FALL 2011 MATH 291 WORKSHOP 1

MARK H. KIM

ABSTRACT. This is a set of notes from the first workshop for Math 291, which took place in September 12, 2011. The first three sections expand on the material from the lecture and are included here for your general edification (and entertainment!) and may be freely skipped if you so desire. The fourth section contains suggested solutions to the challenge problems.

CONTENTS

1. SETS AND FUNCTIONS

A *set* is a clearly-defined collection of mathematical objects: every mathematical object is either in the set, or not in the set. Formally, a set is defined to be any object that satisfies the Zermelo-Fraenkel axioms (ZF)[1]. This is just like saying that money is any object that satisfies certain rules in economics: it could be a coin, paper money, electric signals from your credit card, or a mango. This sort of "functional" definition is at the heart of many subjects that require abstract thinking: instead of sweating over what something is, we instead consider the properties and study that "something" by working with the properties.

Let X and Y be sets. Recall that a *function* $f : X \to Y$ is a "rule" that assigns a unique value $f(x) \in Y$ to each and every element $x \in X$.[2] This means that f assigns an element of Y to each element of X, but not every element of Y is realized as a value of f. The set X is called the *domain* of the function f, and Y the *codomain* of f. The collection of each $f(x)$ is a subset of Y, called the *image* of f. f is *one-to-one*, or *injective*, if $f(x) = f(y)$ implies $x = y$: that is, given any fixed element of Y, only one element of X can have it as the value of the function f. f is *onto*, or *surjective*, if every element $y \in Y$ has a corresponding $x \in X$ such

Date: September 18, 2011. *Revised*: September 21, 2011.

[1]We will not state these axioms here. A course in axiomatic set theory, such as Math 361, will cover this topic.

[2]Formally, a function is a special type of *relation*; we do not pursue this viewpoint here.

that $f(x) = y$. In other words, f is surjective if its image agrees with its codomain. If f is both injective and surjective, then it is *bijective*.

A *Cartesian product* of X and Y is the set

$$\{(x, y) : x \in X \text{ and } y \in Y\}$$

of ordered pairs. With this in mind, we define a *binary operation* on the set X to be a function $* : X \times X \to X$, which we often denote by writing $*(x_1, x_2) = x_! * x_2$. This is to say, a binary operation takes two elements of a set and produces another element of the set.

A simple example of a binary operation is the multiplication operation on $\mathbb{R}$. Let us consider a less trivial example. We let X, Y, and Z be sets, and $f : X \to Y$ and $g : Y \to Z$ be functions. The *composition* f and g is a function $g \circ f : X \to Z$ defined by setting

$$(g \circ f)(x) = g(f(x))$$

for each $x \in X$. Note that *the composition of f and g* is written from right to left: $f \circ g$.

Consider now the set $\text{Maps}(X, X)$ of all functions from X to itself. Given any two functions $f, g \in \text{Maps}(X, X)$, we see that the composition $g \circ f$ is also a function from X to itself, hence in $\text{Maps}(X, X)$. Therefore, the "composition operation" $\circ$ is a binary operation on $\text{Maps}(X, X)$.

2. Linearity

We now restrict our attention to $\mathbb{R}^n$, the n-dimensional Euclidean space. Recall that "linear functions" on $\mathbb{R}$ are of the form

$$f(x) = ax + b.$$

For the sake of simplicity, we consider only the lines going through the origin[3]:

$$f(x) = ax.$$

Note that the above function satisfies the following properties:

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

In fact, if you take *any* function satisfying (i) and (ii), you will end up with a function of the form

$$f(x) = ax,$$

To see this, we let $a = f(1)$, and observe that[4]

$$f(x) = f(x \cdot 1) = xf(1) = xa = ax.$$

Taking a cue from the above characterization of linear functions, we extend the notion to higher dimensions:

Definition 1. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *linear* if

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

---

[3] Functions of the form $f(x) = ax + b$ are known as *affine transformations*, which have slightly different theory of functions associated to them

[4] Note that we did not make use of (i) at all here; this is because (ii) subsumes (i) in the one-dimensional case. (i) is nevertheless necessary for high-dimensional cases.

For example,

$$f(x, y) = x + y$$

is a linear function from $\mathbb{R}^2$ to $\mathbb{R}$. In general,

$$f(x, y) = ax + by$$

is a linear function from $\mathbb{R}^2$ to $\mathbb{R}$, where a and b are any real numbers. If a and b are not both zero, then the image of f is a plane in the coordinate space.

The field of mathematics that deals with linear functions is known as *linear algebra*. For one, it is a basic fact in linear algebra that every linear function on a Euclidean space can be written as

$$f(\boldsymbol{x}) = A \cdot \boldsymbol{x},$$

where A is a matrix, and $\cdot$ denotes the matrix-vector multiplication. If we restrict ourselves to the one-dimensional Euclidean space, then we obtain the familiar fact that every linear function is of the form

$$f(x) = ax.$$

Matrix representations of linear functions come with an array of extremely useful computational principles. Nevertheless, we shall stick to the abstract approach of considering linear functions as "arbitrary functions satisfying certain rules."

Why do we care about linear functions? They are extremely nice functions, after all. One of the key ideas of calculus is to approximate ugly functions with nicer functions, and thus linear functions make frequent appearance. In fact, the process of differentiation may be described as finding the linear approximation of a function.

Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We say that f is *differentiable* on $\mathbb{R}$ if, for each $x \in \mathbb{R}$, the *derivative*

$$f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}.$$

exists. At each point $x_0 \in \mathbb{R}$, this computation gives rise to *the tangent line*

$$y = f'(x_0)(x - x_0) + f(x_0),$$

which is "pretty close[5]" to the original function f near x_0. This is to say that a differentiable function has an approximation by a "linear function" at every point.

