1 The Implicit Function Theorem and Implications Thereof

1. $f(x) = Ax$ can be solved $\iff$ rank$A = n$.
2. $f(x) = Ax + b$ can be solved also if rank$A = n$.
3. Suppose differentiable $f : \mathbb{R}^n \to \mathbb{R}^n$, i.e., $\exists \mathbf{D}f_{x_0} \in L(\mathbb{R}^n, \mathbb{R}^n)$ s.t. $f(x) - f(x_0) - \mathbf{D}f_{x_0}(x - x_0)$ is $o(|x - x_0|)$, where $L(\mathbb{R}^n, \mathbb{R}^n)$ is the set of linear functions. In other words,

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - \mathbf{D}f_{x_0}(x - x_0)}{|x - x_0|} = 0.$$

Then, the purpose of the following discussion is so that we can solve for x given $y = f(x)$, at least locally.

Definition 1.1 (Little "oh" notation)**.** *A function $g(x - x_0)$ is "little oh" of a norm, say, $|x - x_0|$ iff*

$$\lim_{x \to x_0} \frac{g(x - x_0)}{|x - x_0|} = 0.$$

The derivative is defined using o notation, that is, $\mathbf{D}f_x$ is called the derivative of f at x_0 iff

$$f(x) - f(x_0) - \mathbf{D}f_{x_0}(x - x_0) = o(\|x - x_0\|).$$

While we're on the subject,

Definition 1.2 (Big"oh" notation)**.** *A function $f(x - x_0)$ is "big oh" of a norm, say, $|x - x_0|$ iff $\frac{f(x - x_0)}{|x - x_0|}$ is bounded in a neighborhood of x_0.*

We may want to solve $y = f(x)$ for x. Typically, if f is some nonlinear function, so first we hope that we can at least do so locally. Moreover, we can only do so locally if the function is one-to-one in that local neighbordhood. The IFT and various corollaries allow us to do that, and even get some additional local results with the derivative $\mathbf{D}f_{x_0}$ is not invertible (or nonsquare).

Remark 1.3 (Infinite Dimensions)**.** *Implicit in our discussion above is that for a bounded and invertible linear map $\mathbf{D}f_x$, i.e., a square matrix, there exists a bounded inverse. That is not always the case in infinite dimensional (linear) vector spaces, such as $\mathcal{C}$.*

Example 1.4. *An example is, suppose we want to solve following equation in $\mathcal{C}$,*

$$\left(-\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}\right) u + u^3 = F,$$

for u given F. In some sense, we have a function $G(u) = F$ that we wish to invert. We can compute

$$\mathbf{D}G_u(v) = \left(-\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}\right) v + u^2 v,$$

and we have reduced our nonlinear PDE to a linear PDE. The question is, when is $\mathbf{D}G_u$ invertible? If we can answer positively, then we can solve the linearized PDE.

1.1 Linear Algebra: A Review for Infinite Dimensions

Let $\mathcal{X}$ be a normed linear space. This means that there is a function $|\cdot|_{\mathcal{X}} : \mathcal{X} \to \mathbb{R}_+$ s.t.

1. $|0|_{\mathcal{X}} = 0$
2. $|\alpha x|_{\mathcal{X}} = \alpha |x|_{\mathcal{X}}$
3. $|x + y|_{\mathcal{X}} \leq |x|_{\mathcal{X}} + |y|_{\mathcal{X}}$.

For normed linear spaces, $|\cdot|$ defines a distance metric.

Example 1.5. *For the* $\mathcal{C}([0,1],\mathbb{R})$,

$$|f| = \sup_{x\in[0,1]} |f(x)|.$$

For the remainder of this section, we assume $\mathcal{X}, \mathcal{Y}$ are linear spaces unless otherwise stated.

Definition 1.6 (Banach Space)**.** *A Banach space is a normed linear space that is complete.*

Let $\mathcal{B}(\mathcal{X},\mathcal{Y})$ be the set of bounded linear functions from $\mathcal{X}$ to $\mathcal{Y}$. Let $\mathcal{L}(\mathcal{X},\mathcal{Y})$ be the set of any linear function from $\mathcal{X}$ to $\mathcal{Y}$. We have that $\mathcal{B}(\mathcal{X},\mathcal{Y}) \subset \mathcal{L}(\mathcal{X},\mathcal{Y})$, with equality if $\mathcal{X}$ is finite dimensional.

Example 1.7. *Let*

$$\mathcal{X} = \{a_1, a_2, \cdots \mid \text{ only finitely many } a_j > 0\}$$

Note that $\mathcal{X}$ *is a linear vector space. It has the* ℓ_1 *norm, i.e., for a vector* $a \in \mathcal{X}$, $|a|_{\mathcal{X}} = \sum_j |a_j|_{\mathbb{R}}$. *Now define a linear function on* $\mathcal{X}$ *by*

$$La = (a_1, 2a_2, 3a_3, \dots).$$

Clearly, $La \in \mathcal{X}$ *for any* $a \in \mathcal{X}$. *However, it is an unbounded operator. To verify this fact, let*

$$a = (0, \dots, 0, \underbrace{1}_{\text{the } n\text{-th spot}}, 0, \dots).$$

so that

$$La = (0, \dots, 0, n, 0, \dots).$$

Note that L *is thus unbounded. Note that* L^{-1} *is given by*

$$Sa = (a_1, \frac{1}{2}a_2, \frac{1}{3}a_3, \dots).$$

Then, $SL = LS = I$. *We say that* S *is invertible if* **we can find a bounded inverse***, i.e., not just any inverse will do. But* L *is the inverse of* S*, and we just showed that* L *is unbounded. Thus,* $S \notin \mathcal{B}(\mathcal{X},\mathcal{Y})$.

In summary, we have the following definition.

Definition 1.8 (Invertible Linear Function)**.** *A linear function is invertible iff its inverse is a bounded linear function.*

Definition 1.9. *A function* $f : \mathcal{X} \to \mathcal{Y}$ *is* $\mathcal{C}^1$ *if* f *is differentiable for all* $x \in \mathcal{X}$*, and* $\mathbf{D}f : \mathcal{X} \to \mathcal{B}(\mathcal{X},\mathcal{Y})$ *is continuous.*

For $\mathbf{D}f : \mathcal{X} \to \mathcal{B}(\mathcal{X},\mathcal{Y})$ to be a continuous function, obviously we need that $\mathcal{B}(\mathcal{X},\mathcal{Y})$ to a be a metric space (otherwise what would continuity even mean?). Then, we need a norm for $\mathcal{B}(\mathcal{X},\mathcal{Y})$.

$$|A| = \sup_{v\neq 0} \frac{|Av|_{\mathcal{Y}}}{|v|_{\mathcal{X}}} = \sup_{|v|_{\mathcal{X}}=1} |Av|_{\mathcal{Y}}.$$

Lemma 1.10. *If* $B \in \mathcal{B}(\mathcal{X},\mathcal{Y})$ *and* $A \in B(\mathcal{Y},\mathcal{Z})$*, then* $AB \in B(\mathcal{X},\mathcal{Z})$ *and* $|AB| \leq |A|\,|B|$.

Proof. Suppose $Bv \neq 0$ (if it does equal zero then proof is trivial). Then

$$\frac{|ABv|}{|v|} = \frac{|ABv|}{|Bv|}\frac{|Bv|}{|v|} \leq |A|\,|B|$$

□

Example 1.11. *Let*

$$\ell_1(\mathbb{R}) = \left\{ \{a_k\}_{k\in\mathbb{N}} \mid \sum_{k\in\mathbb{N}} |a_k| < \infty \right\}.$$

Define $A \in \mathcal{B}(\ell_1(\mathbb{R}), \ell_1(\mathbb{R}))$ *by*

$$A(v_1, v_2, \dots) = (0, v_1, v_2, \dots)$$

to be the shift. Note that A is injective but not surjective. Let also

$$B(v_1, v_2, \dots) = (v_2, v_3, \dots).$$

It is not injective, but it is surjective. The composition does not commute, i.e.,

$$B \circ A = I, \text{ but } A \circ B \neq I.$$

In what follows, $\mathcal{M}$ is a metric space, $\mathcal{B}$ is the set of bounded linear functions, and $\mathcal{X}$ and $\mathcal{Y}$ are linear vector spaces.

Lemma 1.12. *Let* $A \in \mathcal{C}(\mathcal{M}, \mathcal{B}(\mathcal{X}, \mathcal{Y}))$. *Let* $p \in \mathcal{M}$ s.t. $A(p)$ *is invertible. Then there is a neighborhood* U *of* p *such that* $\forall q \in U$, $q \mapsto A^{-1}(q)$ *is continuous.*

Proof. Let $W(q) = A(p)^{-1}A(q)$ so that $W : \mathcal{M} \to \mathcal{B}(\mathcal{X}, \mathcal{X})$. Pictorially,

$$\mathcal{M} \xrightarrow{A} \mathcal{B}(\mathcal{X}, \mathcal{Y}) \xrightarrow{A(p)^{-1}} \mathcal{B}(\mathcal{Y}, \mathcal{X}).$$

So $W(p) = I$, i.e., $|I - W(p)| = 0$. Also note that $W(q)$ is continuous because $A(q)$ is continuous, and all W is doing is premultiplying by a constant. Thus, $|I - W(q)|$ is a continuous function as a composition of continuous maps, and there is a neighborhood U of p such that $|I - W(q)| \leq \frac{1}{2}$ for $q \in U$. Note

$$W(q) = I - (I - W(q)).$$

Let's try to invert $I - W(q)$. The geometric series

$$\frac{1}{1-r} = 1 + r + r^2 + \cdots$$

suggests that we consider

$$\sum_{k=0}^{\infty} (I - W(q))^k \triangleq PW(q)^{-1}$$

the putative inverse of $W(q)$. Note that the series converges if it converges absolutely and $\left|(I - W(q))^k\right| \leq |I - W(q)|^k \leq (1/2)^k$, which converges. By the M-test, the series $PW(q)^{-1}$ converges. We have that

$$\begin{aligned}
W(q)PW(q)^{-1} &= W(q)\sum_{k=0}^{\infty}(I - W(q))^k \\
&= \lim_{n\to\infty} W(q)\sum_{k=0}^{n}(I - W(q))^k \\
&= \lim_{n\to\infty} \sum_{k=0}^{n}(W(q) + I - I)(I - W(q))^k \\
&= \lim_{n\to\infty} -\sum_{k=0}^{n}(I - W(q))^{k+1} + \sum_{k=0}^{n}(I - W(q))^k \\
&= \lim_{n\to\infty} \sum_{k=0}^{n}(I - W(q))^k - \sum_{k=1}^{n+1}(I - W(q))^k \\
&= \lim_{n\to\infty} I - (I - W(q))^{n+1} \\
&= I
\end{aligned}$$

Similarly, $PW(q)^{-1}$ is the left inverse of $W(q)$. This proves that $PW(q)^{-1}$ is the actual inverse of $W(q)$. From now on we just call it $W(q)^{-1}$. Then,

$$A(q)^{-1} = W(q)^{-1} \circ A(p)^{-1},$$

which is continuous because $W(q)^{-1}$ is a continuous and $A(p)^{-1}$ is a constant, as desired. $\square$

Theorem 1.13 (Inverse Function Theorem)**.** *Let $A \subset \mathbb{R}^n$ be open, $f : A \to \mathbb{R}^n$ a $\mathcal{C}^1$ function. Let $x_0 \in A$. Suppose $\mathbf{D}f_{x_0}$ is invertible. Then we can find an open neighborhood U of x_0 and an open neighborhood W of $f(x_0)$ such that for each $x \in U$, $f(x) \in W$ and there is a $\mathcal{C}^1$ function $\phi : W \to U$ such that ϕ is both a left and right inverse to f, that is*

$$f \circ \phi(y) = y \ \text{ and } \phi \circ f(x) = x$$

i.e, $\phi = f|_U^{-1}$. Moreover, $\mathbf{D}f_x^{-1} = \mathbf{D}\phi_{f(x)}$ and if f is $\mathcal{C}^k$, then so is ϕ.

