Chapter 2

analytic geometry

In this chapter I will describe n-dimensional Euclidean space and its essential properties. Much of this is not much removed from the discussion of vectors in calculus III. However, we will state as many things as possible for arbitrarily many finite dimensions. Also, we will make use of matrices and linear algebra where it is helpful. For those of you who have not yet taken linear algebra, I have included a few exercises in the Problem sets to help elucidate matrix concepts. If you do those exercises it should help. If you need more examples just ask.

2.1 Euclidean space and vectors

Rene Descartes put forth the idea of what we now call *Cartesian coordinates* for the plane several hundred years ago. The Euclidean concept of geometry predating Descartes seems abstract in comparison. Try graphing without coordinates. In any event, the definition of Cartesian coordinates and $\mathbb{R}^n$ are intertwined in these notes. If we talk about $\mathbb{R}^n$ then we have a preferred coordinate system because the zero point is at the origin.[1]

Definition 2.1.1.

We define $\mathbb{R}^n = \{ (x_1, x_2, \dots, x_n) \mid x_i \in \mathbb{R} \text{ for each } i = 1, 2, \dots, n\}$. If $P = (a_1, a_2, \dots, a_n)$ is a **point** in $\mathbb{R}^n$ then the j-th Cartesian coordinate of the point P is a_j.

Notice that[2] in terms of sets we can write $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ and $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. Since points in $\mathbb{R}^n$ are in 1-1 correspondance with vectors in $\mathbb{R}^n$ we can add vectors and rescale them by scalar multiplication. If I wish to emphasize that we are working with vectors I may use the notation $< a, b, c > \in V^3$ etc... However, we will think of $\mathbb{R}^n$ as both a set of points and a set of vectors, which takes precendence depends on the context.

[1]some other authors might use $\mathbb{R}^n$ is refer to abstract Euclidean space where no origin is given apriori by the mathematics. Given Euclidean space $\mathcal{E}$ and a choice of an origin $\mathcal{O}$, one can always set-up a 1-1 correspondance with $\mathbb{R}^n$ by mapping the origin to zero in $\mathbb{R}^n$.

[2]Technically these are ambiguous since the Cartesian product of sets is nonassociative but in these notes we identify $\mathbb{R} \times (\mathbb{R} \times \mathbb{R})$ and $(\mathbb{R} \times \mathbb{R}) \times \mathbb{R}$ as the same object. Btw, my Math 200 notes have more on basics of Cartesian products.

Definition 2.1.2.

We define $V^n = \{< v_1, v_2, \ldots, v_n > \mid v_i \in \mathbb{R} \text{ for each } i = 1, 2, \ldots, n\}$.
If $v =< v_1, v_2, \ldots, v_n >$ is a **vector** in $\mathbb{R}^n$ then the j-th component of the vector v is v_j.
Let $v, w \in V^n$ with $v =< v_i >, w =< w_i >$ and $c \in \mathbb{R}$ then we define:

$$v + w =< v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n > \qquad cv =< cv_1, cv_2, \ldots, cv_n > .$$

I will refer to V^n as the set of n-dimensional real vectors. The dot-product is used to define angles and lengths of vectors in V^n.

Definition 2.1.3.

If $v =< v_1, v_2, \ldots, v_n >$ and $w =< w_1, w_2, \ldots, w_n >$ are vectors in V^n then the **dot-product** of v and w is a **real number** defined by:

$$v \cdot w = v_1 w_1 + v_1 w_1 + \cdots + v_n w_n .$$

The **length (or norm) of a vector** $v =< v_1, v_2, \ldots, v_n >$ is denoted $||v||$ and is the real number defined by:

$$||v|| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_1^2 + \cdots + v_n^2}.$$

If $v =< v_1, v_2, \ldots, v_n > \neq 0$ and $w =< w_1, w_2, \ldots, w_n > \neq 0$ are vectors in V^n then the **angle θ between v and w** is defined by:

$$\theta = \cos^{-1}\left(\frac{v \cdot w}{||v|| \, ||w||} \right)$$

The vectors v, w are said to be **orthogonal** iff $v \cdot w = 0$.

Example 2.1.4. . .

① $V \cdot 0 = V_1(0) + V_2(0) + \cdots + V_n(0) = 0$
$\therefore$ the zero vector is orthogonal to all vectors.

② $\langle a, b, c\rangle \cdot \hat{i} = \langle a, b, c\rangle \cdot \langle 1, 0, 0\rangle = a + 0 + 0 = a$
we can select the x-component by taking dot-product with $\hat{i}$.