In other words, f is differentiable if and only if, for each $x \in \mathbb{R}$, there exists a linear function $L_x : \mathbb{R} \to \mathbb{R}$ and a "remainder function" $R_x : \mathbb{R} \to \mathbb{R}$ such that

$$f(x + h) = f(x) + L_x(h) + R_x(h) \tag{1}$$

for all small $h \in \mathbb{R}$, and

$$\lim_{h \to 0} \frac{|R_x(h)|}{|h|} = 0.$$

To see why the second condition is there, we divide both sides of equation (1) by h and rearrange:

$$\frac{f(x + h) - f(x)}{h} = \frac{L_x(h)}{h} + \frac{R_x(h)}{h}.$$

If f is differentiable, then we may set $L_x(h) = f'(x)h$ and observe that

$$\frac{f(x + h) - f(x)}{h} = f'(x) + \frac{R_x(h)}{h},$$

[5]This is a consequence of Taylor's theorem

whence $|R_x(h)|/|h|$ should go to 0 as h tends to zero.

This observation generalizes naturally to higher dimensions:

Definition 2. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a point $\boldsymbol{x} \in \mathbb{R}^n$ if there exist a linear function $L : \mathbb{R}^m \to \mathbb{R}^n$ and a "remainder function" $R : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$f(\boldsymbol{x}+\boldsymbol{h}) = f(\boldsymbol{x}) + L(\boldsymbol{h}) + R(\boldsymbol{h}) \tag{2}$$

for all $\boldsymbol{h} \in \mathbb{R}^m$ sufficiently close to $\mathbf{0}$,[6] and

$$\lim_{\boldsymbol{h}\to 0} \frac{\|R(\boldsymbol{h})\|}{\|\boldsymbol{h}\|} = 0.$$

We typically write $Df(\boldsymbol{x})$ to denote the linear function T. f is *differentiable* on $\mathbb{R}^n$ if it is differentiable at every point of $\mathbb{R}^n$.

We shall have more to say about this definition later in the semester.

3. Bilinearity

Recall the distributive property of the real numbers, which states that

$$(a+b)c = ac + bc$$

for all $a, b, c \in \mathbb{R}$. Since multiplication is technically a binary operation $\times : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$, we could write

$$\times(a+b, c) = \times(a, c) + \times(b, c)$$

instead. Similarly, we have

$$\times(a, b+c) = \times(a, b) + \times(a, c).$$

We are often interested in binary operations that are linear in each variable, if only to have the nice "distributive" property as above. Formally:

Definition 3. A function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^l$ is *bilinear* if

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

A *bilinear operation* on $\mathbb{R}^n$ is a bilinear function from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^n$.

Note that $f(\boldsymbol{x}, \boldsymbol{y})$ is a linear function of $\boldsymbol{x}$ for a fixed $\boldsymbol{y}$, and a linear function of $\boldsymbol{y}$ for a fixed $\boldsymbol{x}$. This is what we mean by "linear in each variable." Sometimes we want our "bilinear operation" to return a scalar instead of a vector. A slight modification of the above definition gives the following:

Definition 4. A *bilinear form* on $\mathbb{R}^n$ is a bilinear function from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$.

The ordinary multiplication on $\mathbb{R}$ is a bilinear operation on $\mathbb{R}$. Since multiplying to real numbers yields a real number, the ordinary multiplication is also a bilinear form on $\mathbb{R}$. The *dot product*

$$(x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) = x_1 y_x + \cdots + x_n y_n$$

[6] "Sufficiently close to $\mathbf{0}$" means we can find a constant $M > 0$ such that (2) is true for all $h \in \mathbb{R}^n$ satisfying $\|h\| < M$.

is an example of a bilinear form on $\mathbb{R}^n$. The *cross product*

$$(x_1, x_2, x_3) \times (y_1, y_2, y_3) = (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1)$$

is an example of a bilinear operation on $\mathbb{R}^3$.[7]

We say that a bilinear function f is *symmetric* if

$$f(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{y}, \boldsymbol{x}),$$

and *skew-symmetric* if

$$f(\boldsymbol{x}, \boldsymbol{y}) = -f(\boldsymbol{x}, \boldsymbol{y}).$$

The ordinary multiplication on $\mathbb{R}$ and the dot product on $\mathbb{R}^n$ are symmetric; the cross product is skew-symmetric. There is a nice theory of symmetric and skew-symmetric bilinear forms, but we do not pursue it here.

4. Suggested Solutions to Challenge Problems

Recall the definitions of the parallel and orthogonal components of a vector $\boldsymbol{x} \in \mathbb{R}^n$ in the direction of $\boldsymbol{u} \in \mathbb{R}^n$:

$$\begin{aligned} \boldsymbol{x}_\parallel &= (\boldsymbol{x} \cdot \boldsymbol{u})\boldsymbol{u}; \\ \boldsymbol{x}_\perp &= \boldsymbol{x} - \boldsymbol{x}_\parallel = \boldsymbol{x} - (\boldsymbol{x} \cdot \boldsymbol{u})\boldsymbol{u}. \end{aligned}$$

With the above in mind, the *Householder reflection* of $\boldsymbol{x}$ along the direction $\boldsymbol{u}$ is

$$h_{\boldsymbol{u}}(\boldsymbol{x}) = \boldsymbol{x}_\perp - \boldsymbol{x}_\parallel = \boldsymbol{x} - 2\boldsymbol{x}_\parallel = \boldsymbol{x} - 2(\boldsymbol{x} \cdot \boldsymbol{u})\boldsymbol{u}. \tag{3}$$

We also recall that the cross product of two vectors is the zero vector if one vector is a scalar multiple of the other. Indeed, if $\boldsymbol{x} = (x_1, x_2, x_3)$ and $\boldsymbol{y} = \lambda(x_1, x_2, x_3)$, then

$$\begin{aligned} \boldsymbol{x} \times \boldsymbol{y} &= (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1) \\ &= (\lambda x_2x_3 - \lambda x_3x_2, \lambda x_3x_1 - \lambda x_1x_3, \lambda x_1x_2 - \lambda x_2x_1) \\ &= \mathbf{0}. \end{aligned}$$

In particular, this means that two parallel vectors have trivial cross product.

