MATH 291 WORKSHOP 2

MARK H. KIM

Contents

1. Limits and Continuity in Metric Spaces

The motivation for defining an abstract space in mathematics often comes from the desire to generalize a certain property that the Euclidean space $\mathbb{R}^n$ possesses. By doing so, we may minimize redundant effort in developing mathematical tools we may use in a wide variety of contexts. In this section, we take the notion of distance and extract the defining properties thereof.

Definition 1. A set X is a *metric space* if it comes equipped with a *metric*, which is a function $d : X \times X \to [0, \infty)$ satisfying

(i) $d(x, x) = 0$ if and only if $x \in X$;
(ii) $d(x, y) = d(y, x)$ for all $x, y \in X$;
(iii) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

Property (iii) is called the *triangle inequality.* $\mathbb{R}^n$ with the usual vector norm $\|\cdot\|$ is a metric space, with the metric defined by

$$d(\boldsymbol{v}, \boldsymbol{w}) = \|\boldsymbol{v} - \boldsymbol{w}\|.$$

$\mathbb{R}^n$ can be given a number of other metrics. One is the *taxicab metric* (also known as the *Manhattan distance* or the l^1 *norm*)

$$d_t(\boldsymbol{v}, \boldsymbol{w}) = \sum_{i=1}^{n} |v_i - w_i|,$$

and another is the *supremum metric*

$$d_s(\boldsymbol{v}, \boldsymbol{w}) = \max\{|v_1 - w_1|, \ldots, |v_n - w_n|\}.$$

Verifying properties (i) and (ii) for both metrics is straightforward; (iii) follows from the triangle inequality for the usual absolute value in $\mathbb{R}$.

We can, of course, define metrics on other sets as well. For example, the collection $\mathbb{Q}$ of rational numbers with the usual absolute value $|\cdot|$ is a metric space, with the metric define by

$$d(p,q) = |p - q|,$$

and the collection $\mathbb{R}^+$ of positive real numbers with

$$d(x,y) = |\log(y/x)|$$

is a metric space as well. Any set X can be given at least one metric, namely the *discrete metric*

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y. \end{cases}$$

(This is a rather dull metric, of course.)

What can we do with this abstract notion of distance? It turns out that the concept of distance is all that is required to make sense of the notions of limits and continuity.

Definition 2. Let (X, d_1) and (Y, d_2) be metric spaces. A function $f : X \to Y$ has a *limit* $l \in Y$ at $x_0 \in X$ if, for every $\varepsilon > 0$, we can find a $\delta > 0$ such that each $x \in X$ satisfying the inequality

$$d_1(x, x_0) < \delta$$

also satisfies the inequality

$$d_2(f(x), l) < \varepsilon.$$

f is *continuous* at x_0 if f has a limit l at x_0 and $f(x_0) = l$.

This is, of course, a direct generalization of the ε-δ definition of limits and continuity in $\mathbb{R}$. In particular, we can specialize the above definition for functions on a Euclidean space:

Definition 3. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ has a *limit* $\boldsymbol{l} \in \mathbb{R}^m$ at $\boldsymbol{x}_0 \in \mathbb{R}^n$ if, for every $\varepsilon > 0$, we can find a $\delta > 0$ such that each $\boldsymbol{x} \in \mathbb{R}^m$ satisfying the inequality

$$\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$$

also satisfies the inequality

$$\|f(\boldsymbol{x}) - \boldsymbol{l}\| < \varepsilon$$

f is *continuous* at $\boldsymbol{x}_0$ if f has a limit $\boldsymbol{l}$ at $\boldsymbol{x}_0$ and $f(\boldsymbol{x}_0) = \boldsymbol{l}$.

Many fundamental properties of limits and continuity in $\mathbb{R}$ hold in generic metric spaces. Recall, for example, the following characterization of continuity:

Theorem 4. *A function $f : \mathbb{R} \to \mathbb{R}$ is continuous at $x_0 \in \mathbb{R}$ if and only if, for every sequence $(x_n)_{n=1}^\infty$ of real numbers converging to x_0, we have*

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right) = f(x_0).$$

In order to generalize the above theorem, we must define what it means for a sequence in a metric space to converge to some point. The following is, once again, a direct generalization of the ε-neighborhood definition of convergence in $\mathbb{R}$:

Definition 5. A sequence $(x_n)_{n=1}^\infty$ in a metric space (X, d) *converges* to $x \in X$ if, for each $\varepsilon > 0$, we can find a natural number N such that $n > N$ implies $d(x_n, x) < \varepsilon$.