③ if you need to see more

The dot-product has many well-known properties:

Proposition 2.1.5.

Suppose $x, y, z \in \mathbb{R}^{n\times 1}$ and $c \in \mathbb{R}$ then

1. $x \cdot y = y \cdot x$
2. $x \cdot (y + z) = x \cdot y + x \cdot z$
3. $c(x \cdot y) = (cx) \cdot y = x \cdot (cy)$
4. $x \cdot x \geq 0$ and $x \cdot x = 0$ iff $x = 0$

Notice that the formula $\cos^{-1}\left[\frac{x\cdot y}{||x||\,||y||}\right]$ needs to be justified since the domain of inverse cosine does not contain all real numbers. The inequality that we need for it to be reasonable is $\left|\frac{x\cdot y}{||x||\,||y||}\right| \leq 1$, otherwise we would not have a number in the $dom(cos^{-1}) = range(cos) = [-1, 1]$. An equivalent inequality is $|x \cdot y| \leq ||x||\,||y||$ which is known as the **Cauchy-Schwarz** inequality.

Proposition 2.1.6.

If $x, y \in \mathbb{R}^{n\times 1}$ then $|x \cdot y| \leq ||x||\,||y||$

These properties are easy to justify for the norm we defined in this section.

Proposition 2.1.7.

Let $x, y \in \mathbb{R}^{n\times 1}$ and suppose $c \in \mathbb{R}$ then

1. $||cx|| = |c|\,||x||$
2. $||x + y|| \leq ||x|| + ||y||$

Every nonzero vector can be written as a unit vector scalar multiplied by its magnitude.

$$\boxed{v \in V^n \text{ such that } v \neq 0 \;\Rightarrow\; v = ||v||\hat{v} \text{ where } \hat{v} = \frac{1}{||v||}v.}$$

You should recall that we can write any vector in V^3 as

$$v =< a, b, c >= a < 1, 0, 0 > +b < 0, 1, 0 > +c < 0, 0, 1 >= a\hat{i} + b\hat{j} + c\hat{k}$$

where we defined the $\hat{i} =< 1, 0, 0 >$, $\hat{j} =< 0, 1, 0 >$, $\hat{k} =< 0, 0, 1 >$. You can easily verify that distinct Cartesian unit-vectors are orthogonal. Sometimes we need to produce a vector which is orthogonal to a given pair of vectors, it turns out the cross-product is one of two ways to do that in V^3. We will see much later that this is special to three dimensions.

Definition 2.1.8.

If $A =< A_1, A_2, A_3 >$ and $B =< B_1, B_2, B_3 >$ are vectors in V^3 then the **cross-product** of A and B is a **vector** $A \times B$ which is defined by:

$$\vec{A} \times \vec{B} =< A_2B_3 - A_3B_2, A_3B_1 - A_1B_3, A_1B_2 - A_2B_1 > .$$

The magnitude of $\vec{A} \times \vec{B}$ can be shown to satisfy $||\vec{A} \times \vec{B}|| = ||\vec{A}||\,||\vec{B}||\sin(\theta)$ and the direction can be deduced by **right-hand-rule**. The right hand rule for the unit vectors yields:

$$\hat{i} \times \hat{j} = \hat{k}, \quad \hat{k} \times \hat{i} = \hat{j}, \quad \hat{j} \times \hat{k} = \hat{i}$$

If I wish to discuss both the point and the vector to which it corresponds we may use the notation

$$\boxed{P = (a_1, a_2, \dots, a_n) \longleftrightarrow \vec{P} =< a_1, a_2, \dots, a_n >}$$

With this notation we can easily define directed line-segments as the vector which points from one point to another, also the distance bewtween points is simply the length of the vector which points from one point to the other:

Definition 2.1.9.

Let $P, Q \in \mathbb{R}^n$. The directed line segment from P to Q is $\overrightarrow{PQ} = \vec{Q} - \vec{P}$. This vector is drawn from tail Q to the tip P where we denote the direction by drawing an arrowhead. The **distance between** P **and** Q is $d(P, Q) = ||\,\overrightarrow{PQ}\,||$.