Conversely, if $\boldsymbol{x} \times \boldsymbol{y} = \mathbf{0}$, then the two vectors are scalar multiples of each other. Indeed, if $\boldsymbol{x} = (x_1, x_2, x_3)$ and $\boldsymbol{y} = (y_1, y_2, y_3)$, then $\boldsymbol{x} \times \boldsymbol{y} = \mathbf{0}$ implies that

$$\begin{aligned} x_2y_3 &= x_3y_2 \\ x_3y_1 &= x_1y_3 \\ x_1y_2 &= x_2y_1, \end{aligned}$$

[7] Unlike the dot product, the cross product is only defined in the three-dimensional Euclidean space. The generalization of the cross product, known as the *wedge product*, is typically not a binary operation. Wedge products appear in the study of differential forms, which is required to define integration on more general surfaces. If you are a math major, you will encounter them in an upper-division course in multivariable calculus (Math 412 at Rutgers) or a course in differential geometry (such as Math 432 or Math 532). If you are a physics major, you will encounter them in a mathematical physics course.

or

$$\frac{x_3}{x_2} = \frac{y_3}{y_2}$$
$$\frac{x_3}{x_1} = \frac{y_3}{y_1}$$
$$\frac{x_1}{x_2} = \frac{y_1}{y_2}.$$

The ratios of the components are equal, and so the vectors are scalar multiples of each other.

4.1. **Problem 1 (4 points).** Recall that

$$h_{\boldsymbol{u}}(\boldsymbol{a}) \times h_{\boldsymbol{u}}(\boldsymbol{b}) = (\boldsymbol{a}_\perp - \boldsymbol{a}_\parallel) \times (\boldsymbol{b}_\perp - \boldsymbol{b}_\parallel).$$

By the bilinearity[8] of the cross product,

$$(\boldsymbol{a}_\perp - \boldsymbol{a}_\parallel) \times (\boldsymbol{b}_\perp - \boldsymbol{b}_\parallel) = \boldsymbol{a}_\perp \times \boldsymbol{b}_\perp - \boldsymbol{a}_\parallel \times \boldsymbol{b}_\perp - \boldsymbol{a}_\perp \times \boldsymbol{b}_\parallel + \boldsymbol{a}_\parallel \times \boldsymbol{b}_\parallel.$$

Both $\boldsymbol{a}_\parallel$ and $\boldsymbol{b}_\parallel$ are parallel to $\boldsymbol{u}$, hence are parallel to each other. Therefore, $\boldsymbol{a}_\parallel \times \boldsymbol{b}_\parallel = 0$, and so

$$h_{\boldsymbol{u}}(\boldsymbol{a}) \times h_{\boldsymbol{u}}(\boldsymbol{b}) = \boldsymbol{a}_\perp \times \boldsymbol{b}_\perp - \boldsymbol{a}_\parallel \times \boldsymbol{b}_\perp - \boldsymbol{a}_\perp \times \boldsymbol{b}_\parallel.$$

We also recall that

$$\boldsymbol{a} \times \boldsymbol{b} = (\boldsymbol{a}_\perp + \boldsymbol{a}_\parallel) \times (\boldsymbol{b}_\perp + \boldsymbol{b}_\parallel).$$

Once again, bilinearity yields

$$(\boldsymbol{a}_\perp + \boldsymbol{a}_\parallel) \times (\boldsymbol{b}_\perp + \boldsymbol{b}_\parallel) = \boldsymbol{a}_\perp \times \boldsymbol{b}_\perp + \boldsymbol{a}_\perp \times \boldsymbol{b}_\parallel + \boldsymbol{a}_\parallel \times \boldsymbol{b}_\perp + \boldsymbol{a}_\parallel \times \boldsymbol{b}_\parallel.$$

Since $\boldsymbol{a}_\parallel \times \boldsymbol{b}_\parallel = 0$, we have

$$\boldsymbol{a} \times \boldsymbol{b} = \boldsymbol{a}_\perp \times \boldsymbol{b}_\perp + \boldsymbol{a}_\perp \times \boldsymbol{b}_\parallel + \boldsymbol{a}_\parallel \times \boldsymbol{b}_\perp.$$

We now establish a key property of the Householder reflection

Lemma 1 (Linearity of Householder reflection)**.** *The Householder reflection is linear. That is, for any pair of three-dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and a scalar λ we have*

(i) $h_{\boldsymbol{u}}(\boldsymbol{x} + \boldsymbol{y}) = h_{\boldsymbol{u}}(\boldsymbol{x}) + h_{\boldsymbol{u}}(\boldsymbol{y})$*;*
(ii) $h_{\boldsymbol{u}}(\lambda \boldsymbol{x}) = \lambda h_{\boldsymbol{u}}(\boldsymbol{x})$*.*