We are now ready to prove the following

Theorem 6. *Let (X, d_1) and (Y, d_2) be metric spaces, $f : X \to Y$ a function, and $x_0 \in X$ a point. Then f is continuous at x_0 if and only if, for every sequence $(x_n)_{n=1}^{\infty}$ in X converging to x_0, we have*

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right) = f(x_0).$$

Proof. ($\Rightarrow$) Suppose that f is continuous at x_0, and fix a sequence $(x_n)_{n=1}^{\infty}$ converging to x_0. Given a $\varepsilon > 0$, we can find $\delta > 0$ such that $d_2(f(x), f(x_0)) < \varepsilon$ for all $x \in X$ with $d_1(x, x_0) < \delta$. Since $x_n \to x_0$, we can also find $N \in \mathbb{N}$ such that $d_1(x_n, x_0) < \delta$ for all $n > N$. It follows that $d_1(f(x_n), f(x_0)) < \varepsilon$ for all $n > N$, whence the desired result follows.

($\Leftarrow$) Conversely, we now suppose that f is *not* continuous at x_0, which means we can find an $\varepsilon > 0$ such that, for every $\delta > 0$, there exists $x \in X$ such that $d(x, x_0) < \delta$ but $d(f(x), f(x_0)) \geq \varepsilon$. We now construct a sequence which shall serve as a counterexample. For each $n \in \mathbb{N}$, we fix $x_n \in X$ such that $(x_n, x_0) < 1/n$ but $d(f(x_n), f(x_0)) \geq \varepsilon$. Then $(x_n)_{n=1}^{\infty}$ converges to x_0, but $f(x_n)$ does not converge to $f(x_0)$, as was to be shown. □

The "algebraic operations of limits" theorem also holds in metric spaces:

Theorem 7. *Let (X, d) be a metric space, and $f : X \to \mathbb{R}^n$ an $g : X \to \mathbb{R}^n$ be functions. Fix $x_0 \in X$.*

(i) $\lim_{x\to x_0} f(x) + g(x) = \lim_{x\to x_0} f(x) + \lim_{x\to x_0} g(x)$.

(ii) $\lim_{x\to x_0} f(x) \cdot g(x) = \left(\lim_{x\to x_0} f(x)\right) \cdot \left(\lim_{x\to x_0} g(x)\right)$;

(iii) If $n = 1$, then $\lim_{x\to x_0} f(x) \div g(x) = \left(\lim_{x\to x_0} f(x)\right) \div \left(\lim_{x\to x_0} g(x)\right)$ *so long as* $\lim_{x\to x_0} g(x) \neq 0$.

We omit the proof.

2. Linear maps and Matrices

We now restrict our attention to Euclidean spaces. Recall that a function $T : \mathbb{R}^n \to \mathbb{R}^m$ is *linear* if

(i) $T(\boldsymbol{x} + \boldsymbol{y}) = T(\boldsymbol{x}) + T(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$;
(ii) $T(\lambda \boldsymbol{x}) = \lambda T(\boldsymbol{x})$ for all $\lambda \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^n$.

In this section, we discuss a concrete way of representing linear maps.

A (real) *matrix* A is a rectangular array of real numbers, called *entries*. We denote by a_{ij} the entry in the ith row and the jth column. Addition and subtraction of matrices are defined componentwise, just like vector addition and subtraction. In fact, we can think of an m-dimensional vector as a matrix with one row and m columns. Note that this differs from the usual notational convention in multivariable calculus to write vectors horizontally. Considering vectors as "vertical matrices", however, fits nicely into the notational convention for functions, as we shall see.

We say that a matrix is of *size* n by m if it has n rows and m columns. A matrix A of size n by m and another matrix B of size m by p can be *multiplied* as follows: the product $C = A \cdot B$ is defined to be the n by p matrix, whose entries are determined by

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}.$$

For example,

$$\begin{bmatrix} 1 & 3 & 5 \\ 7 & 2 & 3 \\ 2 & 3 & 8 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 + 3 \cdot 3 + 5 \cdot 4 \\ 7 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 \\ 2 \cdot 2 + 3 \cdot 3 + 8 \cdot 4 \end{bmatrix} = \begin{bmatrix} 31 \\ 32 \\ 45 \end{bmatrix}.$$

Given a matrix A, we also define *scalar multiplication* by a real number c by setting $B = cA$, where $b_{ij} = ca_{ij}$. This, of course, coincides with scalar multiplication for vectors whenever A is of size 1 by m.

In general, the matrix multiplication is *noncommutative*: that is, if A and B are matrices, then it is not guaranteed that $AB = BA$. Nevertheless, there are a number of useful algebraic properties that the matrix multiplication satisfies, whose proof we omit:

Theorem 8. *The matrix multiplication satisfies the following properties:*

(i) $A \cdot (B \cdot C) = (A \cdot B) \cdot C$;
(ii) $A \cdot (B + C) = A \cdot B + A \cdot C$;
(iii) $(A + B) \cdot C = A \cdot C + B \cdot C$;
(iv) $(cA) \cdot B = c(A \cdot B) = A \cdot (cB)$.