2.1.1 compact notations for vector arithmetic

I prefer the following notations over the hat-notation of the preceding section because this notation generalizes nicely to n-dimensions.

$$\boxed{e_1 =< 1, 0, 0 > \qquad e_2 =< 0, 1, 0 > \qquad e_3 =< 0, 0, 1 > .}$$

Likewise the Kronecker delta and the Levi-Civita symbol are at times very convenient for abstract calculation:

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \qquad \epsilon_{ijk} = \begin{cases} 1 & (i, j, k) \in \{(1, 2, 3), (3, 1, 2), (2, 3, 1)\} \\ -1 & (i, j, k) \in \{(3, 2, 1), (2, 1, 3), (1, 3, 2)\} \\ 0 & \text{if any index repeats} \end{cases}$$

An equivalent definition for the Levi-civita symbol is simply that $\epsilon_{123} = 1$ and it is antisymmetric with respect to the interchange of any pair of indices;

$$\epsilon_{ijk} = \epsilon_{jki} = \epsilon_{kij} = -\epsilon_{kji} = -\epsilon_{jik} = -\epsilon_{ikj}.$$

Now let us restate some earlier results in terms of the Einstein repeated index conventions[3], let $\vec{A}, \vec{B} \in V^n$ and $c \in \mathbb{R}$ then

$\vec{A} = A_k e_k$	standard basis expansion
$e_i \cdot e_j = \delta_{ij}$	orthonormal basis
$(\vec{A}+\vec{B})_i = \vec{A}_i + \vec{B}_i$	vector addition
$(\vec{A}-\vec{B})_i = \vec{A}_i - \vec{B}_i$	vector subtraction
$(c\vec{A})_i = c\vec{A}_i$	scalar multiplication
$\vec{A} \cdot \vec{B} = A_k B_k$	dot product
$(\vec{A} \times \vec{B})_k = \epsilon_{ijk} A_i B_j$	cross product.

All but the last of the above are readily generalized to dimensions other than three by simply increasing the number of components. However, the cross product is special to three dimensions. I can't emphasize enough that the formulas given above for the dot and cross products can be utilized to yield great efficiency in abstract calculations.

Example 2.1.10. . .

Prove $\vec{A} \cdot (\vec{B} \times \vec{C}) = \vec{C} \cdot (\vec{A} \times \vec{B})$.

$$\begin{aligned}
\vec{A} \cdot (\vec{B} \times \vec{C}) &= A_k \, (\vec{B} \times \vec{C})_k \\
&= A_k \, \epsilon_{ijk} B_i C_j \\
&= \epsilon_{ijk} A_k B_i C_j \\
&= \epsilon_{ijk} C_j A_k B_i \\
&= \epsilon_{kij} C_j A_k B_i \\
&= C_j \, \epsilon_{kij} A_k B_i \\
&= C_j \, (\vec{A} \times \vec{B})_j \\
&= \vec{C} \cdot (\vec{A} \times \vec{B}).
\end{aligned}$$

notice $\epsilon_{ijk} = -\epsilon_{ikj} = -(-\epsilon_{kij})$
$\therefore \epsilon_{ijk} = \epsilon_{kij}$

Another example, you can prove $\epsilon_{ikj}\epsilon_{mpj} = \delta_{im}\delta_{kp} - \delta_{ip}\delta_{km}$ then this identity can be used to evaluate triple cross products,

$$\begin{aligned}
[\vec{A} \times (\vec{B} \times \vec{C})]_k &= \epsilon_{ijk} A_i \, (\vec{B} \times \vec{C})_j \\
&= \epsilon_{ijk} \epsilon_{mpj} A_i B_m C_p \\
&= -\epsilon_{ikj} \epsilon_{mpj} A_i B_m C_p \\
&= -\delta_{im}\delta_{kp} A_i B_m C_p + \delta_{ip}\delta_{km} A_i B_m C_p \\
&= -A_i B_i C_k + A_i B_k C_i \\
&= -(\vec{A} \cdot \vec{B}) C_k + (\vec{A} \cdot \vec{C}) B_k \\
&= \left[(\vec{A} \cdot \vec{C}) \vec{B} - (\vec{A} \cdot \vec{B}) \vec{C} \right]_k
\end{aligned}$$

$\therefore \vec{A} \times (\vec{B} \times \vec{C}) = (\vec{A} \cdot \vec{C})\vec{B} - (\vec{A} \cdot \vec{B})\vec{C}$.

[3] there are more details to be seen in the Appendix if you're curious

2.2 matrices

An $m \times n$ matrix is an array of numbers with m-rows and n-columns. We define $\mathbb{R}^{m\times n}$ to be the set of all $m \times n$ matrices. The set of all n-dimensional column vectors is $\mathbb{R}^{n\times 1}$. The set of all n-dimensional row vectors is $\mathbb{R}^{1\times n}$. A given matrix $A \in \mathbb{R}^{m\times n}$ has mn-components A_{ij}. Notice that the components are numbers; $A_{ij} \in \mathbb{R}$ for all i, j such that $1 \leq i \leq m$ and $1 \leq j \leq n$. We should not write $A = A_{ij}$ because it is nonesense, however $A = [A_{ij}]$ is quite fine.