Proof of lemma. Observe that

$$\begin{aligned} h_{\boldsymbol{u}}(\boldsymbol{x} + \boldsymbol{y}) &= (\boldsymbol{x} + \boldsymbol{y}) - 2((\boldsymbol{x} + \boldsymbol{y}) \cdot \boldsymbol{u})\boldsymbol{u} \\ &= \boldsymbol{x} + \boldsymbol{y} - 2(\boldsymbol{x} \cdot \boldsymbol{u} + \boldsymbol{y} \cdot \boldsymbol{u}) \cdot \boldsymbol{u} \\ &= \boldsymbol{x} + \boldsymbol{y} - 2(\boldsymbol{x} \cdot \boldsymbol{u}) \cdot \boldsymbol{u} + -2(\boldsymbol{y} \cdot \boldsymbol{u}) \cdot \boldsymbol{u} \\ &= \boldsymbol{x} - 2(\boldsymbol{x} \cdot \boldsymbol{u}) \cdot \boldsymbol{u} + \boldsymbol{y} - 2(\boldsymbol{y} \cdot \boldsymbol{u}) \cdot \boldsymbol{u} \\ &= h_{\boldsymbol{u}}(\boldsymbol{x}) + h_{\boldsymbol{u}}(\boldsymbol{x}). \end{aligned}$$

The proof of (ii) is similar. □

Corollary 2. *If $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are three-dimensional vectors, then*

$$h_{\boldsymbol{u}}(\boldsymbol{x}_1 + \cdots + \boldsymbol{x}_n) = h_{\boldsymbol{u}}(\boldsymbol{x}_1) + \cdots + h_{\boldsymbol{u}}(\boldsymbol{x}_1)$$

[8] If you have not read through sections 1-3, just think of this as "expansion is done in an obvious way, without disturbing the order."

Proof of corollary. Apply (i) of linearity $n-1$ times. □

We shall refer to both Lemma 1 and Corollary 2 as "linearity of the Householder reflection." By linearity, we have

$$\begin{aligned} h_{\boldsymbol{u}}(\boldsymbol{a}\times\boldsymbol{b}) &= h_{\boldsymbol{u}}(\boldsymbol{a}_\perp\times\boldsymbol{b}_\perp+\boldsymbol{a}_\perp\times\boldsymbol{b}_\parallel+\boldsymbol{a}_\parallel+\boldsymbol{b}_\perp) \\ &= h_{\boldsymbol{u}}(\boldsymbol{a}_\perp\times\boldsymbol{b}_\perp)+h_{\boldsymbol{u}}(\boldsymbol{a}_\perp\times\boldsymbol{b}_\parallel)+h_{\boldsymbol{u}}(\boldsymbol{a}_\parallel\times\boldsymbol{b}_\perp). \end{aligned}$$

We need another computational observation:

Lemma 3. *If $\boldsymbol{x}$ is orthogonal to $\boldsymbol{u}$, then $h_{\boldsymbol{u}}(\boldsymbol{x})=\boldsymbol{x}$. If $\boldsymbol{x}$ is parallel to $\boldsymbol{u}$, then $h_{\boldsymbol{u}}(\boldsymbol{x})=-\boldsymbol{x}$.*

Proof of lemma. If $\boldsymbol{x}$ and $\boldsymbol{u}$ are orthogonal to each other, then

$$h_{\boldsymbol{u}}(\boldsymbol{x})=\boldsymbol{x}-2(\boldsymbol{x}\cdot\boldsymbol{u})\boldsymbol{u}=\boldsymbol{x}-\boldsymbol{0}=\boldsymbol{x}.$$

If $\boldsymbol{x}$ and $\boldsymbol{u}$ are parallel, then, for some scalar λ, we have $\boldsymbol{u}=\lambda\boldsymbol{x}$, and so

$$\begin{aligned} h_{\boldsymbol{u}}(\boldsymbol{x}) &= \boldsymbol{x}-2(\boldsymbol{x}\cdot\boldsymbol{u})\boldsymbol{u} \\ &= \boldsymbol{x}-2(\boldsymbol{x}\cdot\lambda\boldsymbol{x})\lambda\boldsymbol{x} \\ &= \boldsymbol{x}-2\lambda^2\|\boldsymbol{x}\|^2\boldsymbol{x}. \end{aligned}$$

Since $\boldsymbol{u}=\lambda\boldsymbol{x}$, we have $\|\boldsymbol{u}\|=\lambda\|\boldsymbol{x}\|$, or

$$\|\boldsymbol{x}\|=\frac{\|\boldsymbol{u}\|}{\lambda}.$$

Therefore,

$$h_{\boldsymbol{u}}(\boldsymbol{x})=\boldsymbol{x}-2\|\boldsymbol{u}\|^2\boldsymbol{x}.$$

Recalling that the direction vector $\boldsymbol{u}$ is a unit vector, we now see that

$$h_{\boldsymbol{u}}(\boldsymbol{x})=\boldsymbol{x}-2\boldsymbol{x}=-\boldsymbol{x},$$

as was claimed. □

We now proceed with the computation. $\boldsymbol{a}_\perp\times\boldsymbol{b}_\parallel$ is orthogonal to $\boldsymbol{b}_\parallel$, and $\boldsymbol{b}_\parallel$ is parallel to $\boldsymbol{u}$, so $\boldsymbol{a}_\perp\times\boldsymbol{b}_\parallel$ is orthogonal to $\boldsymbol{u}$. Similarly, $\boldsymbol{a}_\parallel\times\boldsymbol{b}_\perp$ is orthogonal to $\boldsymbol{u}$. $\boldsymbol{a}_\perp\times\boldsymbol{b}_\perp$ is, however, parallel to $\boldsymbol{u}$, for $\boldsymbol{a}_\perp\times\boldsymbol{b}_\perp$ is orthogonal to $\boldsymbol{a}_\perp$ and $\boldsymbol{b}_\perp$, both of which are orthogonal to $\boldsymbol{u}$.

