Appendix 1. Review of Differentiation

1.1 Remark: In this appendix we shall review some of the theory of differentiation of vector valued functions of several variables, as presented in MATH 247, including The Inverse Function Theorem (which is not usually proven in MATH 237).

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

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

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

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

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

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.

1.3 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

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

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 $a \in U$.

1.4 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

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

1.5 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 $|x - a| \le \delta$ we have $\big| f(x) - f(a) - A(x - a) \big| \le \epsilon\, |x - a|$. Let $t \in \mathbb{R}$ with $|t| \le \delta$. Let $x = a + t\, e_\ell$. Then we have $|x - a| = |te_\ell| = |t| \le \delta$ and so $\big| f(x) - f(a) - A(x - a) \big| \le \epsilon\, |x - a|$. Since for any vector $u \in \mathbb{R}^m$ we have $|u_k| \le |u|$, we have

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

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

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

$$\|A\| = \max\big\{|Ax| \, \big| |x| = 1\big\}.$$

1.7 Lemma: *(Properties of the Matrix Norm) Let $A \in M_{m\times n}(\mathbb{R})$. Then*
(1) $|Ax| \leq \|A\| \, |x|$ for all $x \in \mathbb{R}^n$,
(2) if A is invertible then $|Ax| \geq \frac{|x|}{\|A^{-1}\|}$ for all $x \in \mathbb{R}^n$,
(3) $\|A\| \leq \sum\limits_{k=1}^{m} \sum\limits_{\ell=1}^{n} |A_{k,\ell}|$, 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 $0 \neq x \in \mathbb{R}^n$ we have

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

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 $Ax \neq 0$ 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 $|x| = \big|A^{-1}A\,x\big| \leq \|A^{-1}\| \, |Ax|$ so that $|Ax| \geq \frac{|x|}{\|A^{-1}\|}$, as required. To prove Part 3, let $x \in \mathbb{R}^n$ with $|x| = 1$. Then $|x_\ell| \leq |x| \leq 1$ for all indices ℓ, and so

$$\big|Ax\big| = \Big| \sum_{k=1}^{m} (Ax)_k e_k \Big| \leq \sum_{k=1}^{m} \big|(Ax)_k\big| = \sum_{k=1}^{m} \Big| \sum_{\ell=1}^{n} A_{k,\ell} x_\ell \Big| \leq \sum_{k=1}^{m} \sum_{\ell=1}^{n} |A_{k,\ell}| \, |x_\ell| \leq \sum_{k=1}^{m} \sum_{\ell=1}^{n} |A_{k,\ell}|.$$

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

1.8 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

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

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

$$|x-a| \leq \delta \implies \big|f(x) - f(a) - Df(a)(x-a)\big| \leq |x-a|$$

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

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

1.9 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

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

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.

1.10 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 $\alpha(t)$.

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

1.12 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 $\nabla f(a)$.

1.13 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 $|y - a| \leq \delta \Longrightarrow \left|\frac{\partial f_k}{\partial x_\ell}(y) - \frac{\partial f_k}{\partial x_\ell}(a)\right| \leq \frac{\epsilon}{nm}$. 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 we have

$$f(x) - f(a) - Df(a)(x - a) = \big(B - Df(a)\big)(x - a)$$

and so (by Part 2 of Lemma 5.7)

$$\big|f(x) - f(a) - Df(a)(x - a)\big| \leq \big\|B - Df(a)\big\|\,|x - a| \leq \sum_{k,\ell} \left|\frac{\partial f_k}{\partial x_\ell}\big(\alpha_\ell(s_{k,\ell})\big) - \frac{\partial f_k}{\partial x_\ell}(a)\right| \leq \epsilon|x - a|.$$

1.14 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.*

1.15 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.*

1.16 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) $f(x,y) = \frac{x^3-3xy^2}{x^2+y^2}$

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

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

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

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

1.19 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*

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

Proof: Let $\alpha(t) = a + t(b-a)$ and define $g : [0,1] \to \mathbb{R}$ by $g(t) = f\big(\alpha(t)\big) \bullet u$. By the Chain Rule, we have $g'(t) = \big(Df(\alpha(t))\alpha'(t)\big) \bullet u = \big(Df(\alpha(t))(b-a)\big) \bullet u$. 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 $\big(Df(\alpha(s))(b-a)\big) \bullet u = f(b) \bullet u - f(a) \bullet u = \big(f(b) - f(a)\big) \bullet u$. Thus we can take $c = \alpha(s) \in [a,b]$ to get $Df(c)(b-a) \bullet u = \big(f(b) - f(a)\big) \bullet u$.