Suppose $A \in \mathbb{R}^{m\times n}$, note for $1 \leq j \leq n$ we have $col_j(A) \in \mathbb{R}^{m\times 1}$ whereas for $1 \leq i \leq m$ we find $row_i(A) \in \mathbb{R}^{1\times n}$. In other words, an $m \times n$ matrix has n columns of length m and n rows of length m.

Definition 2.2.1.

Two matrices A and B are equal iff $A_{ij} = B_{ij}$ for all i, j. Given matrices A, B with components A_{ij}, B_{ij} and constant $c \in \mathbb{R}$ we define

$$(A + B)_{ij} = A_{ij} + B_{ij} \qquad (cA)_{ij} = cA_{ij} \qquad , \text{ for all } i, j.$$

The **zero matrix** in $\mathbb{R}^{m\times n}$ is denoted 0 and defined by $0_{ij} = 0$ for all i, j. The additive inverse of $A \in \mathbb{R}^{m\times n}$ is the matrix $-A$ such that $A + (-A) = 0$. The components of the additive inverse matrix are given by $(-A)_{ij} = -A_{ij}$ for all i, j. Likewise, if $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$ then the product $AB \in \mathbb{R}^{m\times p}$ is defined by:

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}$$

for each $1 \leq i \leq m$ and $1 \leq j \leq p$. In the case $m = p = 1$ the indices i, j are omitted in the equation since the matrix product is simply a number which needs no index. The identity matrix in $\mathbb{R}^{n\times n}$ is the $n \times n$ square matrix I whose components are the Kronecker delta; $I_{ij} = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$. The notation I_n is sometimes used if the size of the identity matrix needs emphasis, otherwise the size of the matrix I is to be understood from the context.

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let $A \in \mathbb{R}^{n\times n}$. If there exists $B \in \mathbb{R}^{n\times n}$ such that $AB = I$ and $BA = I$ then we say that A is **invertible** and $A^{-1} = B$. Invertible matrices are also called **nonsingular**. If a matrix has no inverse then it is called a **noninvertible** or **singular** matrix. Let $A \in \mathbb{R}^{m\times n}$ then $A^T \in \mathbb{R}^{n\times m}$ is called the **transpose** of A and is defined by $(A^T)_{ji} = A_{ij}$ for all $1 \leq i \leq m$ and $1 \leq j \leq n$. Note **dot-product** of $v, w \in V^n$ is given by $v \cdot w = v^T w$.

Remark 2.2.2.

We will use the convention that points in $\mathbb{R}^n$ are column vectors. However, we will use the somewhat subtle notation $(x_1, x_2, \dots x_n) = [x_1, x_2, \dots x_n]^T$. This helps me write $\mathbb{R}^n$ rather than $\mathbb{R}^{\,n\times 1}$ and I don't have to pepper transposes all over the place. If you've read my linear algebra notes you'll appreciate the wisdom of our convention. Likewise, for the sake of matrix multiplication, we adopt the subtle convention $< x_1, x_2, \dots x_n >= [x_1, x_2, \dots x_n]^T$ for vectors in V^n. Worse yet I will later in the course fail to distinguish between V^n and $\mathbb{R}^n$. Most texts adopt the view that points and vectors can be identified so there is no distinction made between these sets. We also follow that view, however I reserve the right to use V^n if I wish to emphasize that I am using vectors.

Definition 2.2.3.

Let $e_i \in \mathbb{R}^n$ be defined by $(e_i)_j = \delta_{ij}$. The size of the vector e_i is determined by context. We call e_i the i-th standard basis vector.

Example 2.2.4. . .

$$e_1 \in \mathbb{R}^2 \Rightarrow e_1 = (1, 0)$$
$$e_1 \in \mathbb{R}^3 \Rightarrow e_1 = (1, 0, 0)$$

Also, note $\vec{A} = \langle A_1, A_2, \dots, A_n \rangle = \sum_{i=1}^{n} A_i e_i$

and $\vec{A} \cdot e_j = \sum_{i=1}^{n} A_i e_i \cdot e_j = \sum_{i=1}^{n} A_i \delta_{ij} = A_j$.

Definition 2.2.5.

The ij-th **standard basis matrix** for $\mathbb{R}^{\,m\times n}$ is denoted E_{ij} for $1 \le i \le m$ and $1 \le j \le n$. The matrix E_{ij} is zero in all entries except for the (i, j)-th slot where it has a 1. In other words, we define $(E_{ij})_{kl} = \delta_{ik}\delta_{jl}$.