Remark 1.14. *The Inverse Function Theorem holds for infinite dimensional linear spaces. Just replace the finite dimensional spaces $\mathbb{R}^n$ with a Banach space and everything still holds. In the following proof, however, we restrict ourselves to finite dimensions because we have not yet learned the chain rule for infinite dimensions. We'll make a note in the proof where we have to assume finite dimensional Banach space.*

Proof. The first step of the proof is book keeping. Essentially, we'll want to show that if the theorem holds for a function f such that $f(0) = 0$ and $\mathbf{D}f_0 = I$, then it holds for an arbitrary f that also meets the conditions of the theorem. Define

$$\tilde{f}(x) = [\mathbf{D}f_{x_0}]^{-1}(f(x_0 + x) - f(x_0)).$$

We have that $\tilde{f}(0) = 0$, and

$$\begin{aligned} \mathbf{D}\tilde{f}_0 &= \mathbf{D}([\mathbf{D}f_{x_0}]^{-1}\mathbf{D}f(x_0)) \\ &= [\mathbf{D}f_{x_0}]^{-1}\mathbf{D}f(x_0) \\ &= I. \end{aligned}$$

Now suppose g inverts $\tilde{f}$. Then,

$$\begin{aligned} \tilde{f} \circ g(y) &= y \\ \mathbf{D}f_{x_0}^{-1} \circ (f(x_0 + g(y)) - f(x_0)) &= y \\ f(x_0 + g(y)) - f(x_0) &= \mathbf{D}f_{x_0} y \\ f(x_0 + g(y)) &= \mathbf{D}f_{x_0} y + f(x_0) \end{aligned}$$

then, since y was arbitrary,

$$\begin{aligned} f(x_0 + g(\mathbf{D}f_{x_0}^{-1} y)) &= \mathbf{D}f_{x_0}\mathbf{D}f_{x_0}^{-1} y + f(x_0) \\ f(x_0 + g(\mathbf{D}f_{x_0}^{-1} y)) &= y + f(x_0) \end{aligned}$$

Again, since y was arbitrary,

$$f(x_0 + g(\mathbf{D}f_{x_0}^{-1}(y - f(x_0)))) = y.$$

Set $G(y) = x_0 + g(\mathbf{D}f_{x_0}^{-1}(y - f(x_0)))$. Thus, $f \circ G(y) = y$ and $G \circ f(x) = x$. Therefore, to prove the theorem, we only need to consider functions such that $f(0) = 0$ and $\mathbf{D}f_0 = I$.

We proceed proving the theorem assuming that $f(0) = 0$ and $\mathbf{D}f(0) = I$. Essentially, what we are setting out to do in this theorem is, given y in a neighborhood of 0, find x also in a neighborhood of 0 such that $f(x) = y$.

Define $\psi_y(x) = y + x - f(x)$. We have that

$$\psi_y(x) = x \iff y = f(x).$$

First we'll show that ψ_y is a contraction map. For this, let

$$\alpha(t) = \psi_y((1-t)z + tx)$$

so that

$$\psi_y(x) - \psi_y(z) = \alpha(1) - \alpha(0) = \int_0^1 \alpha'(t)dt.$$

Note that

$$|\psi_y(x) - \psi_y(z)| = |\alpha(1) - \alpha(0)| \le \int_0^1 |\alpha'(t)|\, dt.$$

Applying the chain rule (for infinite dimensions, if you like),

$$\alpha'(t) = (\mathbf{D}\psi_y)_{(1-t)z+tx}(x-z).$$

Taking the norm and applying the triangle inequality and taking the sup,

$$|\alpha'(t)| \le \left|(\mathbf{D}\psi_y)_{(1-t)z+tx}\right| |(x-z)| \le \sup_{0\le t\le 1} \left|(\mathbf{D}\psi_y)_{(1-t)z+tx}\right| |(x-z)|$$

Now, we know that $|(\mathbf{D}\psi_y)_x| = I - \mathbf{D}f_0 = I - I = 0$, and we also assumed that the derivative is continuous in the hypothesis. Thus, we can find a $\delta > 0$ such that

$$|(\mathbf{D}\psi_y)_x| < \frac{1}{2} \quad \forall x \in \bar{B}_\delta(0).$$

where $\bar{B}_\delta(0)$ denotes the closed δ-ball centered at zero. Note that $\bar{B}_\delta(0)$ is convex, so if $z, x \in \bar{B}_\delta(0)$, then so is $(1-t)z + tx$ for $0 \le t \le 1$. Thus,

$$|\psi_y(x) - \psi_y(z)| < \frac{1}{2},$$

so ψ_y is a contraction map as long as it maps $\bar{B}_\delta(0)$ onto $\bar{B}_\delta(0)$. To verify this directly

$$\begin{aligned}\psi_y(x) &= y + x - f(x)\\ &\le |y| + |x - f(x)|\\ &= |y| + |\psi_y(x) - \psi_y(0)|\\ &\le |y| + \frac{1}{2}|x|\\ &\le |y| + \frac{1}{2}\delta,\end{aligned}$$

so ψ_y is a contraction map as long as $|y| < \delta$, which is fine because we only claimed the theorem held locally.

Recall the contraction mapping theorem, and apply it to the current state of things. In particular, it tells us that

$$\exists_! x_y \in \bar{B}_{\frac{\delta}{2}}(0) \text{ s.t. } \psi_y(x_y) = x_y, \text{ i.e., } f(x_y) = y.$$

Let $g(y) = x_y$. Thus far, we only showed that

$$\exists g : \bar{B}_{\frac{\delta}{2}}(0) \to \bar{B}_\delta(0) \text{ s.t. } f(g(y)) = y.$$

We need to show that g is continuous before we can prove anything about the differentiability of it. Let $y_1, y_2 \in \bar{B}_{\frac{\delta}{2}}(0)$. For $i = 1, 2$,

$$\psi_y(g(y_i)) = g(y_i)$$

and we can write an identity based on the fact that f inverts g,

$$y_i + g(y_i) - f(g(y_i)) = g(y_i)$$

Then

$$\begin{aligned}|g(y_1)-g(y_2)| &= |y_1+g(y_1)-f(g(y_1))-y_2+g(y_2)-f(g(y_2))| \\ &\le |y_1-y_2|+|g(y_1)-f(g(y_1))+g(y_2)-f(g(y_2))| \\ &= |y_1-y_2|+|\psi_0(g(y_1))-\psi_0(g(y_2))| \\ &\le |y_1-y_2|+\frac{1}{2}\left|g(y_1)-g(y_2)\right|\end{aligned}$$

which implies that $\frac{1}{2}\left|g(y_1)-g(y_2)\right| \le |y_1-y_2|$. Thus g is Lipschitz, which is in fact stronger than continuity. Now we need to show g is differentiable and show that its derivative is equal to the inverse of the derivative of f. We do this by showing that $\mathbf{D}f_x^{-1}$ meets the requirements of the derivative of g. In particular,

$$\begin{aligned}\left|g(y)-g(w)-\mathbf{D}f_x^{-1}(y-x)\right| &= \left|-\mathbf{D}f_x^{-1}(\mathbf{D}f_x(g(y)-g(w))+y-w)\right| \\ &\le \left|\mathbf{D}f_x^{-1}\right|\left|\mathbf{D}f_x(g(y)-g(w))+y-w\right| \\ &= \left|\mathbf{D}f_x^{-1}\right|\left|\mathbf{D}f_x(x-g(w))+f(x)-f(g(w))\right| \\ &= \left|\mathbf{D}f_x^{-1}\right| o(|x-g(w)|),\end{aligned}$$

as desired. □

Remark 1.15. *g is $\mathcal{C}^1$ in a neighborhood of 0. Because $\mathbf{D}f_x$ is continuous in x and a$\mathbf{D}f_0$ is invertible, so is $\mathbf{D}f_x^{-1}$ near 0.*

Now we want to show that, if f is $\mathcal{C}^k$, then g is $\mathcal{C}^k$. To do it, consider the following lemma.

Lemma 1.16. *Let $\mathcal{W}$ be a normed linear space. Let $A : U \subset \mathcal{X} \xrightarrow{\mathcal{C}^k} \mathcal{B}(\mathcal{W},\mathcal{W})$. Assume that for all $p \in \mathcal{X}, A(p)$ is invertible. Then $x \mapsto A^{-1}(p)$ is $\mathcal{C}^k$.*

Proof.

$$\begin{aligned}A^{-1}(x)-A^{-1}(y) &= A^{-1}(x)(I-A(x)A^{-1}(y)) \\ &= A^{-1}(x)\big(A(y)-A(x)\big)A^{-1}(y) \\ &= -A^{-1}(x)\big(A(x)-A(y)\big)A^{-1}(y) \\ &= -A^{-1}(x)\big(A(x)-A(y)-\mathbf{D}A_y(x-y)\big)A^{-1}(y)-A^{-1}(x)\mathbf{D}A_y(x-y)A^{-1}(y)\end{aligned}$$

Stop for a second and consider what the maps are doing.

- $A^{-1}(y) : \mathcal{W} \to \mathcal{W}$.
- $\mathbf{D}A_y : \mathcal{X} \to \mathcal{B}(\mathcal{W},\mathcal{W})$
- $\mathbf{D}A_y(x-y) : \mathcal{W} \to \mathcal{W}$

Now, note that

$$\underbrace{A^{-1}(x)}_{\text{bounded}}\underbrace{\big(A(x)-A(y)-\mathbf{D}A_y(x-y)\big)}_{o(|x-y|)}\underbrace{A^{-1}(y)}_{\text{bounded}}.$$

Thus,

$$A^{-1}(x)-A^{-1}(y) = -A^{-1}(x)\mathbf{D}A_y(x-y)A^{-1}(y)+o(|x-y|)$$

Therefore,

$$A^{-1}(x)-A^{-1}(y)-(-)A^{-1}(x)\mathbf{D}A_y(x-y)A^{-1}(y) = o(|x-y|),$$

and we identify the derivative we are looking for as

$$(\mathbf{D}A^{-1})_y = -A^{-1}(x)\mathbf{D}A_y(x-y)A^{-1}(y)$$

To prove that A^{-1} is $\mathcal{C}^k$, repeat this process $k-1$ more times. □

Definition 1.17. *Let $U \subset \mathcal{X}$ and $V \subset \mathcal{Y}$ with U and V open. Suppose that $f : U \to V$ is $\mathcal{C}^1$ and there is a function $g : V \to U$ such that $f \circ g(y) = x$ and $g \circ f(x) = x$. Then, f is a diffeomorphism and we say that U and V are diffeomorphic.*

Definition 1.18. *It $f : U \subset \mathcal{X} \to \mathcal{Y}$ is $\mathcal{C}^1$ and $\mathbf{D}f_p$ is invertible, then f is a local diffeomorphism.*

Example 1.19. *Let $F(r, \theta) = (r \sin\theta, r\cos\theta)^\top$. Then,*

$$\mathbf{D}F_{(r,\theta)} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}.$$

If $r \neq 0$, then

$$\mathbf{D}f^{-1}_{(r,\theta)} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\frac{1}{r}\sin\theta & \frac{1}{r}\cos\theta \end{bmatrix},$$

which is invertible. So we say that F is a local diffeomorphism.