We now return to the main goal of the section: representing linear functions concretely. More specifically, we shall establish the following

Theorem 9. *Given an n by m matrix A, the function $T(\boldsymbol{x}) = A \cdot \boldsymbol{x}$ is a linear function from $\mathbb{R}^m$ to $\mathbb{R}^n$. Conversely, every linear map $T : \mathbb{R}^m \to \mathbb{R}^n$ can be written as $T(\boldsymbol{x}) = A \cdot \boldsymbol{x}$, where A is an n by m matrix.*

Proof. It follows from properties (ii) and (iv) of Theorem 8 that every function of the form $T(\boldsymbol{x}) = A \cdot \boldsymbol{x}$ is linear. Conversely, if we are given a linear map $T : \mathbb{R}^m \to \mathbb{R}^n$, we can define a matrix A by setting its ith column to be the vector $T(\boldsymbol{e}_i)$, where

$$\boldsymbol{e}_i = (0, \cdots, 1, \cdots, 0)$$

is the ith coordinate vector. Then A is an n by m matrix, and $A \cdot \boldsymbol{e}_i = T(\boldsymbol{e}_i)$ for all i. Given any $\boldsymbol{x} \in \mathbb{R}^m$, we can write

$$\boldsymbol{x} = \sum_{i=1}^{m} \lambda_i \boldsymbol{e}_i,$$

for scalars λ_i, whence by linearity

$$\begin{aligned} A \cdot \boldsymbol{x} &= A \cdot \left(\sum_{i=1}^{m} \lambda_i \boldsymbol{e}_i \right) \\ &= \sum_{i=1}^{m} \lambda_i (A \cdot \boldsymbol{e}_i) \\ &= \sum_{i=1}^{m} \lambda_i T(\boldsymbol{e}_i) \\ &= T\left(\sum_{i=1}^{m} \lambda_m \boldsymbol{e}_i \right) \\ &= T(\boldsymbol{x}), \end{aligned}$$

as was to be shown. □

3. Differentiability and Smoothness Conditions

Since the computation of the derivative of a function is the process of finding a local linear approximation, we generalize the notion of differentiability as follows:

Definition 10. A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is *differentiable* at $\boldsymbol{x} \in \mathbb{R}^m$ if there is a linear function $T : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$\lim_{\boldsymbol{h} \to 0} \frac{f(\boldsymbol{x} + \boldsymbol{h}) - f(\boldsymbol{x}) - T(\boldsymbol{h})}{\|\boldsymbol{h}\|} = \mathbf{0}.$$

If f is differentiable at $\boldsymbol{x}$, then we say that T is the *derivative* of f at $\boldsymbol{x}$, and we write $Df(\boldsymbol{x})$ to denote the linear function T.

Compare the above with the following (perhaps) more "straightforward" generalization:

Definition 11. A function $f : \mathbb{R}^m \to \mathbb{R}^n$ has a *directional derivative* at $\boldsymbol{x} \in \mathbb{R}^m$ along a unit vector $\boldsymbol{u} \in \mathbb{R}^m$ if the limit

$$f'(\boldsymbol{x}; \boldsymbol{u}) = \lim_{t \to 0} \frac{f(\boldsymbol{x} + t\boldsymbol{u}) - f(\boldsymbol{x})}{t}$$

exists. $f'(\boldsymbol{x}; \boldsymbol{u})$ is called the *directional derivative* of f at $\boldsymbol{x}$ along $\boldsymbol{u}$.

The directional derivatives, while easier to compute, are not good candidates for characterizing differentiability. For one, it is possible for a function to have directional derivatives along all directions and still fail to be continuous.

Example. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x, y) = \begin{cases} 0 & \text{if } x = y = 0; \\ \frac{x^2 y}{x^4 + y^2} & \text{otherwise.} \end{cases}$$

Given a unit vector $\boldsymbol{u} = (h, k)$, we see that

$$\begin{aligned} \frac{f(\mathbf{0} + t\boldsymbol{u}) - f(\mathbf{0})}{t} &= \frac{(th)^2 (tk)}{(th)^4 + (tk)^2} \cdot \frac{1}{t} \\ &= \frac{h^2 k}{t^2 h^4 + k^2}, \end{aligned}$$

whence

$$f'(\mathbf{0};\boldsymbol{u}) = \begin{cases} h^2/k & \text{if } k \neq 0; \\ 0 & \text{if } k = 0. \end{cases}$$

Therefore, f has directional derivatives along all directions at $\mathbf{0}$. Nevertheless, we observe that

$$f(t,t^2) = \frac{1}{2}$$

for all $t \neq 0$, whence f is *not* continuous at $\mathbf{0}$.