It thus follows that

$$\begin{aligned} h_{\boldsymbol{u}}(\boldsymbol{a}\times\boldsymbol{b}) &= h_{\boldsymbol{u}}(\boldsymbol{a}_\perp\times\boldsymbol{b}_\perp)+h_{\boldsymbol{u}}(\boldsymbol{a}_\perp\times\boldsymbol{b}_\parallel)+h_{\boldsymbol{u}}(\boldsymbol{a}_\parallel\times\boldsymbol{b}_\perp) \\ &= -\boldsymbol{a}_\perp\times\boldsymbol{b}_\perp+\boldsymbol{a}_\perp\times\boldsymbol{b}_\parallel+\boldsymbol{a}_\parallel\times\boldsymbol{b}_\perp \\ &= -(\boldsymbol{a}_\perp\times\boldsymbol{b}_\perp-\boldsymbol{a}_\perp\times\boldsymbol{b}_\parallel-\boldsymbol{a}_\parallel\times\boldsymbol{b}_\perp) \\ &= -h_{\boldsymbol{u}}(\boldsymbol{a})\times h_{\boldsymbol{u}}(\boldsymbol{b}), \end{aligned}$$

as was to be shown.

Suppose now that $\{u_1,u_2,u_3\}$ is a right-handed orthonormal basis. We claim that $\mathcal{H}=\{u_1,u_2,u_3\}$ is a left-handed orthonormal basis. To this end, we must show that (1) $\mathcal{H}$ is orthonormal, and (2) $\mathcal{H}$ is a left-handed system. We shall need a lemma:

Lemma 4 (Orthogonality of Householder reflection)**.** *$h_{\boldsymbol{u}}$ is orthogonal, viz.,*

$$(h_{\boldsymbol{u}}(x))\cdot(h_{\boldsymbol{u}}(y))=x\cdot y.$$

Proof of lemma. We just do the computation:

$$\begin{aligned}(h_{\boldsymbol{u}}(x))\cdot(h_{\boldsymbol{u}}(y)) &= (\boldsymbol{x}-2\boldsymbol{x}_\parallel)\cdot(\boldsymbol{y}-2\boldsymbol{y}_\parallel)\\ &= \boldsymbol{x}\cdot\boldsymbol{y}-2\boldsymbol{x}\cdot\boldsymbol{y}_\parallel-2\boldsymbol{x}_\parallel\cdot\boldsymbol{y}+4\boldsymbol{x}_\parallel\cdot\boldsymbol{y}_\parallel\\ &= \boldsymbol{x}\cdot\boldsymbol{y}-2(\boldsymbol{x}_\parallel+\boldsymbol{x}_\perp)\cdot\boldsymbol{y}_\parallel-2\boldsymbol{x}_\parallel\cdot(\boldsymbol{y}_\parallel+\boldsymbol{y}_\perp)+4\boldsymbol{x}_\parallel\cdot\boldsymbol{y}_\parallel\\ &= \boldsymbol{x}\cdot\boldsymbol{y}-2\boldsymbol{x}_\perp\cdot\boldsymbol{y}_\parallel-2\boldsymbol{x}_\parallel\cdot\boldsymbol{y}_\perp\\ &= \boldsymbol{x}\cdot\boldsymbol{y}.\end{aligned}$$

□

Since the Householder reflection preserves the dot product, we conclude that $\mathcal{H}$ is an orthonormal basis. It now suffices to observe that

$$[(h_{\boldsymbol{u}}(\boldsymbol{u}_1)\times(h_{\boldsymbol{u}}(\boldsymbol{u}_2)]\cdot(h_{\boldsymbol{u}}(\boldsymbol{u}_3)=-(h_{\boldsymbol{u}}(\boldsymbol{u}_1\times\boldsymbol{u}_2)\cdot(h_{\boldsymbol{u}}(\boldsymbol{u}_3)=-(\boldsymbol{u}_1\times\boldsymbol{u}_2)\cdot\boldsymbol{u}_3=-1,$$

whence $\mathcal{H}$ is left-handed. □

4.2. **Problem 2 (3 points + 1 extra-credit point).** We suppose for a contradiction that $\boldsymbol{u}\times\boldsymbol{v}=0$. This implies that $\boldsymbol{u}$ is a constant multiple of $\boldsymbol{v}$. Since both are unit vectors, we have

$$\boldsymbol{u}=\pm\boldsymbol{v}.$$

If $\boldsymbol{u}=\boldsymbol{v}$, then $h_{\boldsymbol{u}}=h_{\boldsymbol{v}}$, and so

$$\boldsymbol{v}_i=h_{\boldsymbol{v}}(h_{\boldsymbol{u}}(\boldsymbol{u}_i))=\boldsymbol{u}_i,$$

for $i=1,2,3$. This is evidently absurd, for $\{\boldsymbol{u}_1,\boldsymbol{u}_2,\boldsymbol{u}_3\}$ and $\{\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3\}$ were assumed to be distinct. If $\boldsymbol{u}=-\boldsymbol{v}$, then

$$h_{\boldsymbol{v}}(\boldsymbol{x})=\boldsymbol{x}-2(\boldsymbol{x}\cdot\boldsymbol{v})\boldsymbol{v}=\boldsymbol{x}-2(\boldsymbol{x}\cdot-\boldsymbol{u})(-\boldsymbol{u})=\boldsymbol{x}-2(\boldsymbol{x}\cdot\boldsymbol{u})\boldsymbol{u}=h_{\boldsymbol{u}}(\boldsymbol{x})$$

for all $\boldsymbol{x}\in\mathbb{R}^3$. Therefore, $h_{\boldsymbol{v}}=h_{\boldsymbol{u}}$, which results in the same contradiction as above. We thus conclude that $\boldsymbol{u}\times\boldsymbol{v}\neq\boldsymbol{0}$.