Example 1.20. *Let*

$$u(x, y) = (x + xy, y + \frac{x^2}{2} + \frac{y^3}{2}).$$

Then,

$$\mathbf{D}u_{(x,y)} = \begin{bmatrix} 1+y & x \\ x & 1-y \end{bmatrix}.$$

At $(0, 0)$, it is a matrix of ones, so u is not a local diffeomorphism there.

Definition 1.21 (Regular Values)**.** *Let $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$. We say that $y \in \mathbb{R}^m$ is a regular value of F if for all $x \in F^{-1}(y), \mathbf{D}F_x$ is surjective.*

Before we continue, it is germane to review some linear algebra. Suppose $A : \mathbb{R}^n \to \mathbb{R}^m$ is surjective. This implies that $m \leq n$. If $n = m$, then we say that f is an *isomorphism.* We can also decompose the domain of A into

$$\mathbb{R}^n = \ker A \oplus \ker A^\perp.$$

No matter what, we can find a partial inverse of A, denoted by $G : \mathbb{R}^m \to \ker A^\perp$, such that

$$AG = I \text{ and } GA = \Pi_{\ker A^\perp}.$$

Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces. Let $\mathcal{M}$ be a metric space. Define the continuous map $\psi : \mathcal{M} \to \mathcal{B}(\mathcal{X}, \mathcal{Y})$. If for some $p \in \mathcal{M}$, $\psi(p)$ is invertible, then there is a whole neighborhood N of p such that for all $b \in N, \psi^{-1} : N \to \mathcal{B}(\mathcal{Y}, \mathcal{X})$ is continuous.

This gives us a corollary regarding the fact that the rank of a continuous linear map can only go up in a neighborhood. Before stating the corollary, recall the definition of the adjoint operator to A.

$$A^* \text{ is the adjoint of } A \iff \langle Ax, y\rangle = \langle x, A^* y\rangle.$$

if $A : \mathbb{R}^n \to \mathbb{R}^m$, then $A^* : \mathbb{R}^m \to \mathbb{R}^n$.

Corollary 1.22. *Define the continuous map $\psi : \mathcal{M} \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$. If $\psi(p)$ has rank r (obviously, $r \leq \min\{m, n\}$), then there is a neighborhood N of p such that for all $b \in N, \psi(b)$ has rank $\geq r$.*

Proof. This is a result of the fact that, for any matrix, we can find a neighborhood such that small perturbations will not decrease the rank. Since ψ is continuous, the result holds. ∎

Definition 1.23 (Alternate Definition of Regular Value)**.** *Let $F : U \subset \mathcal{X} \to \mathcal{Y}$. We say $p \in U$ is a regular value of F if for all $x \in F^{-1}(p)$, we can find a matrix $A \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$ such that $\mathbf{D}F_x A = I$.*

Definition 1.24 (Submanifold)**.** *$\mathcal{M} \subset \mathbb{R}^N$ is a $\mathcal{C}^k$ n-submanifold of $\mathbb{R}^N$ if we can find a $\mathcal{C}^k$ function $F : \mathbb{R}^N \to \mathbb{R}^{N-n}$ and a regular value $p \in \mathbb{R}^{N-n}$ such that $F^{-1}(p) = \mathcal{M}$.*

Definition 1.25 (Graph of a Function)**.** *Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, U open. The graph of a function is $\Gamma_f = \{(x, f(x)) \subset \mathbb{R}^{n+m} \mid x \in U\}$*

Lemma 1.26. *Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$, U open, and f is $\mathcal{C}^k$. Then, Γ_f is a $\mathcal{C}^k$ submanifold of $\mathbb{R}^{n+m}$.*

Proof. We just need to check that Γ_f is the set of regular points for some function. Let F be defined as a function $F : U \times \mathbb{R}^m \to \mathbb{R}^m$. Write coordinates in $U \times \mathbb{R}^m$ as (x, y). Let $F(x, y) = y - f(x)$, so that $\Gamma_g = F^{-1}(0)$. The derivative is

$$\mathbf{D}F_{(x,y)} = (\mathbf{D}_x F_{(x,y)}, \mathbf{D}_y F_{(x,y)}) = (-\mathbf{D}f_x, I),$$

which is rank m, so 0 is a regular value as desired. □

Finally, we are ready to state and prove the Implicit Function Theorem (IFT).

Theorem 1.27 (Implicit Function Theorem (IFT))**.** *Let $f : \mathbb{R}^{n+m} \to \mathbb{R}^m$ be a $\mathcal{C}^k$ function. Let $c \in \mathbb{R}^m$ be a regular value of f. Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, so that we can write $(x, y) \in \mathbb{R}^{n+m}$. Suppose $(x_0, y_0) \in f^{-1}(c)$. Then there are open neighborhoods $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ of x_0 and y_0 such that we can find a $\mathcal{C}^k$ function $h : U \to V$ that solves the system $f(x, h(x)) = c$. for all $x \in U$.*

Proof. Let $K = \ker \mathbf{D}f_{(x_0,y_0)} \simeq \mathbb{R}^n$. Write $x \in K$ and $y \in K^\perp \simeq \mathbb{R}^m$. Define the function $G(x, y) = (x, f(x, y))$. Then,

$$\mathbf{D}G_{(x_0,y_0)} = (\mathbf{D}_x G_{(x_0,y_0)}, \mathbf{D}_y G_{(x_0,y_0)}) = \begin{bmatrix} I & 0 \\ \mathbf{D}_x f_{(x_0,y_0)} & \mathbf{D}_y f_{(x_0,y_0)} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & \mathbf{D}_y f_{(x_0,y_0)} \end{bmatrix}$$

which is rank $n + m$ by hypothesis. Then we can find an inverse function in the neighborhood of (x_0, y_0). Let $G^{-1}(u, v) = (\phi(u, v), \psi(u, v))$. We have that, locally,

$$\begin{aligned} G^{-1}(G(x, y)) &= (x, y) \\ G^{-1}(x, f(x, y)) &= (x, y) \\ (\phi(x, f(x, y)), \psi(x, f(x, y))) &= (x, y). \end{aligned}$$

Therefore,

$$\begin{aligned} (x, y) = G(G^{-1}(x, y)) &= G(\phi(x, y), \psi(x, y)) \\ &= (\phi(x, y), f(\phi(x, y), \psi(x, y))) \end{aligned}$$

thus, $\phi(x, y) = x$ and

$$f(\phi(x, c), \psi(x, c)) = f(x, \psi(x, c)) = c.$$

Let $h(x) = \psi(x, c)$, and this is the diffeomorphism we desired in the statement of the theorem. □

Corollary 1.28. *Let $\mathcal{M} \subset \mathbb{R}^N$ be an embedded n-submanifold. Let $p \in \mathcal{M}$. Then, there is a diffeomorphism $\psi : B_\delta(p) \to \psi(B_\delta(p))$ such that*

$$\psi(\mathcal{M} \cap B_\delta(p)) = \left\{(x, 0) \mid x \in U \overset{\text{open}}{\subset} \mathbb{R}^n\right\} \cap \psi(B_\delta(p)).$$

Theorem 1.29 (Rank Theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^N$ be a $\mathcal{C}^r$ function such that $\mathbf{D}f_x$ is rank m for all x in a neighborhood of x_0. Then, there are open neighborhoods $U, V \subset \mathbb{R}^n$ with $x_0 \in V$ and a $\mathcal{C}^r$ diffeomorphism $h : V \to U$ such that $f \circ h$ only depends on the first m coordinates of x, that is,*

$$(f \circ h)(x_1, \dots, x_m, x_{m+1}, \dots, x_n) = \tilde{f}(x_1, \dots, x_m)$$

for some $\mathcal{C}^r$ function $\tilde{f}$.

Essentially, the rank theorem tells us that a rank m map depends on only m coordinates.

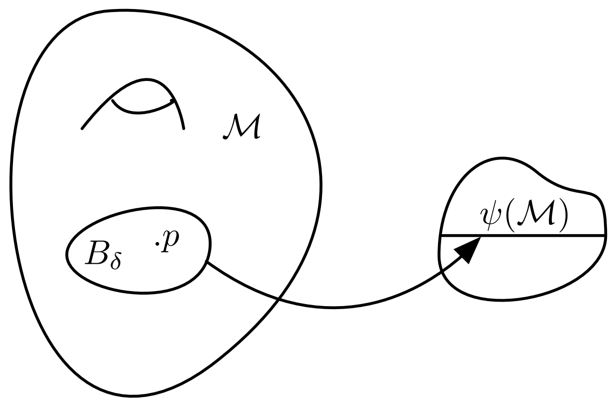

Figure 1: Depicting Corollary 1.28

Proof. Let $K = \ker \mathbf{D} f_q \simeq \mathbb{R}^{n-m}$. Let $M = K^\perp \simeq \mathbb{R}^m$ Then, $\mathbb{R}^n = K \oplus M$. Define $V = \mathrm{Im}\mathbf{D} f_q \simeq \mathbb{R}^m$. Since $V \simeq M$, we can find a map $B : V \to M$ such that $\mathbf{D} f_q \circ B = I$ on all of V. Now extend B to $V^\perp$ by zero, so that B is a projection onto M, that is,

$$B \circ \mathbf{D} f_q = \Pi_M \text{ and } \mathbf{D} f_r \circ B = \Pi_V.$$

Write $x \in K$ and $y \in M$. Now define $g(x, y) = (x, B \circ f(x, y))$. We have that

$$\mathbf{D} g_q = \begin{bmatrix} I & 0 \\ B \circ \mathbf{D}_x f_{(x,y)} & B \circ \mathbf{D}_y f_(x, y) \end{bmatrix} = I_n,$$

so g is locally invertible. Let $g^{-1}(x, y) = (\phi(x, y), \psi(x, y))$. Then

$$\begin{aligned} (x, y) &= g \circ g^{-1}(x, y) \\ (x, y) &= g(\phi(x, y), \psi(x, y)) \\ &= (\phi(x, y), B \circ f(\phi(x, y), \psi(x, y))), \end{aligned}$$

which implies that $\phi(x, y) = x$. Plugging in,

$$y = B \circ f(x, \psi(x, y)) = B \circ f \circ g^{-1}(x, y)$$

We can take the derivative to check x-independence. □

Theorem 1.30 (Domain Straightening)**.** *Let $f : A \overset{\text{open}}{\subset} \mathbb{R}^{n+m} \to \mathbb{R}^m$. Let $x \in A$ s.t. $\mathrm{rk}(\mathbf{D} f_{x_0}) = m$. Then, there is a neighborhood V containing x_0 such that $V \subset A$ and there is $\mathcal{C}^p$ diffeomorphism h s.t.*

$$f \circ h(x_1, \ldots, x_{n+m}) = (x_{n+1}, \ldots, x_{n+m}).$$

and $h : U \overset{\text{open}}{\subset} \mathbb{R}^{n+m} \to V$.