Theorem 2.2.6.

Assume $A \in \mathbb{R}^{\,m\times n}$ and $v \in \mathbb{R}^{\,n\times 1}$ and define $(E_{ij})_{kl} = \delta_{ik}\delta_{jl}$ and $(e_i)_j = \delta_{ij}$ as before then,

$$v = \sum_{i=1}^{n} v_n e_n \qquad A = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}E_{ij}.$$

$$[e_i{}^T A] = row_i(A) \qquad [Ae_i] = col_i(A) \qquad A_{ij} = (e_i)^T A e_j$$

$$E_{ij}E_{kl} = \delta_{jk}E_{il} \qquad E_{ij} = e_i e_j{}^T \qquad e_i{}^T e_j = e_i \cdot e_j = \delta_{ij}$$

You can look in my linear algebra notes for the details of the theorem. I'll just expand one point here: Let $A \in \mathbb{R}^{m\times n}$ then

$$\begin{aligned}
A &= \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} \\
&= A_{11}\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix} + A_{12}\begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix} + \cdots + A_{mn}\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & 0 \\ 0 & 0 & \cdots & 1 \end{bmatrix} \\
&= A_{11}E_{11} + A_{12}E_{12} + \cdots + A_{mn}E_{mn}.
\end{aligned}$$

The calculation above follows from repeated mn-applications of the definition of matrix addition and another mn-applications of the definition of scalar multiplication of a matrix.

Example 2.2.7. . . .

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ if I want to select the (ij)-component then I can multiply by e_i^T on the left & e_j on right ; $A_{ij} = e_i^T A e_j$

For example,

$$\begin{aligned}
e_1^T A e_2 &= [1, 0]\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= [1, 0]\begin{bmatrix} b \\ d \end{bmatrix} \\
&= \underline{b = A_{12}}.
\end{aligned}$$

this idea is useful when we study quadratic forms $Q(x) = x^T A x$.

2.3 linear transformations

We should recall the precise definition of a linear combination: A **linear combination** of objects $A_1, A_2, \dots, A_k$ is a sum $c_1A_1 + c_2A_2 + \cdots + c_kA_k = \sum_{i=1}^{k} c_iA_i$ where $c_i \in \mathbb{R}$ for each i. Essentially, a vector space is simply a set of objects called "vectors" for which any linear combination of the vectors is again in the set. In other words, vectors in a vector space can be added by "vector addition" or rescaled by a so-called "scalar multiplication". A linear transformation is a mapping from one vector space to another which preserves linear combinations.

Definition 2.3.1.

Let V, W be vector spaces. If a mapping $L : V \to W$ satisfies

1. $L(x + y) = L(x) + L(y)$ for all $x, y \in V$,
2. $L(cx) = cL(x)$ for all $x \in V$ and $c \in \mathbb{R}$

then we say L is a linear transformation.

Example 2.3.2. . .

$L(x) = x \cdot x$ for $x \in \mathbb{R}^n$, this means that $L : \mathbb{R}^n \to \mathbb{R}$. Notice that

$$\begin{aligned} L(x+y) &= (x+y)\cdot(x+y) \\ &= x\cdot x + x\cdot y + y\cdot x + y\cdot y \\ &= L(x) + 2x\cdot y + L(y) \end{aligned}$$

$\Rightarrow L$ not linear since we can find x, y with $x \cdot y \neq 0$.

Example 2.3.3. . .

$L(x) = mx + b$ for all $x \in \mathbb{R}$ and m, b are fixed constants.

Note $L(0+0) = b$

but $L(0) + L(0) = b + b = 2b$

The function $L : \mathbb{R} \to \mathbb{R}$ whose graph is a line need not be a linear transformati

Example 2.3.4. . .

$L(x_1, x_2, x_3) = x_1$ is a mapping from $\mathbb{R}^3 \to \mathbb{R}$. We can show this is linear

$$L(x+y) = (x+y)_1 = x_1 + y_1 = L(x) + L(y)$$

$$L(cx) = (cx)_1 = cx_1 = cL(x)$$

for all $x, y \in \mathbb{R}^3$.

Definition 2.3.5.

Let $L : \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$ be a linear transformation, the matrix $A \in \mathbb{R}^{m\times n}$ such that $L(x) = Ax$ for all $x \in \mathbb{R}^{n\times 1}$ is called the **standard matrix** of L. We denote this by $[L] = A$ or more compactly, $[L_A] = A$, we say that L_A is the linear transformation induced by A.