The regular differentiability certainly implies continuity:

Theorem 12. *If $f : \mathbb{R}^m \to \mathbb{R}^n$ is differentiable at $\boldsymbol{x} \in \mathbb{R}^m$, then it is continuous at $\boldsymbol{x}$.*

Proof. Observe that

$$\begin{aligned} \lim_{\boldsymbol{h}\to\mathbf{0}} f(\boldsymbol{x}+\boldsymbol{h}) - f(\boldsymbol{x}) &= \lim_{\boldsymbol{h}\to\mathbf{0}} \|\boldsymbol{h}\| \left(\frac{f(\boldsymbol{x}+\boldsymbol{h}) - f(\boldsymbol{x}) - Df(\boldsymbol{x})(\boldsymbol{h})}{\|\boldsymbol{h}\|} \right) + Df(\boldsymbol{x})(\boldsymbol{h}) \\ &= \mathbf{0} + \mathbf{0} \\ &= \mathbf{0}, \end{aligned}$$

whence f is continuous at $\boldsymbol{x}$. □

We now discuss briefly the method of computing the derivative of $f : \mathbb{R}^m \to \mathbb{R}^n$. Recall that any linear function has a matrix representation, whence we can think of the linear map $Df(\boldsymbol{x}) : \mathbb{R}^m \to \mathbb{R}^n$ as an n by m matrix. How do we find the components of this matrix? We first define a special collection of directional derivatives for real-valued functions of several variables, which are particularly easy to compute:

Definition 13. The *ith partial derivative* of $g : \mathbb{R}^m \to \mathbb{R}$ is the directional derivative of g along the ith coordinate vector $\boldsymbol{e}_i$.

To compute the ith partial derivative of g, we simply consider all but the ith variable as constants and compute the one-variable derivative. We denote the ith partial derivative of g by

$$\frac{\partial g}{\partial x_i}.$$

We have the following convenient computational criterion:

Theorem 14. *Let $f_i : \mathbb{R}^m \to \mathbb{R}$ be the ith component function of $f : \mathbb{R}^m \to \mathbb{R}^n$, so that*

$$f(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \cdots, f_n(\boldsymbol{x})).$$

Then f is differentiable at $\boldsymbol{x} \in \mathbb{R}^m$ if and only if each f_i is differentiable at $\boldsymbol{x}$. Furthermore, if f is differentiable at $\boldsymbol{x}$, then

$$Df(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_m} \end{bmatrix},$$

which is known as the Jacobian matrix.

Proof. See Theorem 5.4 of [Munkres] or Theorem 2-3(3) of [Spivak]. □

We could now ask the converse question: if the partial derivatives exist is the function differentiable? A quick glance at the example

$$f(x,y) = \begin{cases} 0 & \text{if } x = y = 0; \\ \frac{x^2 y}{x^4 + y^2} & \text{otherwise.} \end{cases}$$

above shows that this is not the case. What we need is a slightly stronger condition, commonly referred to as *continuous differentiability.*

Definition 15. A function is *continuously differentiable* if all of its partial derivatives exist and are continuous.

Continuous differentiability provides a converse to Theorem 14:

Theorem 16. *If a function is continuously differentiable at a point, then it is differentiable at the same point.*

Proof. See Theorem 6.2 of [Munkres] or Theorem 2-8 of [Spivak]. □

Continuously differentiable functions are also referred to as functions of *class* $\mathscr{C}^1$. A function is of class $\mathscr{C}^2$ if the partial derivatives of the partial derivatives exist and are continuous. Applying Theorem 16 twice, we see that functions of $\mathscr{C}^2$ are twice-differentiable. A useful theorem concerning $\mathscr{C}^2$-functions is as follows:

Theorem 17 (Clairaut). *If f is of class $\mathscr{C}^2$, then*

$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f = \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} f$$

for all i and j.

Proof. See Theorem 6.3 of [Munkres] or Theorem 2-5 of [Spivak]. □

It is easy to see how functions of class $\mathscr{C}^r$ should be defined for any integer $r > 0$. For notational purposes, we frequently refer to continuous functions as $\mathscr{C}^0$-*functions.* A function is of *class* $\mathscr{C}^\infty$ if the partial derivatives of all orders exist and are continuous. A $\mathscr{C}^\infty$-function is also called *smooth* and is infinitely differentiable. A function is *analytic*, or of *class* $\mathscr{C}^\omega$, if it is infinitely differentiable *and* it has a Taylor series expansion. The $\mathscr{C}^r$ characterizations of functions are collectively referred to as *smoothness conditions.*

Though we have primarily dealt with "very smooth" functions in calculus, it is in general not true that a weaker smoothenss condition implies a stronger smoothness condition. For example, $f(x) = |x|$ is continuous but not differentiable, and $g(x) = \int_{-3}^{x} |t|\, dt$ is differentiable but not twice-differentiable. The function

$$h(x) = \begin{cases} e^{-1/x} & \text{if } x > 0; \\ 0 & \text{if } x \leq 0; \end{cases}$$

is smooth, but not analytic.