Proof. Let $K = \ker \mathbf{D} f_{x_0}$ Write vectors in $\mathbb{R}^{n+m}$ as (x, y), with $x \in K$ and $y \in K^\perp$. Then, define a function $G(x, y) = (x, f(x, y))$. We have that

$$\mathbf{D} G_{(x,y)} = \begin{bmatrix} I & 0 \\ \mathbf{D}_x f_{(x,y)} & \mathbf{D}_y f_{(x,y)} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & \mathbf{D}_y f_{(x,y)} \end{bmatrix}.$$

Clearly $\mathrm{rk}(\mathbf{D} G_{(x,y)}) = n+m$, so we can apply the inverse function theorem. Let $G^{-1}(x, y) = (\phi(x, y), \psi(x, y))$ so that

$$\begin{aligned} (x, y) &= G(\phi(x, y), \psi(x, y)) \\ &= (\phi(x, y), f(\phi(x, y), \psi(x, y))), \end{aligned}$$

and we see that $\phi(x,y) = x$. Therefore, $G^{-1}(x,y) = (x, \psi(x,y))$. thus

$$(x,y) = (x, f \circ G^{-1}(x,y)).$$

Now just set $h = G^{-1}$, and the result follows. $\square$

Now that we can straighten domains, we ought to be able to straighten ranges. Lets try.

Theorem 1.31 (Range Straightening)**.** *Let $f : A \overset{\text{open}}{\subset} \mathbb{R}^m \to \mathbb{R}^n$ for $n \geq m$ be a $\mathcal{C}^r$ map. For some $p \in A$, assume that* $\operatorname{rk}(\mathbf{D}f_p) = m$*. Then, there is a $\mathcal{C}^r$ diffeomorphism g such that*

$$g \circ f(x_1, \ldots, x_m) = (x_1, \ldots, x_m, 0, \ldots, 0).$$

Proof. Let $V = \operatorname{Im}\mathbf{D}f_p \simeq \mathbb{R}^m$. Let $Q = V^\perp \simeq \mathbb{R}^{n-m}$. As usual, write vectors in $\mathbb{R}^n$ as $x \in V$ and $y \in Q$. Define the map $G(x,y) = f(x) + (0,y)$. We have that

$$\mathbf{D}G_{(x,y)} = \begin{bmatrix} \mathbf{D}_x f_{(x,y)} & 0 \\ 0 & I \end{bmatrix}$$

Thus, G is invertible. Then, note that

$$\begin{aligned} (x,y) &= G^{-1}(G(x,y)) \\ &= G^{-1}(f(x) + (0,y)). \end{aligned}$$

Then, $(x,0) = G^{-1} \circ f(x)$. Now just set $g = G^{-1}$, and the result follows. $\square$

Lemma 1.32 (Morse Lemma)**.** *Let $f : A \overset{\text{open}}{\subset} \mathbb{R}^n \to \mathbb{R}$. Let $p \in A$ s.t. $\mathbf{D}f_p = 0$ but $\mathbf{D}^2 f_p$ is nonsingular. Then, there is a local diffeomorphism g and an integer s such that*

$$f \circ g(x) = -x_1^2 - \cdots - x_s^2 + x_{s+1}^2 + \cdots x_n^2 + f(p).$$

We call s the index of the function f at the critical point p.

Proof. First of all, we can make a trivial change of coordinates so that $p = 0$ and $f(0) = 0$. Then, consider that

$$\begin{aligned} f(x) &= \int_0^1 \nabla f(tx) \cdot x dt \\ &= \sum_{i=1}^n x_i g_i(x), \end{aligned}$$

where $g_i(x)$ is the integral of the i-th component of the gradient, i.e.,

$$g_i(x) = \int_0^1 \frac{\partial f}{\partial x_i}\Big|_{tx} dt.$$

We know that $g_i(0) = 0$ from the hypothesis. Then we can do this again, so that

$$f(x) = \sum_{j=1}^n x_i x_j h_{ij}(x).$$

Now, let

$$H_{ij}(x) = \frac{h_{ij} + h_{ji}}{2},$$

and note that $H_{ij}(x) = H_{ji}(x)$ and

$$f(x) = \sum_{j=1}^n x_i x_j H_{ij}(x).$$

Now, if we arrange the terms $H_{ij}(x)$ into a matrix, it will be symmetric. Thus we can diagonalize it. Now we choose a basis of $\mathbb{R}^n$ so that $H_{ij}(0)$ is diagonal. Therefore,

$$H_{ij}(x) = \delta_{ij}\lambda_i + o(|x|).$$

Note also, by the definition of H_{ij},

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(0) = 2H_{ij}(0) = 2\delta_{ij}\lambda_i.$$

Since we assumed that $\mathbf{D}^2 f_0$ was nonsingular, $\lambda_i \neq 0$ for all $i = 1, \ldots, n$. Thus, (adding and subtracting the same thing)

$$f(x) = \underbrace{\lambda_1 x_1^2 + \cdots + \lambda_n x_n^2}_{\text{this is equal to } \sum_{i,j} H_{ij}(0)} + \sum_{i,j} x_j x_j (H_{ij}(x) - H_{ij}(0))$$

Set

$$v_1 = |H_{11}(x)|^{1/2} \left(x_1 + \sum_{i>1} x_i \frac{H_{i1}(x)}{H_{11}(x)} \right)$$

and $v_i = x_i$ for all $i > 1$. Our claim is that $x \mapsto v$ is a local diffeomorphism. To verify this, note that, at 0,

$$\mathbf{D}v_0 = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & I \end{bmatrix},$$

because

$$\sum_{i>1} x_i \frac{H_{i1}(x)}{H_{11}(x)} = o(|x|).$$

Now, note the following calculation, which "completes the square,"

$$v_1^2 = |H_{11}(x)| \Big(\underbrace{x_1^2}_{o(|x|^2)} + \underbrace{2x_1 \sum_{i>1} x_i \frac{H_{i1}(x)}{H_{11}(x)}}_{o(|x|^3)} + \underbrace{\left(\sum_{i>1} x_1 x_i \frac{H_{i1}(x)}{H_{11}(x)} \right)^2}_{o(|x|^4)} \Big)$$

$$\operatorname{sign}(\lambda_1) v_1^2 = H_{11}(x) x_1^2 + 2x_1 \sum_{i>1} x_i \frac{H_{i1}(x)}{H_{11}(x)} + o(|x|^4).$$

Now, we can use this expression to reconsider our original function f. Recalling that H is symmetric, we have that

$$\begin{aligned} f(x) &= \sum_{ij} x_i x_j H_{ij}(x) \\ &= x_1^2 H_{11}(x) + 2\sum_{i>1} x_1 x_i H_{i1}(x) + \sum_{i,j>1} x_i x_j H_{ij}(x) \\ &= \operatorname{sign}(\lambda_1) v_1^2 + \sum_{i,j>1} x_i x_j \tilde{H}_{ij}(x), \end{aligned}$$

where the term $\tilde{H}_{ij}(x)$ absorbs terms that are $o(|x|^4)$. Now we proceed iteratively, doing a similar change of coordinates as above. Suppose we have done this $r-1$ times, so that

$$f(x) = \operatorname{sign}(\lambda_1) v_1^2 + \cdots + \operatorname{sign}(\lambda_{r-1}) v_{r-1}^2 + \sum_{i,j>r} \Theta_{ij}(v, x)$$

where $\Theta_{ij} = \lambda_i \delta_{ij} + o(|v, x|)$. Now define

$$v_r = |\Theta_{rr}|^{1/2} \left(x_r + \sum i > r x_i \frac{\Theta_{ir}(v,x)}{\Theta_{rr}(v,x)} \right).$$

Omitting some calculations,

$$f(x) = \sum_{j=1}^{r} \operatorname{sign}(\lambda_j) v_j^2 + \sum_{i,j>r} x_i x_j \tilde{\Theta}_{ij}(x, v),$$

with $\tilde{\Theta}_{ij}(x, v) = \lambda_i \delta_{ij} + o(|x, v|)$. Repeating the process n times, we get that $f(x) = \sum_{j=1}^{n} \operatorname{sign}(\lambda_j) v_j^2$. $\square$

I have stopped writing proofs for lack of time

Theorem 1.33 (Lagrange Multipliers). *Let $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$. Let c be a regular value of F. Let $f : U \overset{\text{open}}{\subset} \mathbb{R}^n \to \mathbb{R}$. Let $S = F^{-1}(c)$, and suppose $f|_S$ has a local max/min at p. Then,*

$$\nabla f(p) \in \operatorname{span}\{\nabla F_1(p), \ldots, \nabla F_m(p)\}$$

2 Integration

Definition 2.1 (Partition). *A partition P of a box $B \subset \mathbb{R}^n$ is a set of sets of points $P_j = \left\{x_1^j, \ldots, x_{m_j}^j\right\}$ where $m_j > 1$ for $j = 1, \ldots, n$ such that*

$$B = \bigcup_{2 \le i \le m_j} \prod_{j=1}^{n} [x_{i-1}^j, x_i^j]$$

Let $B(P)$ denote the set of boxes defined by the partition P on the box B.

Definition 2.2 (Lower and Upper Sum). *Let $f : B \to \mathbb{R}$, and B be a box. Define the following functions:*

$$L(f, P) = \sum_{R \in B(P)} \inf_{x \in R} f(x) v(R)$$
$$U(f, P) = \sum_{R \in B(P)} \sup_{x \in R} f(x) v(R)$$

Definition 2.3 (Refinement). *We say that P' is a refinement of P if $P \subset P'$.*

Proposition 2.4. *Let $f : B \to \mathbb{R}$. Then, $L(f, P) \le U(f, P)$.*

Definition 2.5 (Riemann Integrable). *Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$. Let $B \supset A$ be a box, and let P be a partition of B. If*

$$\sup_P L(f, P) = \inf_P U(f, P),$$

when we say that f is Riemann integrable (RI), and we define $\int_A f$ to be the limit.

Proposition 2.6. *Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$. Let $B \supset A$ be a box. Extend f to B by zero. Then,*

$$\int_A f = \int_B f,$$

and f is RI on $A \iff f$ is RI on B.

Theorem 2.7 (Darboux's Theorem). *Let $f : A \supset \mathbb{R}^n \to \mathbb{R}$. Assume A is bounded. If f is RI, let I equal the integral of f, and then for any $\epsilon > 0$, there is a partition P such that every $R \in B(P)$ has all side lengths less than δ and there is a set $\{x_R\}_{R \in B(P)}$ with each $x_R \in R$ such that*

$$\left| \sum_{R \in B(P)} f(x_r) v(R) - I \right| \le \epsilon.$$

Let A be a set. The characteristic function of A is denoted as

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{else} \end{cases}.$$

Definition 2.8 (Volume)**.** *Let $A \subset \mathbb{R}^n$ be bounded. We say A has volume if χ_A is RI. We denote the volume by*

$$v(A) = \int_A \chi_A$$

Definition 2.9 (Measure Zero)**.** *We say that a set has measure zero is for any $\epsilon > 0$, we can fund a countable covering of A by sets $\{S_i\}_{i\in\mathbb{N}}$ such that $\sum_{i\in\mathbb{N}} v(S_i) \le \epsilon$.*

Lemma 2.10. *If A has volume zero, the A has measure zero.*

Proposition 2.11. *A countable union of measure zero sets is measure zero.*

Remark 2.12. *If $A \subset B$ and B is measure zero, then A is measure zero.*

Definition 2.13 (Oscillation)**.** *Let $h : \mathbb{R}^n \to \mathbb{R}$. The oscillation of h at a point x_0 is equal to*

$$\operatorname{os}(h, x_0) = \inf_{U \ni x_o} \left\{ \sup_{x,y\in U} |h(x) - h(y)| \mid U \text{ is open} \right\}.$$

Theorem 2.14 (Lebesgue's Theorem)**.** *Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$, A bounded. Let f_A denote the extension of f to $\mathbb{R}^n$ by zero. Then, f is RI on A iff the points at which f_A is discontinuous form a set of measure zero.*

Corollary 2.15. *Let $A \subset \mathbb{R}^n$ be bounded. Then A has volume iff ∂A has measure zero.*

2.1 Improper Integrals

Definition 2.16. *Let $f : A \subset \mathbb{R}^n$ be bounded and assume $f \ge 0$. Let f_A denote the extension of f to all of $\mathbb{R}^n$ by zero. Then, consider the a-cube $[-a, a]^n \subset \mathbb{R}^n$. Assuming the following limit exists, define*

$$\int_A f \triangleq \lim_{a\to\infty} \int_{[-a,a]} f_A.$$

The remaining discussion on Integrals is in my handwritten notes. I will type it here later if I have time.

