

REVIEW OF MULTIVARIABLE CALCULUS AND LINEAR ALGEBRA

COLTON GRAINGER (MATH 6230 DIFFERENTIAL GEOMETRY)

ASSIGNMENT DUE 2019-01-23

[1, No. C.12] The Chain Rule.

From [the chain rule], we can view open subsets of Banach spaces as the objects of a category, whose morphisms are the continuous maps of class C^p . [2, Ch. I.3]

In the appendix, we show that compositions of maps between Banach spaces (e.g. as a special case, Euclidean spaces) are appropriately smooth.

Given. Let $U \subset \mathbf{R}^n$ and $\tilde{U} \subset \mathbf{R}^m$ be open subsets, and let $x = (x^1, \dots, x^n)$ denote the standard coordinates on U and $y = (y^1, \dots, y^m)$ those on \tilde{U} .

To prove.

- (a) A composition of C^1 functions $F: U \rightarrow \tilde{U}$ and $G: \tilde{U} \rightarrow \mathbf{R}^p$ is again of class C^1 , with partial derivatives given by $\frac{\partial(G \circ F)}{\partial x^j} = \sum_{k=1}^m \frac{\partial G^i}{\partial y^k} \frac{\partial F^k}{\partial x^j}$.
- (b) If F and G are smooth, then so is $G \circ F$.

Proof. (a) follows from the chain rule base case (see appendix), the Jacobi matrix representation (w.r.t. the standard basis) of the total derivative for a function between Euclidean spaces (see theorems for Euclidean space in appendix), and the definition of matrix multiplication. (b) follows from the smooth chain rule (proved in the appendix). \square

[1, No. C.39] **Canonical examples with the inverse function theorem.** *Given.* To verify Example C.37 (Polar Coordinates).

As you know from calculus, polar coordinates (r, θ) in the plane are defined implicitly by the relations $x = r \cos \theta$, $y = r \sin \theta$. The map $F: (0, \infty) \times \mathbf{R} \rightarrow \mathbf{R}^2$ defined by $F(r, \theta) = (r \cos \theta, r \sin \theta)$ is smooth and has Jacobian determinant equal to r , which is nonzero everywhere on the domain. Thus, by Corollary C.36

Suppose $U \subset \mathbf{R}^n$ is an open subset, and $F: U \rightarrow \mathbf{R}^n$ is a smooth function whose Jacobian determinant is nonzero at every point in U . Then F is an open map. Moreover, if F is injective, then $F: U \rightarrow F(U)$ is a diffeomorphism.

the restriction of G to any open subset on which it's injective is a diffeomorphism onto its image. One such subset is $\{(r, \theta) : r > 0, -\pi < \theta < \pi\}$, which is mapped bijectively by F onto the complement of the nonpositive part of the x -axis.

Demonstration. The Jacobian of F evaluated at (r, θ) is found to be the linear transformation

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

Such a map is locally invertible as $\det(DF(r, \theta)) = r(\cos^2 \theta + \sin^2 \theta) \neq 0$ for all $r > 0$. Now everywhere on the set

$$U = \{(r, \theta) : r > 0, -\pi < \theta < \pi\}$$

we have an invertible derivative and a restriction of F to an injective function. Note the composite $(r, \theta) \mapsto re^{i\theta} \mapsto \begin{bmatrix} r \cos \theta \\ r \sin \theta \end{bmatrix}$ is clearly injective on U . The inverse function theorem provides a differentiable inverse for each point in the image $F(U)$. So $F|_U$ is a diffeomorphism.

Given. To verify Example C.38 (Spherical Coordinates).

Similarly, spherical coordinates on \mathbf{R}^3 are the functions (ρ, φ, θ) defined by the relations

$$\begin{aligned} x &= \rho \sin \varphi \cos \theta, \\ y &= \rho \sin \varphi \sin \theta, \\ z &= \rho \cos \varphi. \end{aligned}$$

Geometrically, ρ is the distance from the origin, φ is the angle from the positive z -axis, and θ is the angle from the $x > 0$ half of the (x, z) -plane. If we define $G: (0, \infty) \times (0, \pi) \times \mathbf{R} \rightarrow \mathbf{R}^3$ by

$$G(\rho, \varphi, \theta) = (\rho \sin \varphi \cos \theta, \rho \sin \varphi \sin \theta, \rho \cos \varphi),$$

a computation shows that the Jacobian determinant of G is $\rho^2 \sin \varphi \neq 0$. Thus, the restriction of G to any open subset on which it is injective is a diffeomorphism onto its image. One such subset is

$$\{(\rho, \varphi, \theta) : \rho > 0, 0 < \varphi < \pi, -\pi < \theta < \pi\}.$$