Example 2.3.6. . .

$$L(x, y) = (x+y,\ x-y,\ y) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

note $L : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ need a 3×2 matrix.

$$\therefore\quad [L] = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}$$

Example 2.3.7. . .

$$L(x_1, x_2, x_3, x_4) = (x_1 + x_4,\ x_2 - 3x_3,\ x_4 + 3) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \end{bmatrix}$$

This mapping $L : \mathbb{R}^4 \longrightarrow \mathbb{R}^3$ is <u>not</u> linear due to the $(0,0,0,3)$ vector.

$$L(0+0) = L(0) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \end{bmatrix}$$

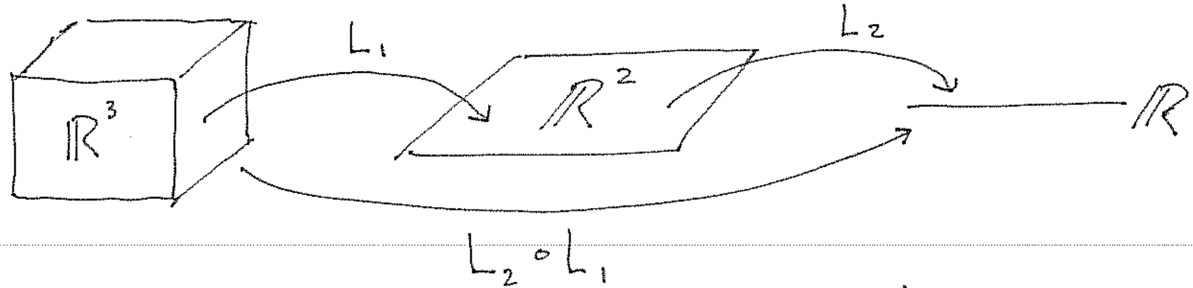
$$L(0) + L(0) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 6 \end{bmatrix} \neq L(0+0)$$

<u>Remark</u>: if $L(0) \neq 0$ then L is <u>not</u> linear transformation. This is a useful criteria since 0 is easy to check.

Proposition 2.3.8.

Let V_1, V_2, V_3 be vector spaces and suppose $L_1 : V_1 \to V_2$ and $L_2 : V_2 \to V_3$ are linear transformations then $L_2 \circ L_1 : V_1 \to V_3$ is a linear transformation and if V_1, V_2 are column spaces then $[L_2 \circ L_1] = [L_2][L_1]$.

Example 2.3.9. . .



If $L_1(x) = (x_1, x_2)$ and $L_2(x_1, x_2) = x_1 + x_2$ then $L_1(x_1, x_2, x_3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ & $L_2(x_1, x_2) = [1, 1]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Hence, $[L_1] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ and $[L_2] = [1, 1]$.

$$\Rightarrow \underline{[L_2][L_1] = [1, 1]\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = [1, 1, 0]}. \quad \text{I}$$

In contrast we can calculate the composite directly,

$$\begin{aligned} (L_2 \circ L_1)(x_1, x_2, x_3) &= L_2(L_1(x_1, x_2, x_3)) \\ &= L_2(x_1, x_2) \\ &= x_1 + x_2 \\ &= [1 \quad 1 \quad 0]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{aligned}$$

$$\therefore \underline{[L_2 \circ L_1] = [1, 1, 0]}. \quad \text{II}$$

Comparing I & II you see $[L_2 \circ L_1] = [L_2][L_1]$.

This algebra $\Rightarrow$ chain rule of multivariate calculus, we'll see it.

2.4 orthogonal transformations

Orthogonal transformations play a central role in the study of geometry.

Definition 2.4.1.

If $T : \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ is a linear transformation such that $T(x)\cdot T(y) = x\cdot y$ for all $x, y \in \mathbb{R}^{n\times 1}$ then we say that T is an **orthogonal transformation**. The matrix R of an orthogonal transformation is called an **orthogonal matrix** and it satisfies $R^TR = I$. The set of orthogonal matrices is $O(n)$ and the subset of rotation matrices is denoted $SO(n) = \{R \in O(n) | det(R) = 1\}$.