3 Calculus on Curved Spaces (Multilinear Algebra)

For all of this section, V is an n dimensional vector space.

Definition 3.1 (Dual Space)**.** *We denote by V^* the dual of V. It is defined as*

$$V^* \triangleq \{\textit{linear maps} : V \to \mathbb{R}\} = \operatorname{Hom}(V, \mathbb{R}).$$

Proposition 3.2. *If $(V, \langle , \rangle)$ is an inner product space, then $V \simeq V^*$ ($\simeq$ denotes an isomorphism).*

Proof. Let $\phi \in V^*$ and $\{e_1, \dots, e_n\}$ be an orthonormal basis for V. Thus, for $w \in V$, we may write $w = \sum_{i=1}^n w_i e_i$. By linearity, $\phi(w) = \sum_{i=1}^n w_i \phi(e_i) = \langle w, \sum_{i=1}^n \phi(e_i) e_i \rangle$. Thus, $\phi \mapsto V_\phi \triangleq \sum_i \phi(e_i) e_i$ defines an isomorphism $V^* \to V$. □

Remark 3.3 (Dual Basis)**.** *Observe that if $\{b_1, \dots, b_n\}$ is a basis for V, then it determines a unique dual basis $\{\omega^1, \dots, \omega^n\}$ for V^* by the description*

$$\omega^i(b_j) = \delta_i^j.$$

Definition 3.4 (k-linear)**.** *A function $\phi : V^k \to \mathbb{R}$ is called k-linear if it is linear in each argument.*

Example 3.5 (Bilinear)**.** *A function $\phi : V \times V \to \mathbb{R}$ is called bilinear if*

$$\phi(\alpha v + \beta w, u) = \alpha\phi(v, u) + \beta\phi(w, u) \quad \textit{and}$$
$$\phi(v, \alpha u + \beta w) = \alpha\phi(v, u) + \beta\phi(v, w).$$

Definition 3.6 (Alternating Linear Map)**.** *A k-linear map $\phi : V^k \to \mathbb{R}$ is called alternating if*

$$\phi(v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots, v_{j-1}, v_j, v_{j+1}, \dots, v_n) = -\phi(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots, v_n)$$

for any $1 \le i < j \le n$.

Example 3.7 (Alternating Bilinear Map)**.** *$\phi(v, w) = -\phi(w, v)$ for all $v, w \in V$.*

Example 3.8 (Alternating Trilinear Map)**.** *$\phi(u, v, w) = -\phi(u, w, v) = \phi(w, u, v)$ for all $u, v, w \in V$.*

Definition 3.9 (Permutation on k Letters)**.** *A map $\sigma : \{1, \dots, k\} \to \{1, \dots, k\}$ is called a permutation on k letters if is is bijective. We define* $\mathrm{sign}(\sigma)$ *as $\phi(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = \mathrm{sign}(\sigma)\phi(v_1, \dots, v_k)$. if ϕ is an alternating k linear map.*

Let $\bigwedge^k V^*$ denote the alternating k linear maps from $V^k \to \mathbb{R}$.

Proposition 3.10. *If $m > n$, then all maps in $\bigwedge^m V^*$ are identically zero.*

Proof. Suppose $m = n + 1$. Rather than proving it, we offer a Lemma, the result of which implies the result of this proposition immediately.

Lemma 3.11. *Let $\phi \in \bigwedge^k V^*$, and $w_1, \dots, w_k$ a linearly dependent set of vectors. Then, $\phi(w_1, \dots, w_k) = 0$.*

Proof. First consider the simple case that $w_k = w_{k-1}$. Then, $\phi(w_1, \dots, w_{k-1}, w_{k-1}) = -\phi(w_1, \dots, w_{k-1}, w_{k-1})$. The only number that equals its opposite is zero. Thus, $\phi(w_1, \dots, w_k) = 0$. Now, consider the more general case. Since $w_1, \dots, w_k$ is a linearly dependent set of vectors, there is a set $\alpha_1, \dots, \alpha_n$ that are not identically zero and $w_k = \sum_{i=1}^{k} \alpha_i w_i = 0$ Assume w.l.o.g. $\alpha_k \neq 0$. Then, we have that

$$\begin{aligned}
\phi(w_1, \dots, w_k) &= \phi\left(w_1, \dots, \sum_{i=1}^{k-1} \frac{\alpha_i}{\alpha_n} w_i\right) \\
&= \sum_{i=1}^{k-1} \frac{\alpha_i}{\alpha_n} \phi\left(w_1, \dots, w_i\right) \\
&= \sum_{i=1}^{k-1} \frac{\alpha_i}{\alpha_n} 0 = 0,
\end{aligned}$$

by linearity and the result of the first part of the proof with $w_{k-1} = w_k$. ☐

The result of Lemma 3.11 proves the Proposition 3.10 ☐

Proposition 3.12. *$\bigwedge^n V^* \simeq \mathbb{R}$ and if we fix a basis $\{e_1, \dots, e_n\}$ for V then $\bigwedge^n V^*$ has basis* $\det()$.

Proof. We already know $\det()$ to be an alternating n-linear map. Since these sorts of elements populate $\bigwedge^n V^*$, we only need to prove that $\bigwedge^n V^*$ is 1-dimensional. To see this, let $\{b_1, \dots, b_n\}$ be a basis for V. Consider $\phi \in \bigwedge^n V^*$ operating on a set of n vectors $\{v_\alpha\}_{\alpha=1}^n$. For each α, we can always write

$$v_\alpha = \sum_{i=1}^{n} v_{\alpha i} b_i.$$

Then,

$$\phi(v_1, \dots, v_n) = \sum_{j=1}^{n} v_{1j_1} \cdots v_{nj_n} \phi(b_{j_1}, \dots, b_{j_n}).$$

So, ϕ is determined up to scalar multiplication by $\phi(b_{j_1}, \dots, b_{j_n})$, which is itself equal (up to a sign) to ϕ of any other ordering of the basis vectors, this its essentially one scalar, and thus $\bigwedge^n V^*$ is 1 dimensional. ☐

In class, I asked about the multi indices in the above proof, and I received the following justification for why they are necessary and why the summation above makes sense.

Example 3.13 (n=2)**.** *Let $\{b_1, b_2\}$ be a basis for $V = \mathbb{R}^2$. We have that*

$$\begin{aligned}\phi(v_1, v_2) &= \phi(v_{11}b_1 + v_{12}b_2, v_{21}b_1, v_{22}b_2)\\ &= v_{11}v_{21}\underbrace{\phi(b_1, b_1)}_{=0} + v_{11}v_{22}\phi(b_1, b_2) + v_{12}v_{21}\phi(b_2, b_1) + v_{12}v_{22}\underbrace{\phi(b_2, b_2)}_{=0}\\ &= v_{11}v_{22}\phi(b_1, b_2) + v_{12}v_{21}\phi(b_2, b_1)\\ &= (v_{11}v_{22} - v_{12}v_{21})\phi(b_1, b_2)\end{aligned}$$

My question was about the multi indices $j_1, \ldots, j_n$. The example shows us that, for $j = 1$ and $n = 2$ as in the example, $1_1 = 1$ and $1_2 = 2$ and for $j = 2$, $2_1 = 2$ and $2_2 = 1$. Also, note that if the basis is normalized, we seem to have recovered the determinant of a 2×2 matrix, as predicted by Proposition 3.12.

Proposition 3.14. $\dim\left(\bigwedge^k V^*\right) = \begin{pmatrix} n \\ k \end{pmatrix}$.

Proof. By a similar argument present in the proof of Proposition 3.12, we know that any $\phi \in \bigwedge^k V^*$ is determined up to scalar multiplication by $\phi(b_{i_1}, \ldots, b_{i_k})$. Therefore, the question reduces to: how many ways are there to choose k basis vectors from the original n that form a basis for V? The answer of course is the binomial. $\square$

Let $w \in V$. Define $i_w : \bigwedge^k V^* \to \bigwedge^{k-1} V^*$ by

$$(i_w\phi)(v_1, \ldots, v_{k-1}) = \phi(w, v_1, \ldots, v_{k-1}).$$

Note that for $\phi \in \bigwedge^k V^*$, we have that $(i_w\phi) \in \bigwedge^{k-1} V^*$ as desired. Essentially, i_w pulls w out of the argument of ϕ.

Question 1. *So does i_w just take away the first argument? If so then what if that argument isn't w? Does it somehow effect the other arguments by taking components that are parallel to w?*

Because ϕ depends only on the basis elements, note that there are only n independent maps $i_{b_j} : \bigwedge^n V^* \to \bigwedge^{n-1} V^*$. Thus, $\bigwedge^{n-1} V^*$ has basis $\{i_{b_1} \det, \ldots, i_{b_n} \det\}$ and there are $\begin{pmatrix} n \\ 2 \end{pmatrix}$ maps $\bigwedge^n V^* \to \bigwedge^{n-2} V^*$ defined by

$$i_{b_j}(i_{b_k} \det),$$

where $j < k$, since once we include the case of $j < k$, $j = k$ sets the function to zero and $k > j$ just has opposite sign w.r.t. $j < k$. In this set, det is a polynomial function; it can be multiplied by compositions of functions $i_{b_{j_1}} \circ i_{b_{j_k}}$ to form bases for the spaces $\bigwedge^{n-1} V^*, \ldots, \bigwedge^{n-k} V^*$.

3.1 Wedge Products

Let $\{b_1, \ldots, b_n\}$ be a basis for V. Recall Remark 3.3, which tells us how to find a dual basis $\left\{\omega^1, \ldots, \omega^n\right\}$ for V^* given the basis for V. The wedge product of two dual basis elements is an alternating bilinear map $(\omega^i \wedge \omega^j) \in \bigwedge^2 V^*$, that is,

$$\begin{aligned}(\omega^i \wedge \omega^j)(u, v) &= (\omega^i \wedge \omega^j)\left(\sum_{k=1}^n u_k b_k, \sum_{\ell=1}^n v_\ell b_\ell\right)\\ &= (\omega) \wedge \omega^j)\,(u_i b_i + u_j b_j, v_i b_i + v_j b_j)\\ &= u_i(\omega^i \wedge \omega^j)\,(b_i, v_i b_i + v_j b_j) + u_j(\omega^i \wedge \omega^j)\,(b_j, v_i b_i + v_j b_j)\\ &= u_i\left(v_i(\omega^i \wedge \omega^j)\,(b_i, b_i) + v_j(\omega^i \wedge \omega^j)\,(b_i, b_j)\right) + u_j(\omega^i \wedge \omega^j)\,(b_j, v_i b_i + v_j b_j)\\ &= u_i\left(v_i(\omega^i \wedge \omega^j)\,(b_i, b_i) + v_j(\omega^i \wedge \omega^j)\,(b_i, b_j)\right) + u_j\left(v_i(\omega^i \wedge \omega^j)\,(b_j, b_i) + (v_j(\omega^i \wedge \omega^j)\,(b_j, b_j)\right)\\ &= (u_i v_j - u_j v_i)(\omega^i \wedge \omega^j)\,(b_i, b_j)\end{aligned}$$

Hereafter, let I be an index set $I = \{i_1, \ldots, i_k\}$. To understand the notation a bit better, note that $\bigwedge^1 V^* = V^*$, and $\omega^1, \ldots, \omega^n$ is a basis for V^* and a basis for $\bigwedge^2 v^*$ is $\left\{\omega^i \wedge \omega^j \mid i < j\right\}$.