Demonstration. The Jacobian of G (and determinant) evaluated at (ρ, φ, θ) is

$$\begin{aligned} DG(\rho, \varphi, \theta) &= \begin{bmatrix} \sin \varphi \cos \theta & \rho \sin \varphi \cos \theta & -\rho \sin \varphi \sin \theta \\ \sin \varphi \sin \theta & \rho \cos \varphi \sin \theta & \rho \sin \varphi \cos \theta \\ \cos \varphi & -\rho \sin \varphi & 0 \end{bmatrix}, \\ \det(DG(\rho, \varphi, \theta)) &= \rho^2 \sin \varphi. \end{aligned}$$

Because G is injective (with an invertible derivative) on the set

$$V = \{(\rho, \varphi, \theta) : \rho > 0, 0 < \varphi < \pi, -\pi < \theta < \pi\},$$

each point in the image $G(V)$ is locally invertible. So there's a differentiable inverse from $G(V)$ to V . Whence $G|_V$ is a diffeomorphism. \square

Metric tensors. *Given.* A metric g on a vector space V is a symmetric, positive definite, bilinear form $g: V \times V \rightarrow \mathbf{R}$. Given a basis $[e_i]$ for V , we define the components of g w.r.t. this basis to be $g_{ij} = g(e_i, e_j)$. That g is symmetric gives both $g(v, w) = g(w, v)$ for all $v, w \in V$, and also $g_{ij} = g_{ji}$. Bilinearity of g gives a means to write $g(v, w)$ as a linear combination of multiplies of the a^i , the e_i , the b^j , and the e_j . That is

$$v = a^i e_i \quad \text{and} \quad w = b^j e_j \quad \text{implies} \quad g(v, w) = g_{ij} a^i b^j.$$

To demonstrate.

- The matrix $[g_{ij}] = A_g$ is transformed as a rank 2 tensor of type $(1, 1)$ by a change of basis for V .
- A_g represents a self-adjoint linear operator, and gives an isomorphism between V and its dual V^* .

Demonstration. Let $A_g = [g_{ij}]$ where $f_{ij} = g(e_i, e_j)$ w.r.t. the basis $[e_i]$ for V . Consider an arbitrary $R \in GL_n(\mathbf{R})$, along with $v, w \in V$. Say $v = [e_i][a^i]$ and $w = [e_j][b^j]$. Let $[f_i]$ be another basis for V given by $[f_i] = [e_i]R$. Then

$$\begin{aligned} v &= [e_i]RR^{-1}[a^i] \\ &= [f_i]R^{-1}[a^i] \\ w &= [e_j]RR^{-1}[b^j] \\ &= [f_j]R^{-1}[b^j]. \end{aligned}$$

Now the metric $g(v, w)$ as a scalar product is independent of basis, so w.r.t. the basis $[e_i]$ we have

$$\begin{aligned}[a^i]^T A_g [b^j] &= [a^i] [g_{ji} b^j] \\ &= a^i g_{ji} b^j \\ &= g_{ij} a^i b^j.\end{aligned}$$

It should also be that

$$\underbrace{g(v, w)}_{\text{coord. indep.}} = \underbrace{g_{ij} a^i b^j}_{\text{coord. dep.}}$$

given our expression for v and w in the basis $[f_i]$. That is, if \tilde{A}_g is the matrix representing the metric g w.r.t. the basis $[f_i]$, then

$$\begin{aligned}g([f_i] R^{-1} [a^i], [f_j] R^{-1} [b^j]) &= [R^{-1} [a^i]]^T \tilde{A}_g [R^{-1} [b^j]] \\ &= [a^i]^T (R^{-1})^T \tilde{A}_g R^{-1} [b^j].\end{aligned}$$

We conclude that $[a^i]^T A_g [b^j] = [a^i]^T (R^{-1})^T \tilde{A}_g R^{-1} [b^j]$ for all coefficients $[a^i]$ and $[b^j]$, hence $R^T A_g R = \tilde{A}_g$. We say that the matrices A_g and \tilde{A}_g are *congruent*.

The matrix A_g *could* represent a linear transformation (in the sense that a self-adjoint matrix of course *could be associated* with a linear transformation) but not each linear transformation has a representative of the form A_g .