4. Suggested Solutions to the Challenge Problems

Let $\boldsymbol{b}(t)$ be a piecewise $\mathscr{C}^1$ vector-valued function[1] into $\mathbb{R}^3$. The goal of the problem set is to find the solution $\boldsymbol{v}(t)$ to the system

$$\begin{aligned}\boldsymbol{v}'(t) &= \boldsymbol{b}(t) \times \boldsymbol{v}(t)\\ \boldsymbol{v}(t_0) &= \boldsymbol{v}_0,\end{aligned}$$

known as the *rotation equation.* To this end, we first need a slight modification of Theorem 26 in Chapter 2 of [Carlen]:

Theorem 18. *For any given vectors $\boldsymbol{b}$ and $\boldsymbol{v}_0$ in $\mathbb{R}^3$ and the initial point t_0 in $\mathbb{R}$, there is precisely one continuous function $\boldsymbol{v}(t)$ satisfying the rotation equation. Furthermore, if $\boldsymbol{b} = \boldsymbol{0}$ or $\boldsymbol{b} = \lambda \boldsymbol{v}_0$ for some $\lambda \in \mathbb{R}$, then*

$$\boldsymbol{v}(t) = \boldsymbol{v}_0;$$

otherwise,

$$\boldsymbol{v}(t) = \cos(\|\boldsymbol{b}\|[t-t_0])\|(\boldsymbol{v}_0)_\perp\|\boldsymbol{u}_1 + \sin(\|\boldsymbol{b}\|[t-t_0])\|(\boldsymbol{v}_0)_\perp\|\boldsymbol{u}_2 + \frac{\boldsymbol{b}\cdot\boldsymbol{v}_0}{\|\boldsymbol{b}\|}\boldsymbol{u}_3,$$

where $(\boldsymbol{v}_0)_\perp$ is the orthogonal component of $\boldsymbol{v}_0$ with respect to $\boldsymbol{b}$ and

$$\begin{aligned}\boldsymbol{u}_1 &= \frac{(\boldsymbol{v}_0)_\perp}{\|(\boldsymbol{v}_0)_\perp\|};\\ \boldsymbol{u}_2 &= \boldsymbol{u}_3 \times \boldsymbol{u}_1;\\ \boldsymbol{u}_3 &= \frac{\boldsymbol{b}(t_0)}{\|\boldsymbol{b}(t_0)\|}.\end{aligned}$$

We shall use the above theorem in a piecewise fashion in what follows. We therefore caution the reader that if $\boldsymbol{v}_0$, t_0, and $\boldsymbol{b}$ change, then the normalization and the vectors $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, and $\boldsymbol{u}_3$ change as well.

4.1. **Problem 1.**

Before we attempt to generalize Theorem 18, we consider a simple example of a "piecewise constant" vector $\boldsymbol{b}$. Let

$$\begin{aligned}\boldsymbol{b}(t) &= \begin{cases}(2,1,2) & \text{if } 0 < t \le \pi;\\ (1,2,2) & \text{if } \pi < t < 2\pi.\end{cases}\\ \boldsymbol{v}_0 &= (3,3,0).\end{aligned}$$

We shall find a solution to the rotation equation with the given $\boldsymbol{b}$. For $0 < t \le \pi$, the vector $\boldsymbol{b}$ is not a constant multiple of $\boldsymbol{v}_0$, and so

$$\boldsymbol{v}(t) = \cos(\|\boldsymbol{b}\|[t-t_0])\|(\boldsymbol{v}_0)_\perp\|\boldsymbol{u}_1 + \sin(\|\boldsymbol{b}\|[t-t_0])\|(\boldsymbol{v}_0)_\perp\|\boldsymbol{u}_2 + \frac{\boldsymbol{b}\cdot\boldsymbol{v}_0}{\|\boldsymbol{b}(0)\|}\boldsymbol{u}_3$$

---

[1] If you have not read the first three sections, a $\mathscr{C}^1$ function is a differentiable function whose derivative has continuous components

by Theorem 18. We have

$$\begin{aligned}
(\boldsymbol{v}_0)_\perp &= \boldsymbol{v}_0 - \left(\boldsymbol{v}_0 \cdot \frac{\boldsymbol{b}(0)}{\|\boldsymbol{b}(0)\|}\right)\frac{\boldsymbol{b}(0)}{\|\boldsymbol{b}(0)\|} \\
&= (3,3,0) - \frac{1}{9}(6+3+0)(2,1,2) \\
&= (3,3,0) - (2,1,2); \\
&= (1,2,-2); \\
\|(\boldsymbol{v}_0)_\perp\| &= 3; \\
\|\boldsymbol{b}(0)\| &= 3; \\
\boldsymbol{b}(0)\cdot\boldsymbol{v}_0 &= 3; \\
t_0 &= 0;
\end{aligned}$$

and so

$$\boldsymbol{v}(t) = 3\cos(3t)\boldsymbol{u}_1 + 3\sin(3t)\boldsymbol{u}_2 + 3\boldsymbol{u}_3$$

satisfies the rotation equation on $(0,\pi]$.