3.1.1 A Tensor Perspective on Wedge Products

We can also consider the tensor perspective. Let $V^{*\otimes 2}$ denote the k-linear maps $V^2 \to \mathbb{R}$. As an example for $k = 2$, $V^{*\otimes 2} = V^* \otimes V^*$ denotes the bilinear maps $V \times V \to \mathbb{R}$. Then, for $\phi, \psi \in V^*$, we define their tensor product as $\phi \otimes \psi(v, w) = \phi(v)\psi(w)$. Thus, $\left\{\omega^i \otimes \omega^j\right\}_{i,j=1}^n$ is a basis for $V^* \otimes V^*$. Then, we can define the wedge product using the tensor notation as

$$\omega^i \wedge \omega^j = (\omega^i \otimes \omega^j - \omega^j \otimes \omega^i),$$

which just recovers the alternating property of wedge products that we expect. We can similarly define bigger wedge products, but in order to get the alternating property (as we have done in the $k = 2$ case), the expressions become unwieldy. In particular,

$$\omega^{i_1} \wedge \cdots \wedge \omega^{i_k} = \sum_{\sigma \in \{\text{permutations}\}} \operatorname{sign}(\sigma) \omega^{\sigma_{i_1}} \otimes \omega^{\sigma_{i_n}}.$$

3.2 Tangent Spaces

Let $M \subset \mathbb{R}^n$ be a k-dimensional submanifold, i.e., $M = \phi^{-1}(0)$ for some $\phi : U \overset{\text{open}}{\subset} \mathbb{R}^n \to \mathbb{R}^{n-k}$ such that 0 is a regular value. Note that for any $p \in M$, there is a $\delta > 0$ s.t. $B(p, \delta) \cap M = \Gamma_F$ for some function $F : V \overset{\text{open}}{\subset} \mathbb{R}^k \to \mathbb{R}^{n-k}$, there is a map $X : V \overset{\text{open}}{\subset} \mathbb{R}^k \to B(p, \delta) \subset \mathbb{R}^n$ s.t. $X(x) = (x, F(x))$ and $X(0) = p$. Fig. 2 shows a picture of the above manifold and related functions.

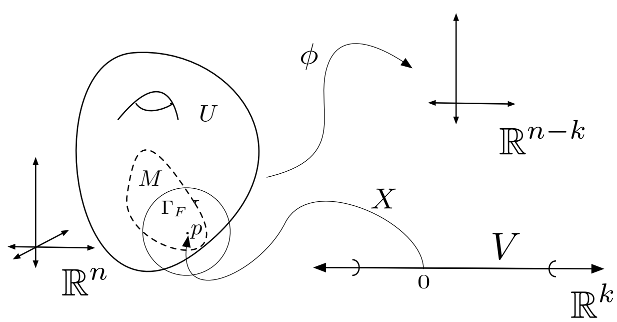

Figure 2: Depicting the manifold M and the functions that it implies.

Definition 3.15 (Tangent Space (first definition)). *Using the functions defined above, the tangent space to a manifold M at the point p is defined as*

$$T_pM = \operatorname{Im}(\mathbf{D}X_0).$$

We note that $X : V \overset{\text{open}}{\subset} \mathbb{R}^k \to \mathbb{R}^n$, so $\mathbf{D}X_0 : \mathbb{R}^k \to \mathbb{R}^n$. Note also that, clearly $\mathbf{D}X_0$ is full rank. Note that $\phi(M) = \{0\}$. Then, $\mathbf{D}\phi \circ \mathbf{D}X = 0$. Note that $\mathbf{D}\phi$ has rank $n - k$ on $\operatorname{Im}X$, so $\ker \mathbf{D}\phi$ has dimension k. Thus $\operatorname{Im}\mathbf{D}X = \ker \mathbf{D}\phi$, giving us an alternate definition of the tangent space.

Definition 3.16 (Tangent Space (second definition)). *Using the functions defined above, the tangent space to a manifold M at the point p is defined as*

$$T_pM = \ker(\mathbf{D}\phi_p).$$

Remark 3.17. *I view the manifold M as a level set for the function ϕ. This makes sense in Definition 3.16, because it is saying that the tangent space is the set of directions we can move at p that do not change the value of the function ϕ, which is the exact intuition for a level set.*

3.3 Alternate Perspective on Vectors

Definition 3.18 (Derivation). *Let $p \in \mathbb{R}^n$. A derivation X_p at p is a collection of $\mathbb{R}$-linear maps from every open neighborhood U of p such that $X_p : C^\infty(U) \to \mathbb{R}$ that satisfy*

1. $X_p(sf + tg) = sX_p(f) + tX_p(g) \quad \forall s, t \in \mathbb{R},$ *and* $f, g \in C^\infty(U)$.
2. $X_p(fg) = f(p)X_p(g) + g(p)X_p(f)$ *(Liebniz)*
3. *If* $p \in V \subset U, f \in C^\infty(U)$, *then* $X_p(f) = X_p(f|_V)$.

Consider some function $f \in C^\infty(U), p \in U$. Taylor's theorem tells us that

$$f(x) = f(p) + \sum_{i=1}^{n}(x^i - p^i)\frac{\partial f}{\partial x^i}(p) + \sum_{i,j=1}^{n}(x^i - p^j)(x^j - p^i)g_{ij}(x),$$

where the functions g_{ij} are similar to the functions used in our proof Lemma 1.32. Then, using the definition of derivations, we note that

$$X_p(f) = \sum_{i=1}^{n}\frac{\partial f}{\partial x^i}(p)X_p(x^i - p^i),$$

since by definition $X_p(\text{constant}) \equiv 0$ and

$$X_p\left(\sum_{i,j=1}^{n}(x^i - p^j)(x^j - p^i)g_{ij}(x)\right) = 0.$$

The above equality holds because all terms vanish. Therefore, *all derivations* can be written as maps

$$f \mapsto \sum_{i=1}^{n} a_i\left(\frac{\partial}{\partial x^i}f\right)(p).$$

Recall that by fixing a basis $\{e_1, \dots, e_n\}$ for $\mathbb{R}^n$, we get a set of coordinates $\left\{x^1, \dots, x^n\right\}$ for $\mathbb{R}^n$. Thus, there is an isomorphism from the set of derivations at p to $\mathbb{R}^n$, given by

$$a_1\frac{\partial}{\partial x^1}\bigg|_p, \dots, a_n\frac{\partial}{\partial x^n}\bigg|_p \mapsto a_1e_1, \dots, a_ne_n.$$

$T_p\mathbb{R}^n$ is simply the set of derivations at $p \in \mathbb{R}^n$. A basis for $T_p\mathbb{R}^n$ is $\left\{\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}\right\} \sim \{e_1, \dots, e_n\}$.

3.4 The Chain Rule

Given

$$X_p = a_1\frac{\partial}{\partial x^1}\bigg|_p + \dots + a_n\frac{\partial}{\partial x^n}\bigg|_p, \text{ let}$$
$$\tilde{X}_p = a_1e_1 + \dots + a_ne_n$$

and observe that $X_pf = \mathbf{D}f_p\tilde{X}_p$ since $\tilde{X}_p = (a_1, \dots, a_n)^\top$ and $\mathbf{D}f_p$ is $\left(\frac{\partial f}{\partial x^1}\Big|_p, \dots, \frac{\partial f}{\partial x^n}\Big|_p\right)$.

Proposition 3.19. *Let $h : \mathbb{R}^n \to \mathbb{R}^m, p \in \mathbb{R}^n$, Then, define $\mathbf{D}h_p : T_p\mathbb{R}^n \to T_{h(p)}\mathbb{R}^m$ as follows. Let ϕ be a smooth function defined in a neighborhood of $h(p)$. Then,*

$$\mathbf{D}h_pX_p\phi \triangleq X_p(\phi \circ h).$$

The following calculation confirms that our notion of the chain rule still applies in the new notation.

$$\begin{aligned} X_p(\phi \circ h) &= \mathbf{D}\left(\phi \circ h\right)_p \tilde{X}_p \\ &= \mathbf{D}\phi_{h(p)} \circ \mathbf{D}h_p \tilde{X}_p \\ &= \mathbf{D}\phi_{h(p)} \mathbf{D}h_p \tilde{X}_p \\ &= \mathbf{D}\phi_{h(p)} X_p h \end{aligned}$$

Note that we can write $h(x) = (y^1(x), \ldots, y^m(x))$, as coordinates in $\mathbb{R}^m$. Then,

$$\begin{aligned} \mathbf{D}(f \circ h)_p &= \left(\left.\frac{\partial (f \circ h)}{\partial x^1}\right|_p, \ldots, \left.\frac{\partial (f \circ h)}{\partial x^n}\right|_p \right) \\ &= \left(\sum_{k=1}^m \left.\frac{\partial f}{\partial y^k}\right|_{h(p)} \left.\frac{\partial y^k}{\partial x^1}\right|_p, \ldots, \sum_{k=1}^m \left.\frac{\partial f}{\partial y^k}\right|_{h(p)} \left.\frac{\partial y^k}{\partial x^n}\right|_p \right) \end{aligned}$$

Example 3.20. *What is* $\mathbf{D}h_p \left(\left.\frac{\partial}{\partial x^i}\right|_p \right)$? *How does it act on a function?*

We have that

$$\begin{aligned} \mathbf{D}h_p \left(\left.\frac{\partial}{\partial x^i}\right|_p \right) f &= \left.\frac{\partial (f \circ h)}{\partial x^i}\right|_p \\ &= \mathbf{D}(f \circ h)_p e_i \\ &= \left(\sum_{k=1}^m \frac{\partial f}{\partial y^k} \frac{\partial y^k}{\partial x^1}, \ldots, \sum_{k=1}^m \frac{\partial f}{\partial y^k} \frac{\partial y^k}{\partial x^n} \right) e_i \\ &= \sum_{k=1}^m \frac{\partial f}{\partial y^k} \frac{\partial y^k}{\partial x^i}. \end{aligned}$$

Now we can see that no matter what we chose for f, the operator is the same:

$$\mathbf{D}h_p \left(\frac{\partial}{\partial x^i} \right) = \sum_{k=1}^m \frac{\partial y^k}{\partial x^i} \frac{\partial}{\partial y^k}.$$

where we omit the reference to the fact that we are evaluating the partial derivatives on the right at the points p and $h(p)$. They are present throughout anytime we are looking at the space $T_p\mathbb{R}^n$. Consider now that

$$\begin{aligned} \frac{\partial}{\partial x^i} x^j &= \delta_i^j \\ \mathbf{D}x^j \left(\frac{\partial}{\partial x^i} \right) &= \delta_i^j, \end{aligned}$$

so that

$$\left(\mathbf{D}x^1, \ldots, \mathbf{D}x^n \right) \text{ is a basis for } \left(T_p\mathbb{R}^n \right)^*.$$

As convention, we write $\left(T_p\mathbb{R}^n\right)^* = T_p^*\mathbb{R}^n$ and we use lower case d as the derivative, so the above statement is the same as

$$\left(dx^1, \ldots, dx^n \right) \text{ is a dual basis for } T_p^*\mathbb{R}^n.$$

Then, $\left\{ dx^I \mid |I| = p, i_1 < \cdots < i_p \right\}$ forms a basis for $\bigwedge^p T_q^*\mathbb{R}^n$, where recall the notation that

$$dx^I = dx^{i_1} \wedge \cdots \wedge dx^{i_p}.$$

3.5 Differential Forms

Definition 3.21 (p-form)**.** *Consider*

$$f(x) = \sum_{|I|=p} f_I(x) dx^I,$$

where $f_I \in C^k$. Then, f is called a C^k p-form.