The definition above is made so that an orthogonal transformation preserves the lengths of vectors and the angle between pairs of vectors. Since both of those quantities are defined in terms of the dot-product it follows that lengths and angles are invariant under a linear transformation since the dot-product is unchanged. In particular,

$$||T(x)||^2 = T(x)\cdot T(x) = x\cdot x = ||x||^2 \quad \Rightarrow \quad \boxed{||T(x)|| = ||x||}$$

Likewise, defining θ to be the angle between x, y and θ_T the angle between $T(x), T(y)$:

$$T(x)\cdot T(y) = x\cdot y \;\Rightarrow\; ||T(x)||||T(y)||\cos\theta_T = ||x||||y||\cos\theta \;\Rightarrow\; \cos\theta_T = \cos\theta \;\Rightarrow\; \boxed{\theta_T = \theta}$$

2.5 orthogonal bases

Definition 2.5.1.

A set S of vectors in $\mathbb{R}^{n\times 1}$ is **orthogonal** iff every pair of vectors in the set is orthogonal. If S is orthogonal and all vectors in S have length one then we say S is **orthonormal**.

It is easy to see that an orthogonal transformation maps an orthonormal set to another orthonormal set. Observe that the standard basis $\{e_1, e_2, \dots, e_n\}$ is an orthonormal set of vectors since $e_i \cdot e_j = \delta_{ij}$. When I say the set is a **basis** for $\mathbb{R}^n$ this simply means that it is a set of vectors which **spans** $\mathbb{R}^n$ by finite linear combinations and is also **linearly independent**. In case you haven't had linear,

Definition 2.5.2.

1. $S = \{v_1, v_2 \dots, v_k\}$ is **linearly independent** iff $\sum_{i=1}^k c_iv_i = 0$ implies $c_i = 0$ for $i = 1, 2, \dots k$.

2. $S = \{v_1, v_2 \dots, v_k\}$ is **spans** W iff for each $w \in W$ there exist constants $w_1, w_2, \dots, w_k$ such that $w = \sum_{i=1}^k w_iv_i$.

3. β is a **basis** for a vector space V iff it is a linearly independent set which spans V. Moreover, if there are n vectors in β then we say $dim(V) = n$.

In fact, since the dimension of $\mathbb{R}^n$ is known to be n either spanning or linear independence of a set of n vectors is a sufficient condition to insure a given set of vectors is a basis for $\mathbb{R}^n$. In any event, we can prove that an orthonormal set of vectors is linearly independent. So, to summarize, if we have a linear transformation T we can construct a new orthonormal basis from the standard basis:

$$T(\{e_1, \dots, e_n\}) = \{T(e_1), \dots T(e_n)\}$$

Example 2.5.3. *In calculus III you hopefully observed (perhaps not in this langauge, but the patterns were there just waiting to be noticed):*

1. *a line through the origin is spanned by its direction vector.*
2. *a plane through the origin is spanned by any two non-paralell vectors that lie in that plane.*
3. *three dimensional space is spanned by three non-coplanar vectors. For example, $\hat{i}, \hat{j}, \hat{k}$ span $\mathbb{R}^3$.*

2.6 coordinate systems

Definition 2.6.1.

A **coordinate system** of $\mathbb{R}^n$ is a set of n functions $\bar{x}_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, 2, \dots, n$ such that we can invert the equations

$$\bar{x}_i = \bar{x}_i(x_1, x_2, \dots, x_n) \qquad \text{to obtain} \qquad x_i = x_i(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$$

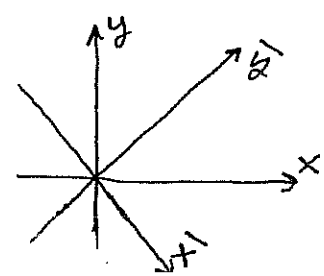
on most of $\mathbb{R}^n$ In other words, we can group the functions into a coordinate map $\Phi = \bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$ and $\bar{x}$ is a 1-1 correspondance on most of $\mathbb{R}^n$. We call $\bar{x}_j$ the **j-th coordinate** of the $\bar{x}$ coordinate system. For a particular coordinate system we also define the **j-th coordinate axis** to be the set of points such that all the other coordinates are zero. If the coordinate axis is a line for each coordinate then the coordinate system is said to be **rectilinear**. If the coordinate axis is not a line for all the coordinates then the coordinate system is said to be **curvelinear**. If the coordinate axes have orthgonal unit vectors then the coordinate system is said to be an **orthogonal coordinate system.** Likewise, if the coordinate curves of a curvelinear coordinate system have orthogonal tangent vectors at all points then the curvelinear coordinate system is said to give an **orthogonal coordinate system.**

Example 2.6.2. .