From [3]:

A vector space becomes related to its dual space by the metric. Given a vector space V , with metric g , and a vector v , we have a linear map $g(v, \cdot): V \rightarrow \mathbf{R}$. But $g(v, \cdot)$ is itself linear in v , so the map defined by $V \rightarrow V^*, v \mapsto g(v, \cdot)$ is linear. Because g is positive definite, this map is an isomorphism.

Such a relation between vectors $v \in V$ and covectors in V^* implies that g is a rank 2 tensor of type $(1, 1)$. \square

Appendix: Differential Calculus in Banach Spaces.

As before, we prefer to work in a coordinate-free representation. In other words, we develop the differential calculus for maps between Banach spaces. This representation is conceptually simple and actually makes many expressions look much simpler. The classical formulas for the derivatives in the usual coordinate representation follow easily from the general results using the product structure of finite-dimensional Euclidean spaces. [4, Ch. VII]

Road map. We aim to consider finite dimensional real vector spaces as Banach spaces. We need to:

- Define the Fréchet derivative in Banach spaces.¹
- Prove a composition of p -times continuously differentiable maps is again C^p .
- Define partial derivatives in the special case that the codomain is a finite dimensional vector space.
- Argue the Jacobian matrix representing the derivative function with respect to the standard bases between finite dimensional real vector spaces has entries given by

$$\frac{\partial(G^i \circ F)}{\partial x^j} = \sum_{k=1}^m \frac{\partial G^i}{\partial y^k} \frac{\partial F^k}{\partial x^j}.$$

Definitions. (Paraphrased from Lang [2, Ch. I]) A *Banach space* is a topological vector space that's complete and whose topology can be defined by a norm. The set of *continuous linear maps* from one Banach space E to another F will be denoted by $L(E, F)$, whereas the set of *continuous r -multilinear maps*

$$\psi: E \times \cdots \times E \rightarrow F$$

¹Idea from <https://math.stackexchange.com/questions/1851553>, <https://math.stackexchange.com/questions/2826748>.

of E into F will be denoted by $L^r(E, F)$. We put a topology on $L^n(E, F)$ for all $n \geq 0$ by declaring $L^0(E, F) := F$ and defining the norm of an r -multilinear map $\psi \in L^r(E, F)$ to be

$$\sup\{K : |\psi(x^1, \dots, x^r)| \leq K|x^1| \cdots |x^r|\}.$$

The $L^n(E, F)$ are thus Banach spaces.

Proposition. If E_1, \dots, E_r are Banach spaces, then the canonical map $L(E_1, L(E_2, \dots, L(E_r, F), \dots))$ to $L^r(E_1, \dots, E_r, F)$ is a norm-preserving toplinear isomorphism.

Definition. Say E and F are Banach spaces, $U \subset E$ is open and $a \in U$. A map $f: U \rightarrow F$ is said to be *differentiable at a* if there's a linear approximation $\lambda_a \in L(E, F)$ such that

$$f(a + v) = f(a) + \lambda_a(v) + r(v) \quad \text{where } \frac{|r(v)|}{|v|} \rightarrow 0 \text{ as } v \rightarrow 0.$$

Remark. λ_a is uniquely determined by f and a . We say λ_a is the *derivative of f at a* (also called the *total* or *Fréchet* derivative of f at a). We denote λ_a by $\partial f(a)$, $Df(a)$, etc.

If f is differentiable at each $a \in U$, then we have a map

$$Df : U \rightarrow L(E, F), \quad a \mapsto Df(a),$$

which we say is *the derivative function of f* (known also as the *Fréchet*, or *total* derivative of f).

Proposition. (Chain rule) Say E , F , and G are Banach spaces and we've maps defined on open subsets

$$U \subset E \xrightarrow{f} V \subset F \xrightarrow{g} G.$$

If f and g are differentiable at $a \in U$ and $f(a) \in V$ (resp.), then $g \circ f$ is differentiable at a with derivative

$$\underbrace{D(g \circ f)(a)}_{\text{in } L(E, G)} = \underbrace{Dg(f(a))}_{\text{in } L(F, G)} \underbrace{Df(a)}_{\text{in } L(E, F)}.$$

Proof. (Adapted from Folland [5, Ch. 2.6]) Let $b = f(a)$. By definition of differentiability at a point, both

$$\begin{aligned} g(b + h) &= g(b) + Dg(b)h + E_1(h), & \frac{|E_1(h)|}{|h|} &\rightarrow 0 \text{ as } h \rightarrow 0 \\ f(a + k) &= f(a) + Df(a)k + E_2(k), & \frac{|E_2(k)|}{|k|} &\rightarrow 0 \text{ as } k \rightarrow 0. \end{aligned}$$