Example 3.22 ($p = 2$, $n = 3$)**.** *For this, we have that $dx^1 \wedge dx^2$, $dx^2 \wedge dx^3$, and $dx^1 \wedge dx^3$ form a basis for $\bigwedge^2 T_q^*\mathbb{R}^3$.*

Now let V and W be vector spaces. Let $\phi : W \to V$ be a linear map. Then there is a dual map $\phi^* : V^* \to W^*$ so that for $\psi \in V^*$, $u \in W$,

$$\phi^*\psi(w) \triangleq \psi(\phi(u)),$$

that is, $\phi^*\psi = \psi \circ \phi$.

Definition 3.23 (Pullback)**.** *Let $h : \mathbb{R}^n \to \mathbb{R}^m$ be C^k. Let $h^* : \bigwedge^p T_{h(q)}^*\mathbb{R}^m \to \bigwedge^p T_q^*\mathbb{R}^n$ be given as follows: For $\psi \in \bigwedge^p T_{h(q)}^*\mathbb{R}^m$ and $X_1, \ldots, X_p \in T_q\mathbb{R}^n$,*

$$(h^*\psi)(X_1, \ldots, X_p) = \psi\left(\mathbf{D}h(X_1), \ldots, \mathbf{D}h(X_p)\right).$$

Example 3.24 ($p = 1$)**.** *We have that*

$$\begin{aligned}
\left(h^* dy^i\right)\left(\frac{\partial}{\partial x^j}\right) &= dy^i\left(\mathbf{D}h\left(\frac{\partial}{\partial x^j}\right)\right) \\
&= dy^i\left(\sum_{k=1}^m \frac{\partial y^k}{\partial x^j}\frac{\partial}{\partial y^k}\right) \\
&= \frac{\partial y^i}{\partial x^j}.
\end{aligned}$$

Then, noting that the above calculation was just acting on the basis element $\frac{\partial}{\partial x^j}$, we see that, generally, h^* of the 1-form dy^i is just

$$h^* dy^i = \sum_{j=1}^n \frac{\partial y^i}{\partial x^j} dx^j.$$

Example 3.25. *Let $h(r, \theta) = (r\cos(\theta), r\sin(\theta))$. Compute $h^*(dx \wedge dy)$.*

It is a fact that $h^*(dx \wedge dy) = h^*(dx) \wedge h^*(dy)$. So, the first step is to compute each of these objects.

$$\begin{aligned}
h^*(dx) &= \frac{\partial x}{\partial r} dr + \frac{\partial x}{\partial \theta} d\theta \\
&= \cos\theta dr - r\sin\theta d\theta
\end{aligned}$$

and

$$\begin{aligned}
h^*(dy) &= \frac{\partial y}{\partial r} dr + \frac{\partial y}{\partial \theta} d\theta \\
&= \sin\theta dr + r\cos\theta d\theta.
\end{aligned}$$

Then, we have that

$$\begin{aligned}
h^*(dx \wedge dy) &= h^*(dx) \wedge h^*(dy) \\
&= (\cos\theta dr - r\sin\theta d\theta) \wedge (\sin\theta dr + r\cos\theta d\theta) \\
&= \cos\theta dr \wedge \sin\theta dr + \cos\theta dr \wedge r\cos\theta d\theta - r\sin\theta d\theta \wedge \sin\theta dr - r\sin\theta d\theta \wedge r\cos\theta d\theta \\
&= r\cos^2\theta dr \wedge d\theta - r\sin^2\theta d\theta \wedge dr \\
&= r dr \wedge d\theta
\end{aligned}$$

This is a problem in Homework 6. It also came up during lecture that $h^* f dx = (f \circ h) h^* dx$.

Suppose $V \subset W$ and $i : V \hookrightarrow W$ is the inclusion map. Consider $i^* : W^* \to V^*$. If W and V are finite dimensional, then i^* is surjective. In other words, all linear functionals on V are restrictions of linear functionals on W.

Now consider M a k dimensional submanifold of $\mathbb{R}^n$. Let $i : M \hookrightarrow \mathbb{R}^n$. Then, for $p \in M$, the derivative of i is a map between tangent spaces $\mathbf{D}i_p : T_pM \hookrightarrow T_p\mathbb{R}^n$ and therefore induces a pullback $i_p^* : \bigwedge^j T_p^*\mathbb{R}^n \to \bigwedge^j T_p^*M$.

Let $\phi : U \overset{\text{open}}{\subset} \mathbb{R}^n \to \mathbb{R}^{n-k}$ s.t. 0 is a regular value and define $M = \phi^{-1}(0)$. If $p \in M$, then $\mathbf{D}\phi_p$ is rank $n-k$ thus $\mathbf{D}\phi$ is rank $n-k$ in a neighborhood of p. Theorem 1.30 implies that there is a local diffeomorphism h such that

$$(\phi \circ h)(x_1, \dots, x_n) = (x_{k+1}, \dots, x_n),$$

in this locality. Then, $\phi \circ h(x) = 0$, where x is a vector in $\mathbb{R}^k$ and we assume the usual embedding $x \mapsto (x, 0) \in \mathbb{R}^n$. In other words, $\{x \mid \phi \circ h(x) = 0\}$ is a k dimensional subspace of $\mathbb{R}^n$, so,

$$h|_{\mathbb{R}^k} : U \overset{\text{open}}{\subset} \mathbb{R}^k \mapsto h(U) \subset M \subset \mathbb{R}^n. \tag{1}$$

Consider $\mathbb{R}^k \subset \mathbb{R}^n$. If $b \in \mathbb{R}^k$, then $T_b\mathbb{R}^k$ is spanned by $\left\langle \frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^k} \right\rangle$ and $T_p\mathbb{R}^n$ is spanned by $\left\langle \frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n} \right\rangle$. This defines dual bases $\left\langle dx^1, \dots, dx^k \right\rangle$ and $\left\langle dx^1, \dots, dx^n \right\rangle$. Note that the dx^1 in $T_b^*\mathbb{R}^k$ and dx^1 in $T_p^*\mathbb{R}^n$ are two fundamentally different objects, but we just write them in the same notation out of convenience so that we can do computations. They are related by

$$i^* dx^1_{\mathbb{R}^n} = dx^1_{\mathbb{R}^k}.$$

Definition 3.26 (Coordinate Map)**.** *Let $M^k \subset \mathbb{R}^n$ be a C^∞ submanifold. A coordinate map is a C^∞ function $X_\alpha : U_\alpha \overset{\text{open}}{\subset} \mathbb{R}^k \to \mathbb{R}^n$ s.t.*

1. *$V_\alpha \triangleq X_\alpha(U_\alpha) \subset M$,*
2. *X_α is bijective*
3. *$\mathbf{D}X_\alpha$ is injective*

Note that h from (1) is such a function.

Proposition 3.27. *Theorem 1.27 implies that for every $p \in M$ there is such a $V_\alpha = X_\alpha(U_\alpha)$ containing p.*

Proof. (Obvious from Theorem 1.30) □

Definition 3.28 (Coordinate Neighborhood)**.** *V_α in Definition 3.26 is known as a coordinate neighborhood.*

Definition 3.29 (Differential Form)**.** *A C^∞ j-form on M is a choice of element $\psi \in \bigwedge^j T_p^*\mathbb{R}^n$ such that $X_\alpha^*\psi$ is a smooth j-form on U_α for all coordinate maps α.*

A simple exercise is to show that for any two coordinate maps α and β, there is a local diffeomorphism $U_{\alpha\beta} \subset \mathbb{R}^k \to \mathbb{R}^k$ given by $X_\beta \circ X_\alpha^{-1}$, $X_\beta^*\psi = (X_\beta \circ X_\alpha^{-1})^* X_\alpha^*\psi$.

Consider now a map $f : \mathbb{R}^n \to \mathbb{R}^n$. We shall write coordinates x and y to denote points in the domain and range respectively, i.e., $f(x) = y$. Consider

$$f^*(dy^1 \wedge \cdots \wedge dy^n),$$

an object in $\bigwedge^n T_x\mathbb{R}^n$. It is equal to $\lambda_x(dx^1 \wedge \cdots \wedge dx^n)$ for some $\lambda_x \in \mathbb{R}$. To compute λ_x, observe the following,

$$\begin{aligned} f^*(dy^1 \wedge \cdots \wedge dy^n)\left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right) &= \lambda_x \\ (dy^1 \wedge \cdots \wedge dy^n)\left(\mathbf{D}f\frac{\partial}{\partial x^1}, \ldots, \mathbf{D}f\frac{\partial}{\partial x^n}\right) &= \\ dy^1 \wedge \cdots \wedge dy^n \left(\sum_{j=1}^{n} \frac{\partial y^{j_1}}{\partial x^1}\frac{\partial}{\partial x^1}, \ldots, \sum_{j=1}^{n} \frac{\partial y^{j_n}}{\partial x^n}\frac{\partial}{\partial x^n}\right) &= \\ dy^1 \wedge \cdots \wedge dy^n \left(\frac{\partial f}{\partial x^1}, \ldots, \frac{\partial f}{\partial x^n}\right) &= \lambda_x, \end{aligned}$$

which uses the notation $f(x) = (f^1(x), \ldots, f^n(x)) = (y^1(x), \ldots, y^n(x))$. In other words,

$$\begin{bmatrix} \frac{\partial y^1}{\partial x^1} \\ \vdots \\ \frac{\partial y^n}{\partial x^1} \end{bmatrix} = \frac{\partial}{\partial x^1}\begin{bmatrix} y^1 \\ \vdots \\ y^n \end{bmatrix} = \frac{\partial}{\partial x^1}\begin{bmatrix} f^1 \\ \vdots \\ f^n \end{bmatrix}.$$

We have that

$$dy^1 \wedge \cdots \wedge dy^n = \det(\mathbf{D}f),$$

so that

$$f^*(dy^1 \wedge \cdots \wedge dy^n) = \det(\mathbf{D}f_x)dx^1 \wedge \cdots \wedge dx^n.$$

An extension of this result is that $f^*(g(y)dy^1 \wedge \cdots \wedge dy^n) = g \circ f \det(\mathbf{D}f_x)dx^1 \wedge \cdots \wedge dx^n$.

3.6 Orientation

We need to stud y orientation because we want to do integration. Note that the formula relating the wedge products above is almost the change of variables formula, save the absolute value of the determinant, which is related to the orientation of manifolds.