Set

$$\boldsymbol{a}=\frac{\boldsymbol{u}\times\boldsymbol{v}}{\|\boldsymbol{u}\times\boldsymbol{v}\|}.$$

By linearity, we have

$$h_{\boldsymbol{v}}(h_{\boldsymbol{u}}(\boldsymbol{a}))=\frac{1}{\|\boldsymbol{u}\times\boldsymbol{v}\|}h_{\boldsymbol{v}}(h_{\boldsymbol{u}}(\boldsymbol{u}\times\boldsymbol{v})).$$

Since $\boldsymbol{v}\times\boldsymbol{u}$ is orthogonal to $\boldsymbol{u}$, Lemma 3 implies that

$$h_{\boldsymbol{v}}(h_{\boldsymbol{u}}(\boldsymbol{u}\times\boldsymbol{v}))=h_{\boldsymbol{v}}(\boldsymbol{u}\times\boldsymbol{v}).$$

Likewise, $\boldsymbol{u}\times\boldsymbol{v}$ is orthogonal to $\boldsymbol{v}$, and so

$$h_{\boldsymbol{v}}(\boldsymbol{u}\times\boldsymbol{v})=\boldsymbol{u}\times\boldsymbol{v}$$

by Lemma 3. It thus follows that

$$h_{\boldsymbol{v}}(h_{\boldsymbol{u}}(\boldsymbol{a}))=\boldsymbol{a}.$$

We define $\boldsymbol{a}_1=\boldsymbol{a}$, $\boldsymbol{a}_2=\boldsymbol{u}$, and $\boldsymbol{a}_3=\boldsymbol{a}_1\times\boldsymbol{a}_2$. By construction, $\boldsymbol{a}_1$ is orthogonal to $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$, and $\boldsymbol{a}_2$ orthogonal to $\boldsymbol{a}_3$ and

$$\boldsymbol{a}_1=\frac{\boldsymbol{u}\times\boldsymbol{v}}{\|\boldsymbol{u}\times\boldsymbol{v}\|}.$$

We also have $\|\boldsymbol{a}_1\|=\|\boldsymbol{a}_2\|=1$. To compute the norm of $\|\boldsymbol{a}_3\|$, we recall that

$$\|\boldsymbol{a}_1\times\boldsymbol{a}_2\|=\|\boldsymbol{a}_1\|\|\boldsymbol{a}_2\|\sin\theta,$$

where θ is the angle between $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$. Since $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are orthogonal to each other, we have $\sin\theta = \sin(\pi/2) = 1$.[9] We thus conclude that

$$\|\boldsymbol{a}_3\| = \|\boldsymbol{a}_1 \times \boldsymbol{a}_2\| = \|\boldsymbol{a}_1\|\|\boldsymbol{a}_2\| \sin\theta = 1,$$

whence $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ is an orthonormal system.

We now define $\Theta \in (-\pi, \pi]$ by setting

$$\Theta = \begin{cases} \arccos(\boldsymbol{v} \cdot \boldsymbol{u}) & \text{if } \boldsymbol{v} \cdot \boldsymbol{a}_3 \geq 0; \\ -\arccos(\boldsymbol{v} \cdot \boldsymbol{u}) & \text{if } \boldsymbol{v} \cdot \boldsymbol{a}_3 < 0; \end{cases}$$

and set

$$\boldsymbol{v}(t) = \cos(t\Theta)\boldsymbol{u} + \sin(t\Theta)\boldsymbol{a}_3$$

for each $t \in [0, 1]$. The continuity of $\boldsymbol{v}(t)$ follows immediately from the continuity of $\cos(t\Theta)$ and $\sin(t\Theta)$. Plugging in $t = 0$ and $t = 1$ yields $\boldsymbol{v}(0) = \boldsymbol{u}$ and

$$\boldsymbol{t}(1) = \cos(\Theta)\boldsymbol{u} + \sin(\Theta)\boldsymbol{a}_3.$$

We claim that $\boldsymbol{t}(1) = \boldsymbol{v}$. To see this, we write

$$\boldsymbol{v} = (\boldsymbol{v} \cdot \boldsymbol{a}_1)\boldsymbol{a}_1 + (\boldsymbol{v} \cdot \boldsymbol{a}_2)\boldsymbol{a}_2 + (\boldsymbol{v} \cdot \boldsymbol{a}_3)\boldsymbol{a}_3.$$

Since $\boldsymbol{v}$ is orthogonal to $\boldsymbol{a}_1$, we have

$$\boldsymbol{v} = (\boldsymbol{v} \cdot \boldsymbol{a}_2)\boldsymbol{a}_2 + (\boldsymbol{v} \cdot \boldsymbol{a}_3)\boldsymbol{a}_3.$$