Let $h := f(a + k) - f(a)$, then so too $h = Df(a)k + E_2(k)$. Consider

$$\begin{aligned} (g \circ f)(a + k) &= g(b + h) = g(b) + Dg(b)h + E_1(h) \\ &= g(f(a)) + Dg(b)[Df(a)k + E_2(k)] + E_1(h) \\ &= (g \circ f)(a) + Dg(b)Df(a)k + E_3(k), \end{aligned}$$

where $E_3(k) = Dg(b)E_2(k) + E_1(h)$. We claim $E_3(k)/|k| \rightarrow 0$ as $k \rightarrow 0$, in which case

$$(g \circ f)(a + k) = (g \circ f)(a) + Dg(b)Df(a)k + \text{remainder}$$

confirms that the linear map $Dg(b)Df(a)$ is the total derivative of $g \circ f$ at a . For the claim, consider each term in the error summand. By our definition of a norm for linear maps,

$$\frac{|Dg(b)E_2(k)|}{|k|} \leq |Dg(b)| \frac{|E_2(k)|}{|k|} \rightarrow 0, \quad \text{as } k \rightarrow 0.$$

This also implies when k is a vector close to 0, then $|E_2(k)| \leq |k|$, hence by the triangle inequality and our definition of a norm for linear maps

$$|h| = |Df(a)k + E_2(k)| \leq \underbrace{(|Df(a)| + 1)}_{\text{constant}} |k|.$$

Since $|k|$ is bounded below by a constant multiple of $|k|$, as $k \rightarrow 0$, it's true also that

$$\frac{|E_1(h)|}{|k|} \leq \frac{|E_1(h)|}{\text{const.} \cdot |h|} \rightarrow 0.$$

□

Definitions. (Directional and partial derivatives) Consider a map $f: U \rightarrow F$ from open U in E , choose a base point $a \in U$, and a vector $v \in E \setminus \{0\}$. Because U is open, there's an $\epsilon > 0$ such that $a + tv \in U$ for all $t \in (-\epsilon, \epsilon)$. We dub the *directional derivative of f at a in the v direction to be

$$\partial_v f(a) := \frac{d}{dt} f(a + tv)|_{t=0} = \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t}.$$

To quote [4]:

If $E = \mathbf{R}^n$, the derivatives in the direction of the coordinate axes are particularly important, for practical and historical reasons. ... We thus write ∂_k or $\partial/\partial x^k$ for the derivatives in the direction of the standard basis vectors e_k for $k = 1, \dots, n$.

Theorems for Euclidean space. [4, No. VII.2.8]

- i. Suppose $E = \mathbf{R}^n$ and $f: U \rightarrow F$ is differentiable at $a \in U$. Then

$$Df(a)h = \sum_{k=1}^n \partial_k f(a) h^k \quad \text{for } h = (h^1, \dots, h^k) \in \mathbf{R}^n.$$

- ii. Suppose E is a Banach space and $f = (f^1, \dots, f^m): U \rightarrow \mathbf{F}^m$ (an m dimensional \mathbf{F} -vector space). The f is differentiable at a if and only if all the coordinate functions f^j are differentiable at a . Then

$$Df(a) = (Df^1(a), \dots, Df^m(a)).$$

- iii. Suppose U is open in \mathbf{R}^n and $f = (f^1, \dots, f^m): U \rightarrow \mathbf{R}^m$ is differentiable at a . The matrix representation (in the standard basis) of the derivative of f is the Jacobi matrix of f :

$$[Df(a)] = [\partial_k f^j(a)].$$

Definition. (Continuously differentiable) Let U be open in a Banach space E and say $f: U \rightarrow F$ is differentiable. When the derivative function $Df: U \rightarrow L(E, F)$ is continuous, we say that f is *continuously differentiable*, or of *class C^1* . To define class C^p , we need to inductively define higher total derivatives. The p th derivative of f is $D(D^{p-1}f)$, which is a map

$$D^p f: U \rightarrow L(E, L(E, \dots, L(E, F), \dots)) \cong L^p(E, F) \quad \text{by isometry.}$$

A maps is said to be of *class C^p* if its k th derivative exists $1 \leq k \leq p$.

Smooth Chain Rule (Base Case). If $U \xrightarrow{f} V \xrightarrow{g} V \xrightarrow{G}$ are maps of class C^1 between open sets of Banach spaces E, F , and G , then the composite $g \circ f$ is of class C^1 .