We now let

$$\boldsymbol{v}_1 = \boldsymbol{v}(\pi) = -3\boldsymbol{u}_1 + 3\boldsymbol{u}_3 = -(1,2,-2) + (2,1,2) = (1,-1,4).$$

For $\pi < t < 2\pi$, the vector $\boldsymbol{b}$ is not a constant multiple of $\boldsymbol{v}_1$, and so

$$\boldsymbol{v}(t) = \cos(\|\boldsymbol{b}(\pi)\|[t-t_0])\|(\boldsymbol{v}_0)_\perp\|\boldsymbol{u}_1 + \sin(\|\boldsymbol{b}(\pi)\|[t-t_0])\|(\boldsymbol{v}_0)_\perp\|\boldsymbol{u}_2 + \frac{\boldsymbol{b}(\pi)\cdot\boldsymbol{v}_0}{\|\boldsymbol{b}(\pi)\|}\boldsymbol{u}_3$$

by Theorem 18. We have

$$\begin{aligned}
(\boldsymbol{v}_1)_\perp &= \boldsymbol{v}_1 - \left(\boldsymbol{v}_1 \cdot \frac{\boldsymbol{b}(\pi)}{\|\boldsymbol{b}(\pi)\|}\right)\frac{\boldsymbol{b}(\pi)}{\|\boldsymbol{b}(\pi)\|} \\
&= (1,-1,4) - \frac{1}{9}(2-1+8)(2,1,2) \\
&= (1,-1,4) - (2,1,2); \\
&= (-1,-2,2); \\
\|(\boldsymbol{v}_1)_\perp\| &= 3; \\
\|\boldsymbol{b}\| &= 3; \\
\boldsymbol{b}\cdot\boldsymbol{v}_1 &= 3; \\
t_0 &= \pi;
\end{aligned}$$

and so

$$\boldsymbol{v}(t) = 3\cos(3(t-\pi))\boldsymbol{u}_1' + 3\sin(3(t-\pi))\boldsymbol{u}_2' + 3\boldsymbol{u}_3',$$

with

$$\begin{aligned}
\boldsymbol{u}_1' &= \frac{(\boldsymbol{v}_1)_\perp}{\|(\boldsymbol{v}_1)_\perp\|}; \\
\boldsymbol{u}_2' &= \boldsymbol{u}_3' \times \boldsymbol{u}_1'; \\
\boldsymbol{u}_3' &= \frac{\boldsymbol{b}(\pi)}{\|\boldsymbol{b}(\pi)\|}.
\end{aligned}$$

satisfies the rotation equation on $(\pi, 2\pi)$.

Let us check that the two solutions agree at $t = \pi$. Indeed,

$$\begin{aligned} 3\cos(3\pi)\boldsymbol{u}_1 + 3\sin(3\pi)\boldsymbol{u}_2 + 3\boldsymbol{u}_3 &= -(1,2,-2) + (2,1,2) \\ &= (1,-1,4) \\ &= (-1,-2,2) + (2,1,2) \\ &= 3\cos(3(\pi-\pi))\boldsymbol{u}_1' + 3\sin(3(\pi-\pi))\boldsymbol{u}_2' \\ &\quad +3\boldsymbol{u}_3', \end{aligned}$$

whence

$$\boldsymbol{v}(t) = \begin{cases} 3\cos(3t)\boldsymbol{u}_1 + 3\sin(3t)\boldsymbol{u}_2 + 3\boldsymbol{u}_3 & \text{if } 0 < t \leq \pi; \\ 3\cos(3(t-\pi))\boldsymbol{u}_1' + 3\sin(3(t-\pi))\boldsymbol{u}_2' + 3\boldsymbol{u}_3' & \text{if } \pi < t < 2\pi; \end{cases}$$

is the unique curve solving the rotation equation on $(0, 2\pi)$ for the given $\boldsymbol{b}(t)$ with the initial condition $\boldsymbol{v}(0) = (3,3,0)$. □

4.2. **Problem 2.** We now suppose that $\boldsymbol{b}(t)$ is piecewise constant with finitely many discontinuities. Applying Theorem 18 to each part of $\boldsymbol{b}(t)$ furnishes a solution, which, by construction, is well-defined at each point of discontinuity. This solution is unique, for any other solution much agree with what Theorem 18 furnishes at each part of $\boldsymbol{b}(t)$, whence it must agree everywhere. □

4.3. **Problem 3.** The general version of Theorem 18 is the following:

Theorem 19. *For any vector $\boldsymbol{v}_0 \in \mathbb{R}^3$ and a piecewise $\mathscr{C}^1$ vector-valued function into $\mathbb{R}^3$, there exists precisely one curve $\boldsymbol{v}$ satisfying the rotation equation for $\boldsymbol{b}$ with the initial condition $\boldsymbol{v}(0) = \boldsymbol{v}_0$.*