1.20 Corollary: *(Vanishing Derivative) Let $U \subseteq \mathbb{R}^n$ be open and connected and let $f : U \to \mathbb{R}^n$ 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 $\big(Df(d)(c-b)\big) \bullet u = \big(f(c) - f(b)\big) \bullet u$. Then we have

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

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.

1.21 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 $\left|\frac{\partial f_k}{\partial x_\ell}(x) - \frac{\partial f_k}{\partial f_\ell}(a)\right| \le \frac{1}{2n^2\|A^{-1}\|}$ for all k, ℓ. Let $U_0 = B(a,r)$ and note that for all $x \in U_0$ we have $\big\|Df(x) - A\big\| \le \frac{1}{2\|A^{-1}\|}$.

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 $Df(a)u = 0$. But then we have

$$\big\|Df(x) - A\big\| \ge \big|(Df(x) - A)u\big| = \big|Au\big| \ge \frac{|u|}{\|A^{-1}\|} = \frac{1}{\|A^{-1}\|}$$

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

Claim 2: for all $b, c \in U_0$ we have $\big|f(c) - f(b) - A(c-b)\big| \le \frac{\|c-b\|}{2\|A^{-1}\|}$.
Let $b, c \in U_0$. Let $\alpha(t) = b + t(c-b)$ and note that we $\alpha(t) \in U_0$ 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 $L(a) = f(a) + Df(a)(x-a)$. 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

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

By the Mean Value Theorem we have $\big|\phi(1) - \phi(0)\big| \le \max_{0 \le t \le 1} \big|\phi'(t)\big| \le \frac{|c-b|}{2\|A^{-1}\|}$ and note that $\phi(1) - \phi(0) = \big(f(c) - L(c)\big) - \big(f(b) - L(b)\big) = f(c) - f(b) - A(c-b)$, and so

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

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

$$\big|f(c) - f(b) - A(c-b)\big| \ge \big|A(c-b)\big| - \big|f(c) - f(b)\big| \ge \frac{|c-b|}{\|A^{-1}\|} - \big|f(c) - f(b)\big|$$

and so, by Claim 3, we have

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

It follows that when $b \ne c$ we have $f(b) \ne f(c)$, 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 $b \ne c$ we have $\big|f(c) - f(b)\big| \ge \frac{|c-b|}{2\|A^{-1}\|} > 0$ so that $f(b) \ne f(c)$. Thus the restriction of f to U_0 is injective, as claimed.

Claim 5: the inverse $g = f^{-1} : V_0 \to U_0$ is continuous (indeed uniformly continuous).
Let $p, q \in V_0$. Let $b = g(p)$ and $c = g(q)$ so that $p = f(b)$ and $q = f(c)$. By Claim 3 we have $|c-b| \le 2\|A^{-1}\|\,\big|f(c) - f(b)\big|$, that is $\big|g(q) - g(p)\big| \le 2\|A^{-1}\|\,|q-p|$. It follows that g is uniformly continuous in V_0. (We remark that this claim is not used anywhere in the proof and we included it simply because it fits neatly nestled at the bottom of the page).

Claim 6: 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 6(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

$$\psi(b) = \big|f(b) - q\big| = |p - q| < \tfrac{s}{4\|A^{-1}\|} \text{ and, using Claim 3,}$$

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

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 6(b): we have $f(c) = q$.
Suppose, for a contradiction, that $f(c) \neq q$ so we have $\psi(c) > 0$. Let $v = f(c) - q$ 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 7: 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

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

and, 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.

Claim 8: 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.

1.22 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 $(x, y) = \big(x(t), y(t)\big) = \big(x, g(x)\big) \in \text{Graph}\,(g)$.

1.23 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

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

1.24 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

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