Proof. Consider a point $a \in U$. By the chain rule, defining $\varphi(a) = (Dg(f(a)), Df(a))$ and $\psi(A, B) = AB$ (as composition of linear maps), TFDC.

$$\begin{array}{ccc} U & \xrightarrow{\varphi} & L(E, F) \times L(F, G) \\ & \searrow D(g \circ f) & \downarrow \psi \\ & & L(E, G) \end{array}$$

Since φ is continuous in each coordinate, φ is a continuous. At the same time, ψ is just the product of linear functions $(A, B) \mapsto AB$, hence is continuous.² Thence $g \circ f$ is of class C^1 .

Lemma. Any continuous multilinear map $\psi \in L^r(E, F)$ is infinitely differentiable.

Proof. Let (a^1, \dots, a^r) be a basepoint and (h^1, \dots, h^r) be a step vector in $E \times \dots \times E$. Observe $\psi(a^1 + h^1, \dots, a^r + h^r)$ can be written as 2^n terms

$$\psi(a^1, \dots, a^r) + \sum_{i=1}^r \psi(a^1, \dots, h^i, \dots, a^r) + \sum \text{terms with at least two coordinates } h_i, h_j.$$

Assuming ψ is continuous, the last $2^r - (r+1)$ terms converge to 0_F superlinearly as $(h^1, \dots, h^r) \rightarrow 0$. Note that in whatever norm we choose for $E \times \dots \times E$, that norm is equivalent to, say, the max norm, so that

$$\frac{|\psi(a^1, \dots, h^i, \dots, h^j, \dots, a^r)|}{(|h^1, \dots, h^r|)} \leq \frac{|\psi| \cdot \text{const.} \cdot |h^i| |h^j|}{|h^{\max}|} \rightarrow 0.$$

So we have a continuous $(r-1)$ -multilinear map

$$D\psi(a): h \mapsto \sum_{i=1}^r \psi(a^1, \dots, h^i, \dots, a^r),$$

which implies that the map $D\psi: a \mapsto D\psi(a)$ is r -multilinear and continuous as well. Given continuous bilinear maps as a base case, by induction on r we've show all continuous r -multilinear maps are infinitely differentiable. \square

Smooth Chain Rule (Inductive Step). If $U \xrightarrow{f} V \xrightarrow{g} W \xrightarrow{G}$ are maps of class C^p between open sets of Banach spaces E, F , and G , then the composite $g \circ f$ is of class C^p .

Proof. Let our inductive hypothesis be that the composite of any two compatible C^{p-1} maps on Banach spaces is also C^{p-1} . Assume now that $f \in C^p(U, F)$ and $g \in C^p(V, G)$. Say a point $a \in U$. By the chain rule, defining as before $\varphi(a) = (Dg(f(a)), Df(a))$ and $\psi(A, B) = AB$, TFDC.

$$\begin{array}{ccc} U & \xrightarrow{\varphi} & L(E, F) \times L(F, G) \\ & \searrow D(g \circ f) & \downarrow \psi \\ & & L(E, G) \end{array}$$

By assumption, the map $a \mapsto Df(a)$ is in $C^{p-1}(U, L(E, F))$ and $b \mapsto Dg(b)$ is in $C^{p-1}(V, L(F, G))$. By inductive hypotheses, the composition $a \mapsto f(a) \mapsto Dg(f(a))$ of C^{p-1} maps is in $C^{p-1}(U, L(E, G))$. By its coordinates φ is in $C^{p-1}(U, L(E, F) \times L(F, G))$. By lemma, the continuous bilinear map ψ is in $C^\infty(L(E, F) \times L(F, G), L(E, G))$. By the inductive hypothesis, then $\psi \circ \varphi$ is in $C^{p-1}(U, L(E, G))$. so $D(g \circ g)$ is of class C^{p-1} , and we conclude $g \circ f$ is of class C^p . \square

REFERENCES

- [1] J. M. Lee, *Introduction to Smooth Manifolds*. New York: Springer-Verlag, 2003.
- [2] S. Lang, *Fundamentals of Differential Geometry*. New York: Springer-Verlag, 1999.
- [3] A. Lahiri, "Lecture Notes on Differential Geometry for Physicists," 2011.
- [4] H. Amann and J. Escher, Eds., "Multivariable differential calculus," in *Analysis II*, Basel: Birkhäuser Basel, 2008, pp. 115–278.
- [5] G. B. Folland, *Advanced Calculus*. Pearson, 2002.

²Idea from <https://math.stackexchange.com/questions/2826748>.