We will not prove this theorem, but we will prove a crucial lemma:

Lemma 20. *Fix $\boldsymbol{v}_0 \in \mathbb{R}^3$. Let $\boldsymbol{b}_1(t)$ and $\boldsymbol{b}_2(t)$ be two piecewise $\mathscr{C}^1$ vector-valued functions into $\mathbb{R}^3$ with only finitely many discontinuities, and $\boldsymbol{v}_1(t)$ and $\boldsymbol{v}_2(t)$ be the piecewise differentiable solutions of the rotation equations for $\boldsymbol{b}_1(t)$ and $\boldsymbol{b}_2(t)$, respectively, with the initial condition $\boldsymbol{v}_i(0) = \boldsymbol{v}_0$; we are, of course, assuming for a moment that the solutions exist. If $\varepsilon > 0$ satisfies $\|\boldsymbol{b}_1(t) - \boldsymbol{b}_2(t)\| \leq \varepsilon$ for all $0 \leq t \leq 1$, then*

$$\|\boldsymbol{v}_1(t) - \boldsymbol{v}_2(t)\| \leq \varepsilon \|\boldsymbol{v}_0\|.$$

Proof of Lemma 20. Let $\boldsymbol{w}(t) = \boldsymbol{v}_1(t) - \boldsymbol{v}_2(t)$. We shall compute

$$\frac{d}{dt}\|\boldsymbol{w}(t)\|^2.$$

We first note that

$$\begin{aligned} \boldsymbol{w}'(t) &= \boldsymbol{v}_1'(t) - \boldsymbol{v}_2'(t) \\ &= \boldsymbol{b}_1(t) \times \boldsymbol{v}_1(t) - \boldsymbol{b}_2 \times \boldsymbol{v}_2(t) \\ &= \boldsymbol{b}_1(t) \times (\boldsymbol{v}_1(t) - \boldsymbol{v}_2(t)) + \boldsymbol{b}_1 \times \boldsymbol{v}_2(t) - \boldsymbol{b}_2(t) \times \boldsymbol{v}_2(t) \\ &= \boldsymbol{b}_1(t) \times (\boldsymbol{v}_1(t) - \boldsymbol{v}_2(t)) + (\boldsymbol{b}_1(t) - \boldsymbol{b}_2(t)) \times \boldsymbol{v}_2(t) \\ &= \boldsymbol{b}_1(t) \times \boldsymbol{w}(t) + (\boldsymbol{b}_1(t) - \boldsymbol{b}_2(t)) \times \boldsymbol{v}_2(t). \end{aligned}$$

□

Therefore,

$$\begin{aligned}\frac{d}{dt}\|\boldsymbol{w}(t)\|^2 &= \frac{d}{dt}(\boldsymbol{w}(t)\cdot\boldsymbol{w}(t))\\ &= 2\boldsymbol{w}'(t)\cdot\boldsymbol{w}(t)\\ &= 2[\boldsymbol{b}_1(t)\times\boldsymbol{w}(t)+(\boldsymbol{b}_1(t)-\boldsymbol{b}_2(t))\times\boldsymbol{v}_2(t)]\cdot\boldsymbol{w}(t)\\ &= 2[(\boldsymbol{b}_1(t)\times\boldsymbol{w}(t))\cdot\boldsymbol{w}(t)+((\boldsymbol{b}_1(t)-\boldsymbol{b}_2(t))\times\boldsymbol{v}_2(t))\cdot\boldsymbol{w}(t)].\end{aligned}$$

Since $\boldsymbol{b}_1(t)\times\boldsymbol{w}\perp\boldsymbol{w}$, we have

$$\begin{aligned}\frac{d}{dt}\|\boldsymbol{w}(t)\|^2 &= 2[(\boldsymbol{b}_1(t)\times\boldsymbol{w}(t))\cdot\boldsymbol{w}(t)+((\boldsymbol{b}_1(t)-\boldsymbol{b}_2(t))\times\boldsymbol{v}_2(t))\cdot\boldsymbol{w}(t)]\\ &= 2(\boldsymbol{b}_1(t)-\boldsymbol{b}_2(t))\times\boldsymbol{v}_2(t))\cdot\boldsymbol{w}(t),\end{aligned}$$

whence by the Cauchy-Schwarz inequality

$$\begin{aligned}\left|\frac{d}{dt}\|\boldsymbol{w}(t)\|^2\right| &\leq 2\|(\boldsymbol{b}_1(t)-\boldsymbol{b}_2(t))\times\boldsymbol{v}_2(t)\|\|\boldsymbol{w}(t)\|\\ &\leq 2\|\boldsymbol{b}_1(t)-\boldsymbol{b}_2(t)\|\|\boldsymbol{v}_2(t)\|\sin\theta\|\boldsymbol{w}(t)\|\\ &\leq 2\varepsilon\|\boldsymbol{v}_2(t)\|\|\boldsymbol{w}(t)\|.\end{aligned}$$

