)}$ denote the upper-left $k \times k$ sub matrix of A. Then*
(1) A is positive-definite if and only if $\det(A^{(k)}) > 0$ for all k with $1 \leq k \leq n$, and
(2) A is negative-definite if and only if $(-1)^k \det(A^{(k)}) > 0$ for all k with $1 \leq k \leq n$.

Proof: Part (2) follows easily from Part (1) by noting that A is negative-definite if and only if $-A$ is positive-definite. We shall prove one direction of Part (1). Suppose that A is positive-definite. Let $1 \leq k \leq n$. Since $u^T A u > 0$ for all $0 \neq u \in \mathbb{R}^n$, we have $\begin{pmatrix} u^T & 0 \end{pmatrix} A \begin{pmatrix} u \\ 0 \end{pmatrix} = 0$, or equivalently $u^T A^{(k)} u > 0$, for all $0 \neq u \in \mathbb{R}^k$. This shows that $A^{(k)}$ is positive definite. By the previous theorem, all of the eigenvalues of $A^{(k)}$ are positive. Since $\det(A^{(k)})$ is equal to the product of its eigenvalues, we see that $\det(A^{(k)}) > 0$.

The proof of the other direction of Part (1) is more difficult. We shall omit the proof. It is often proven in a linear algebra course.

6.42 Exercise: Let $A = \begin{pmatrix} 3 & -1 & 2 \\ -1 & 2 & 1 \\ 2 & 1 & 5 \end{pmatrix}$. Determine whether A is positive-definite.

6.43 Definition: Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $a \in A$. We say that f has a **local maximum value** at a when there exists $r > 0$ such that $f(a) \geq f(x)$ for all $x \in B_A(a,r)$. We say that f has a **local minimum value** at a when there exists $r > 0$ such that $f(a) \leq x$ for all $x \in B_A(a,r)$.

6.44 Exercise: Show that when $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ where U is open in $\mathbb{R}^n$ with $a \in U$, if f has a local maximum or minimum value at a then either $Df(a) = 0$ or $Df(a)$ does not exist (that is one of the partial derivatives $\frac{\partial f}{\partial x_k}(a)$ does not exist).

6.45 Definition: Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ where U is open in $\mathbb{R}^n$. For $a \in U$, we say that a is a **critical point** of f when either $Df(a) = 0$ or $Df(a)$ does not exist. When $a \in U$ is a critical point of f but f does not have a local maximum or minimum value at a, we say that a is a **saddle point** of f.

6.46 Theorem: *(The Second Derivative Test) Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ with U open in $\mathbb{R}^n$ and let $a \in U$. Suppose that f is $\mathcal{C}^2$ in U with $Df(a) = 0$. Then*
(1) if $Hf(a)$ is positive definite then f has a local minimum value at a,
(2) if $Hf(a)$ is negative definite then f has a local maximum value at a, and
(3) if $Hf(a)$ is indefinite then f has a saddle point at a.

Proof: Suppose that $Hf(a)$ is positive-definite. Then $\det\big(Hf(a)^{(k)}\big) > 0$ for $1 \leq k \leq n$. Since each determinant function $\det(A^{(k)})$ is continuous as a function in the entries of the matrix A, the set $V = \big\{x \in U \,\big|\, Hf(x)^{(k)} > 0 \text{ for } k = 1, 2, \cdots, n\big\}$ is open. Choose $r > 0$ so that $B(a,r) \subseteq V$. Then we have $u^T Hf(c)\, u > 0$ for all $0 \neq u \in \mathbb{R}^n$ and all $c \in B(a,r)$. Let $x \in B(a,r)$ with $x \neq a$. By Taylor's Theorem, we have

$$f(x) - f(a) - Df(a)(x-a) = (x-a)^T Hf(c)\,(x-a)$$

for some $c \in [a,x]$. Since $Df(a) = 0$ and $Hf(c)$ is positive-definite, we have $f(x) - f(a) > 0$. Thus f has a local minimum value at a. This proves Part (1) and Part (2) is similar.

Let us prove Part (3). Suppose there exists $0 \neq u \in \mathbb{R}^n$ such that $u^T Hf(a)\, u > 0$. Let $r > 0$ with $B(a,r) \subseteq U$ and scale the vector u if necessary so that $[a,u] \subseteq B(a,r)$. Let $\alpha(t) = a + tu$ and let $g(t) = f\big(\alpha(t)\big)$ for $0 \leq t \leq 1$. As in the proof of Taylor's Theorem, we have

$$g'(t) = \sum_{i=1}^{n} \tfrac{\partial f}{\partial x_i}\big(\alpha(t)\big)\, u_i = Df\big(\alpha(t)\big)\, u \text{ , and}$$

$$g''(t) = \sum_{i,j=1}^{n} \tfrac{\partial^2 f}{\partial x_i \partial x_j}\big(\alpha(t)\big)\, u_i u_j = u^T Hf\big(\alpha(t)\big)\, u\,.$$

Since $g(0) = f(a)$, $g'(0) = Df(a)\, u = 0$ and $g''(0) = u^T Hf(a)\, u > 0$, it follows from single-variable calculus that we can choose t_0 with $0 < t_0 < 1$ so that $g(t_0) > g(0)$. When $x = \alpha(t_0)$ we have $x \in B(a,r)$ and $f(x) = f\big(\alpha(t_0)\big) = g(t_0) > g(0) = f(a)$, and so f does not have a local maximum value at a. Similarly, if there exists $0 \neq v \in \mathbb{R}^n$ such that $v^T Hf(a)\, v < 0$ then f does not have a local minimum value at a. Thus when $Hf(a)$ is indefinite, f has a saddle point at a.

6.47 Exercise: Find and classify the critical points of the following functions $f : \mathbb{R}^2 \to \mathbb{R}$.
(a) $f(x,y) = x^3 + 2xy + y^2$ (b) $f(x,y) = x^3 + 3x^2y - 6y^2$ (c) $f(x,y) = x^2 y\, e^{-x^2 - 2y^2}$