Definition 3.30. *Let V be an n dimensional vector space. Define an equivalence relation on the set of ordered bases of V as follows. We say that $\{e_1, \ldots, e_n\} \sim \{f_1, \ldots, f_n\}$ if, for the matrix A that satisfies*

$$f_j = \sum_l a_{jk}e_k$$

(A is the change of basis matrix), we have that $\det A > 0$.

Proposition 3.31. *$\sim$ is an equivalence relation*

Proof.

1. $\det I > 0 \implies$ reflexivity.
2. $\det A^{-1} > 0 \iff \det A > 0$, so symmetric.
3. $e \sim f, f \sim g \implies e \sim g$. To verify, suppose A is change of basis matrix from e to f and B is the change of basis matrix from f to g. Then, BA has positive determinant is A and B do.

$\square$

Definition 3.32 (Orientation)**.** *An orientation O_V for V is an element of $^{\{ordered\ bases\}}/_{\sim}$.*

Proposition 3.33. *$^{\{ordered\ bases\}}/_{\sim}$ has only two elements.*

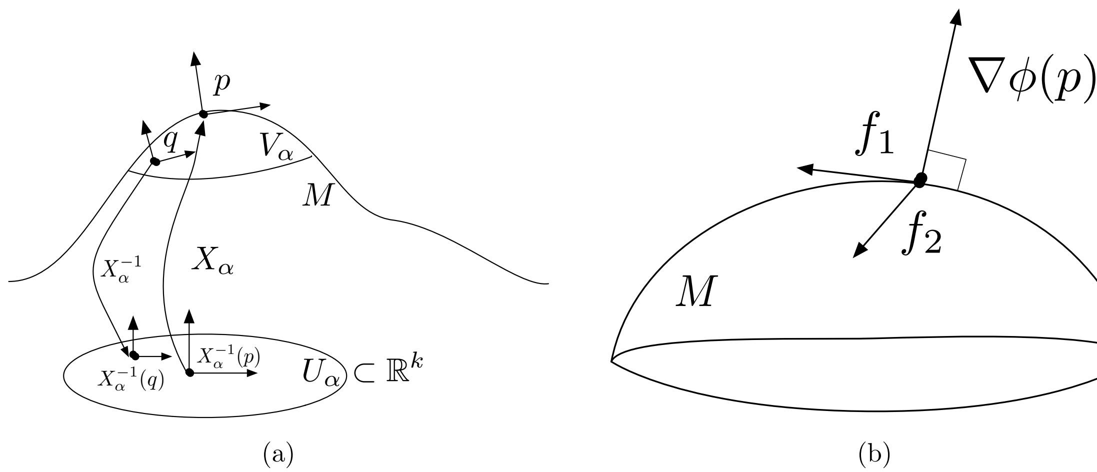

Figure 3: (a) Showing an orientation preserving coordinate map X_α. (b) Showing how an orientation in M in O_p if when we compleet the basis by adding the gradient, the orientation is preserved coordinate in $\mathbb{R}^n$

Here is an alternate view of orientations. Let V be an n dimensional vector space. Let $\{e_1, \dots, e_n\}$ and $\{f_1, \dots, f_n\}$ be ordered basis, which induce $\{\omega^1, \dots, \omega^n\}$ and $\{\psi^1, \dots, \psi^n\}$ their dual bases. Let A be the matrix s.t.

$$\begin{aligned} I_{ik} &= A\psi(e_j) \\ \psi^k(e_j) &= (A^{-1})_{jk} \\ \psi^k &= \sum_j (A^{-1})_{jk}\omega^j \end{aligned}$$

It is instructive to look at how $\omega^1 \wedge \cdots \wedge \omega^n$ relates to $\psi^1 \wedge \cdots \wedge \psi^n$. We have that

$$\{e_1, \dots, e_n\} \sim \{f_1, \dots, f_n\} \iff \operatorname{sign}(\omega^1 \wedge \cdots \wedge \omega^n) = \operatorname{sign}(\{\psi^1 \wedge \cdots \wedge \psi^n\}).$$

Definition 3.34 (Orientation)**.** *A manifold $M^k \subset \mathbb{R}^n$ is oriented if for all $p \in M^k$, there is an orientation O_p of T_pM s.t. if $p \in X_\alpha(U_\alpha) = V_\alpha$, with X_α a coordinate map $X_\alpha : U_\alpha \overset{\text{open}}{\subset} \mathbb{R}^k \to M^k \subset \mathbb{R}^n$. s.t.*

$$\left\{(\mathbf{D}X_\alpha)_{X_\alpha^{-1}(p)}\left(\frac{\partial}{\partial x^1}\right), \dots, (\mathbf{D}X_\alpha)_{X_\alpha^{-1}(p)}\left(\frac{\partial}{\partial x^k}\right)\right\} \in O_p$$
$$\left\{(\mathbf{D}X_\alpha)_{X_\alpha^{-1}(q)}\left(\frac{\partial}{\partial x^1}\right), \dots, (\mathbf{D}X_\alpha)_{X_\alpha^{-1}(q)}\left(\frac{\partial}{\partial x^k}\right)\right\} \in O_q$$

for all $q \in V_\alpha$.

This definition is depicted in Fig. 3(a).

Proposition 3.35. *For all V_α, there is a never zero smooth k-form ϕ on V_α s.t. the orientation determined by $\phi(p)$ agrees with the orientation O_p for all $p \in V_\alpha$.*

Proof. (Left as an exercise in Homework 7). □

Definition 3.36 (Orientable Manifold)**.** *A manifold is orientable if it has an orientation.*

Example 3.37. *Let M be an $n-1$ dimensional sub manifold of $\mathbb{R}^n$. Then, M is orientable.*

Proof. First of all, we know tere is a map $\phi : U \overset{\text{open}}{\subset} \mathbb{R}^n \to \mathbb{R}$ such that $M = \phi^{-1}(0)$, 0 is a regular value. Also note that $\mathbb{R}^n$ is orientable by giving it an ordered basis. Then $T_p\mathbb{R}^n$ inherits this orientation because its basis functions

$$\left\{\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}\right\}$$

are defined by the choice of coordinates for for $\mathbb{R}^n$. Let $p \in M$. Define O_p of T_pM as follows. If $\{f_1, \dots, f_{n-1}\} \in O_p$ if

$$\{f_1, \dots, f_{n-1}, \nabla\phi(p)\} \in O_p$$

is oriented as a basis for $T_p\mathbb{R}^n$. $\square$

This definition is depicted in Fig. 3(b).

3.7 Differentiation of Forms

Let $U \overset{\text{open}}{\subset} \mathbb{R}^n$. Let $\mathcal{a}^k(U)$ denote the set of smooth k-forms on U, i.e., if $f \in \mathcal{a}^k(U)$, we can write

$$f(x) = \sum_{|I|=k} f_I(x) dx^I.$$

Definition 3.38 (Differentiation)**.** *Let $d : \mathcal{a}^0(U) \to \mathcal{a}^1(U)$ be defined by*

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} dx^i,$$

i.e., $df\left(\frac{\partial}{\partial x^j}\right) \frac{\partial f}{\partial x^j} = \mathbf{D}f\left(\frac{\partial}{\partial x^j}\right)$.

Proposition 3.39. *There exists a unique extension of d, given by $d : \mathcal{a}^p(U) \to \mathcal{a}^{p+1}(U)$ s.t.*

1. *$d^2 = 0$.*
2. *If $\phi \in \mathcal{a}^j(U)$ and $\psi \in \mathcal{a}^k(U)$, then $d(\phi \wedge \psi) = d\phi \wedge \psi + (-1)^j \phi \wedge d\psi$.*

d is linear over $\mathbb{R}$.

Proof. Suppose it exists.

1. $dx^i = d(X^i)$. (recall that $d(X^i) = \sum_{k=1}^n \frac{\partial X^i}{\partial x^k} dx^k$.)
2. $d(dx^i) = d^2(X^i) = 0$.
3. $$\begin{aligned} d(dx^i \wedge dx^I) &= \underbrace{ddx^i \wedge dx^J}_{=0} + (-1)dx^i \wedge d(dx^J) \\ &= -dx^i \wedge d(dx^J) \end{aligned}$$

   Now consider the induction to prove the LHS =0. If $|J| = 1$, then clearly $d(dx^J) = 0$. Thus, $d(dx^i \wedge dx^j) = d(dx^I) = 0$ if $|I| = 2$. Now let $|J| = p+1$, and write $J = \{j_1, J'\}$. we have htat

   $$\begin{aligned} d(dx^J) &= d(dx^{j_1} \wedge dx^{J'}) \\ &= dx^{j_1} \wedge d(dx^{J'}) \\ &= 0, \end{aligned}$$

   as desired.

Let $f \in \mathcal{a}^k(U)$, and write $f = \sum_{|I|=k} f_I dx^I$. So

$$\begin{aligned} df &= \sum_{|I|=k} df_I \wedge dx^I \\ &= \sum_{|I|=k,k} \frac{\partial f_I}{\partial x^k} \wedge dx^I \end{aligned}$$

This proves uniqueness of the extension. Defining $df = \sum_{|I|=k,k} \frac{\partial f_I}{\partial x^k} \wedge dx^I$ satisfis linearity and the Liebniz rule because

$$d^2 f = \sum_{|I|=k,k,j} \frac{\partial^2 f_I}{\partial x^k \partial x^j} dx^j \wedge dx^k \wedge dx^I.$$

Note that $\frac{\partial^2 f_I}{\partial x^k \partial x^j} = \frac{\partial^2 f_I}{\partial x^j \partial x^k}$ and $dx^j \wedge dx^k = -dx^k \wedge dx^j$, so all terms in the above summation vanish, i.e., $d^2 f = 0$. $\square$

Example 3.40. $\mathcal{a}^0(\mathbb{R}^3) \xrightarrow{d} \mathcal{a}^1(\mathbb{R}^3) \xrightarrow{d} \mathcal{a}^2(\mathbb{R}^3) \xrightarrow{d} \mathcal{a}^3(\mathbb{R}^3)$.

Before we learned differential forms, we liked to write vectors in the above example instead of differential forms, i.e., we mapped

$$\sum_I f_I dx^I \mapsto (f_1, f_2, f_3) \in \mathbb{R}^3.$$

For example, we might have mapped a two form to $\mathbb{R}^3$ via

$$f = P dx \wedge dy + Q dy \wedge dz + R dz \wedge dx \mapsto (Q, R, P).$$

Differentiating it, we see that

$$\begin{aligned} df &= \frac{\partial P}{\partial z} dx \wedge dy \wedge dz + \frac{\partial Q}{\partial x} dx \wedge dy \wedge dz + \frac{\partial R}{\partial y} dx \wedge dy \wedge dz \\ &= \left(\frac{\partial P}{\partial z} + \frac{\partial Q}{\partial x} + \frac{\partial R}{\partial y} \right) dx \wedge dy \wedge dz \\ &= \operatorname{div}(Q, R, P) dx \wedge dy \wedge dz. \end{aligned}$$

Interestingly, we can recover the grad and curl in this way,

$$\mathcal{a}^0(\mathbb{R}^3) \xrightarrow{\nabla} \mathcal{a}^1(\mathbb{R}^3) \xrightarrow{\text{curl}} \mathcal{a}^2(\mathbb{R}^3) \xrightarrow{\text{div}} \mathcal{a}^3(\mathbb{R}^3).$$