By the chain rule, we also have

$$\frac{d}{dt}\|\boldsymbol{w}(t)\|^2 = 2\|\boldsymbol{w}(t)\|\frac{d}{dt}\|\boldsymbol{w}(t)\|,$$

whence we have

$$\left|\frac{d}{dt}\|\boldsymbol{w}(t)\|\right| \leq \varepsilon\|\boldsymbol{v}_2(t)\|.$$

Now we observe that

$$\frac{d}{dt}\|\boldsymbol{v}_2(t)\|^2 = 2\boldsymbol{v}_2'(t)\cdot\boldsymbol{v}_2(t) = 2(\boldsymbol{b}_2(t)\times\boldsymbol{v}_2(t))\cdot\boldsymbol{v}_2(t) = 0,$$

whence $\|\boldsymbol{v}_2(t)\| = \|\boldsymbol{v}_2(0)\| = \|\boldsymbol{v}_0\|$, and so

$$\left|\frac{d}{dt}\|\boldsymbol{w}(t)\|\right| \leq \varepsilon\|\boldsymbol{v}_0\|$$

as was to be shown.

Since $\|\boldsymbol{w}(0)\| = 0$, it now follows from the fundamental theorem of calculus that

$$\begin{aligned}\|\boldsymbol{w}(t)\| &= \left|\|\boldsymbol{w}(0) + \int_0^t \frac{d}{ds}\|\boldsymbol{w}(s)\|\,ds\right|\\ &\leq \int_0^t \left|\frac{d}{ds}\|\boldsymbol{w}(s)\|\right|\,ds\\ &\leq \int_0^t \varepsilon\|\boldsymbol{v}_0\|\,ds\\ &\leq \varepsilon\|\boldsymbol{v}_0\|,\end{aligned}$$

as was to be shown. □

4.4. **Problem 4.** We now suppose that $\boldsymbol{b}(t)$ is a $\mathscr{C}^1$ curve on $[0,1]$ mapping into $\mathbb{R}^3$. Since $\boldsymbol{b}'(t)$ is continuous, we can find a constant C such that $\|\boldsymbol{b}'(t)\| \leq C$ for all $t \in [0,1]$, whence, for any $(t, t_0) \subseteq (0,1)$, we have

$$\|\boldsymbol{v}(t) - \boldsymbol{v}(t_0)\| = \left\| \int_{t_0}^{t} \boldsymbol{b}'(s)\, ds \right\| \leq \int_{t_0}^{t} \|\boldsymbol{b}'(s)\|\, ds \leq C(t - t_0).$$

We fix a positive integer k and define a piecewise constant function

$$\boldsymbol{b}_1(t) = \boldsymbol{b}\left(\frac{j}{2^k}\right) \qquad \text{for} \qquad \frac{j}{2^k} \leq t < \frac{j+1}{2^k}.$$

We let $\boldsymbol{b}_2 = \boldsymbol{b}$ and claim that

$$\|\boldsymbol{b}_1(t) - \boldsymbol{b}_2(t)\| \leq C2^{-k}.$$

Indeed,

$$\boldsymbol{b}_2(t) = \boldsymbol{b}_1(t) + (t - t_0)\boldsymbol{b}'(t_0) + \mathcal{O}((t - t_0)^2)$$

by Taylor's theorem, whence

$$\|\boldsymbol{b}_1(t) - \boldsymbol{b}_2(t)\| \leq \|t - t_0\| \|\boldsymbol{b}'(t_0)\| + \|\mathcal{O}((t - t_0)^2)\| \leq C2^{-k} + M2^{-2k}$$

for some constant M. We can therefore bound this by $C_1 2^{-k}$, where C_1 is "slightly larger" than C.

By Lemma 20, the solutions $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ associated to $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$, respectively, satisfy the following inequality on $[0,1]$:

$$\|\boldsymbol{v}_1(t) - \boldsymbol{v}_2(t)\| \leq C_1 2^{-k} \|\boldsymbol{v}_0\|.$$

This estimate can be made as small as desired by enlarging k, hence $\boldsymbol{v}_1$ is a "very close" approximation of the true solution $\boldsymbol{v}_2$. Furthermore, $\boldsymbol{v}_1$ can be constructed explicitly via a repeated application of Theorem 18. □

References

[Carlen] Carlen, Eric. *Multivariable Calculus, Linear Algebra and Differential Equations.*
[Munkres] Munkres, James. *Analysis on Manifolds.*
[Rudin] Rudin, Walter. *Principles of Mathematical Analysis* (3e).
[Spivak] Spivak, Michael *Calculus on Manifolds.*