We have

$$\boldsymbol{v} \cdot \boldsymbol{a}_2 = \boldsymbol{v} \cdot \boldsymbol{u} = \cos\Theta$$

by definition, and

$$\begin{aligned}
\boldsymbol{v} \cdot \boldsymbol{a}_3 &= \boldsymbol{v} \cdot (\boldsymbol{a}_1 \times \boldsymbol{u}) \\
&= -\boldsymbol{v} \cdot (\boldsymbol{u} \times \boldsymbol{a}_1) \\
&= -(\boldsymbol{v} \times \boldsymbol{u}) \cdot \boldsymbol{a}_1 \\
&= -(\boldsymbol{v} \times \boldsymbol{u}) \cdot \left(\frac{\boldsymbol{u} \times \boldsymbol{v}}{\|\boldsymbol{u} \times \boldsymbol{v}\|} \right) \\
&= \frac{(\boldsymbol{u} \times \boldsymbol{v}) \cdot (\boldsymbol{u} \times \boldsymbol{v})}{\|\boldsymbol{u} \times \boldsymbol{v}\|} \\
&= \frac{\|\boldsymbol{u} \times \boldsymbol{v}\|^2}{\|\boldsymbol{u} \times \boldsymbol{v}\|} \\
&= \|\boldsymbol{u} \times \boldsymbol{v}\| \\
&= \sin\Theta.
\end{aligned}$$

It thus follows that

$$\begin{aligned}
\boldsymbol{t}(1) &= \cos(\Theta)\boldsymbol{u} + \sin(\Theta)\boldsymbol{a}_3 \\
&= (\boldsymbol{v} \cdot \boldsymbol{a}_2)\boldsymbol{u} + (\boldsymbol{v} \cdot \boldsymbol{a}_3)\boldsymbol{a}_3 \\
&= (\boldsymbol{v} \cdot \boldsymbol{a}_2)\boldsymbol{a}_2 + (\boldsymbol{v} \cdot \boldsymbol{a}_3)\boldsymbol{a}_3 \\
&= \boldsymbol{v},
\end{aligned}$$

as was claimed.

---

[9] Alternatively, we may use the *Binet-Cauchy identity*: $(\boldsymbol{x}_1 \times \boldsymbol{x}_2) \cdot (\boldsymbol{y}_1 \times \boldsymbol{y}_2) = (\boldsymbol{x}_1 \cdot \boldsymbol{x}_2)(\boldsymbol{y}_1 \cdot \boldsymbol{y}_2) - (\boldsymbol{x}_1 \cdot \boldsymbol{y}_2)(\boldsymbol{x}_2 \cdot \boldsymbol{y}_1)$. This identity is not very hard to prove, so you should try to verify it yourself if you're feeling ambitious!

For each $t \in [0,1]$, we define

$$\boldsymbol{f}_t(\boldsymbol{x}) = h_{\boldsymbol{v}(t)}(h_{\boldsymbol{u}}(\boldsymbol{x})),$$

and set

$$\boldsymbol{u}_{j(t)} = \boldsymbol{f}_t(\boldsymbol{u}_j)$$

for $j = 1, 2, 3$. We first note that

$$\boldsymbol{u}_{j(0)} = \boldsymbol{f}_0(\boldsymbol{u}_j) = h_{\boldsymbol{v}(0)}(h_{\boldsymbol{u}}(\boldsymbol{u}_j)) = h_{\boldsymbol{u}}(h_{\boldsymbol{u}}(\boldsymbol{u}_j)) = \boldsymbol{u}_j$$

and that

$$\boldsymbol{u}_{j(1)} = \boldsymbol{f}_1(\boldsymbol{u}_j) = h_{\boldsymbol{v}(1)}(h_{\boldsymbol{u}}(\boldsymbol{u}_j)) = h_{\boldsymbol{v}}(h_{\boldsymbol{u}}(\boldsymbol{u}_j)) = \boldsymbol{v}_j.$$

It then follows from the continuity of $\boldsymbol{v}(t)$ that $\boldsymbol{f}_t$ continuously "interpolates" from $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ to $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ as t goes from 0 to 1.

We claim that $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is a right-handed orthonormal basis. By the assumption, we already know that it is an orthonormal basis. To see that it is a right-handed system, we define the "orientation detection function"

$$d(t) = (\boldsymbol{f}_t(\boldsymbol{u}_1) \times \boldsymbol{f}_t(\boldsymbol{u}_2)) \cdot \boldsymbol{f}_t(\boldsymbol{u}_3),$$

which is a continuous function of t. Since the value of $d(t)$ is either 1 or -1, it follows from the continuity of $d(t)$ that $d(t)$ must be a constant function, and so

$$d(1) = d(0) = 1.$$

Noting that $d(1)$ is the "orientation" of $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$, we conclude that $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is a right-handed orthonormal basis.

4.3. **Problem 3 (3 points + 1 extra-credit point).** By Lemma 3, we have

$$\boldsymbol{f}_t(\boldsymbol{a}_1) = h_{\boldsymbol{v}(t)}(h_{\boldsymbol{u}}(\boldsymbol{a}_1)) = h_{\boldsymbol{v}(t)}(\boldsymbol{a}_1) = \boldsymbol{a}_1,$$

for $\boldsymbol{a}_1$ is orthogonal to both $\boldsymbol{v}(t)$ and $\boldsymbol{u}$.

We claim that

$$\boldsymbol{f}_t(\boldsymbol{a}_2) = \cos(t2\Theta)\boldsymbol{a}_2 + \sin(t2\Theta)\boldsymbol{a}_3.$$

To see this, we first observe that

$$h_{\boldsymbol{u}}(\boldsymbol{a}_2) = h_{\boldsymbol{u}}(\boldsymbol{u}) = \boldsymbol{u} - 2(\boldsymbol{u} \cdot \boldsymbol{u})\boldsymbol{u} = \boldsymbol{u} - 2\boldsymbol{u} = -\boldsymbol{u}.$$

Then,

$$h_{\boldsymbol{v}(t)}(-\boldsymbol{u}) = -(\boldsymbol{u} - 2(\boldsymbol{u} \cdot \boldsymbol{v}(t))\boldsymbol{v}(t)).$$

Since

$$\boldsymbol{u} \cdot \boldsymbol{v}(t) = (1\boldsymbol{u} + 0\boldsymbol{a}_3) \cdot (\cos(t\Theta)\boldsymbol{u} + \sin(t\Theta)\boldsymbol{a}_3) = \cos(t\Theta),$$

we have

$$\begin{aligned} -2(\boldsymbol{u} \cdot \boldsymbol{v}(t))\boldsymbol{v}(t) &= -2\cos(t\Theta)\cos(t\Theta)\boldsymbol{u} - 2\cos(t\Theta)\sin(t\Theta)\boldsymbol{a}_3 \\ &= -2\cos^2(t\Theta)\boldsymbol{u} - \sin(t2\Theta)\boldsymbol{a}_3. \end{aligned}$$

Therefore,

$$\begin{aligned} h_{\boldsymbol{u}}(\boldsymbol{a}_2) &= h_{\boldsymbol{v}(t)}(-\boldsymbol{u}) \\ &= -(\boldsymbol{u} + (-2\cos^2(t\Theta)\boldsymbol{u} - \sin(t2\Theta)\boldsymbol{a}_3)) \\ &= (2\cos^2(t\Theta) - 1)\boldsymbol{u} + \sin(t2\Theta)\boldsymbol{a}_3 \\ &= \cos(t2\Theta)\boldsymbol{u} + \sin(t2\Theta)\boldsymbol{a}_3, \end{aligned}$$

as was claimed.

We also show that

$$\boldsymbol{f}_t(\boldsymbol{a}_3) = -\sin(t2\Theta)\boldsymbol{a}_2 + \cos(t2\Theta)\boldsymbol{a}_3.$$

Since $\boldsymbol{a}_3$ is orthogonal to $\boldsymbol{u}$, we have

$$h_{\boldsymbol{v}(t)}(h_{\boldsymbol{u}}(\boldsymbol{a}_3)) = h_{\boldsymbol{v}(t)}(\boldsymbol{a}_3)$$

by Lemma 3. We observe that

$$\boldsymbol{a}_3 \cdot \boldsymbol{v}(t) = (0\boldsymbol{u} + 1\boldsymbol{a}_3) \cdot (\cos(t\Theta)\boldsymbol{u} + \sin(t\Theta)\boldsymbol{a}_3) = \sin(t\Theta),$$

whence

$$\begin{aligned} h_{\boldsymbol{v}(t)}(\boldsymbol{a}_3) &= \boldsymbol{a}_3 - 2(\boldsymbol{a}_3 \cdot \boldsymbol{v}(t))\boldsymbol{v}(t) \\ &= \boldsymbol{a}_3 - 2\sin(t\Theta)(\cos(t\Theta)\boldsymbol{u} + \sin(t\Theta)\boldsymbol{a}_3) \\ &= \boldsymbol{a}_3 - 2\sin(t\Theta)\cos(t\Theta))\boldsymbol{u} - 2\sin^2(t\Theta))\boldsymbol{a}_3 \\ &= -\sin(t2\Theta)\boldsymbol{u} + \cos(t2\Theta)\boldsymbol{a}_3, \end{aligned}$$

as was to be shown.

We shall now explain why $\boldsymbol{f}_t$ is a rotation by the angle $t2\Theta$ in the plane through $\boldsymbol{0}$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$. For each $\boldsymbol{x} \in \mathbb{R}^3$, we write

$$\boldsymbol{x} = \tilde{x}\boldsymbol{a}_1 + \tilde{y}\boldsymbol{a}_2 + \tilde{z}\boldsymbol{a}_3.$$

By linearity of the Householder reflection, we have

$$\begin{aligned} \boldsymbol{f}_t(\boldsymbol{x}) &= \tilde{x}\boldsymbol{f}_t(\boldsymbol{a}_1) + \tilde{y}\boldsymbol{f}_t(\boldsymbol{a}_2) + \tilde{z}\boldsymbol{f}_t(\boldsymbol{a}_3) \\ &= \tilde{x}\boldsymbol{a}_1 + [\tilde{y}\cos(t2\Theta) - \tilde{z}\sin(t2\Theta)]\boldsymbol{a}_2 + [\tilde{y}\sin(t2\Theta) + \tilde{z}\cos(t2\Theta)]\boldsymbol{a}_3, \end{aligned}$$

Therefore, the "$\boldsymbol{a}_1$ coordinate" is fixed, and the other two coordinates are rotated by the angle $t2\Theta$.[10]

4.4. **Problem 4 (1 extra-credit point).** Suppose that we are considering a rotation from $\boldsymbol{x}$ to $\boldsymbol{y}$. We find a right-handed orthonormal basis $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ so that the axis of rotation is the line through $\boldsymbol{0}$ and $\boldsymbol{a}_1$. Then, by our previous work, we know that the composition $h_{\boldsymbol{v}(t)}(h_{\boldsymbol{u}})$ of two Householder reflections represents a rotation from $\boldsymbol{x}$ to $\boldsymbol{y}$. Since these two vectors were arbitrary, we conclude that any rotation in $\mathbb{R}^3$ can be written as a composition of two Householder reflections.

[10]If you know what a *rotation matrix* is, what we have for the other two coordinates is basically the following:

$$\begin{pmatrix} \cos t2\Theta & -\sin t2\Theta \\ \sin t2\Theta & \cos t2\Theta \end{pmatrix} \begin{pmatrix} \tilde{y} \\ \tilde{z} \end{pmatrix}$$