Let $\bar{x} = x - y$ and $\bar{y} = x + y$.
Notice that $\bar{x} + \bar{y} = 2x$ whereas $\bar{x} - \bar{y} = -2y$ hence we have inverse relations : $x = \frac{1}{2}(\bar{x} + \bar{y})$ & $y = \frac{1}{2}(-\bar{x} + \bar{y})$.
The coordinate axes are found by setting $\bar{x} = 0$ or $\bar{y} = 0$,

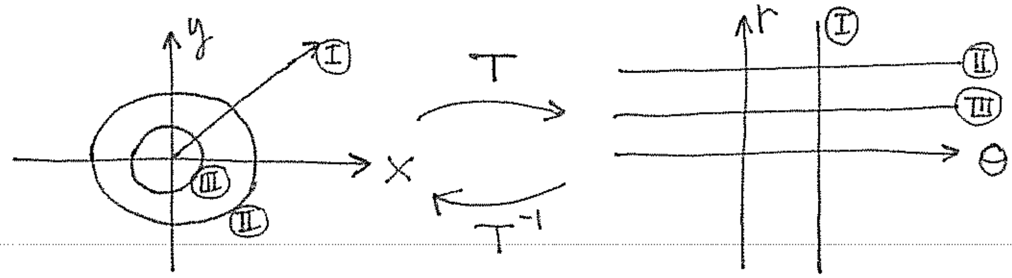
$\bar{x}$ axis has $\bar{y} = 0$ $\to$ $x = (1/2)\bar{x}$, $y = (-1/2)\bar{x}$ $\to$ $\underline{y = -x}$ eq^n of $\bar{x}$ axis

$\bar{y}$ axis has $\bar{x} = 0$ $\to$ $x = \frac{1}{2}\bar{y}$, $y = \frac{1}{2}\bar{y}$ $\to$ $\underline{y = x}$ eq^n of $\bar{y}$ axis



The case of Cartesian coordinates has $\Phi = Id$. Conceptually we think of the codomain as a different space than the domain in general. For example, in the case of polar coordinates on the plane we have a mapping $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ where a circle in the domain becomes a line in the range. The line in $r\theta$ space is a representation of the circle in the view of polar coordinates. Students often confuse themselves by implicitly insisting that the domain and range of the coordinate map are the same copy of $\mathbb{R}^n$ but this is the wrong concept. Let me illustrate with a few mapping pictures:

Example 2.6.3. .



$$T(x,y) = (\theta, r) = \left(\tan^{-1}(y/x),\ \sqrt{x^2+y^2}\right)$$

$$T^{-1}(\theta, r) = (r\cos\theta, r\sin\theta)$$

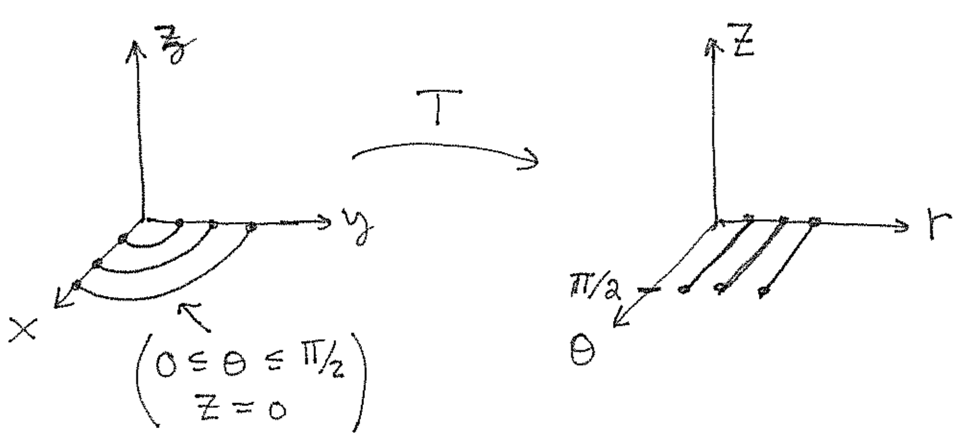
$$T : \mathbb{R}^2_{xy} \longrightarrow \mathbb{R}^2_{\theta r}, \quad \text{polar coordinates are curvelinear.}$$

Generally I admit that I'm being a bit vague here because the common useage of the term coordinate system is a bit vague. Later I'll define a patched manifold and that structure will give a refinement of the coordinate concept which is unambiguous. That said, common coordinate systems such as polar, spherical coordinates fail to give coordinates for manifolds unless we add restrictions on the domain of the coordinate which are not typically imposed in applications. Let me give a few coordinate systems commonly used in applications so we can constrast those against the coordinate systems given from orthonormal bases of $\mathbb{R}^n$.

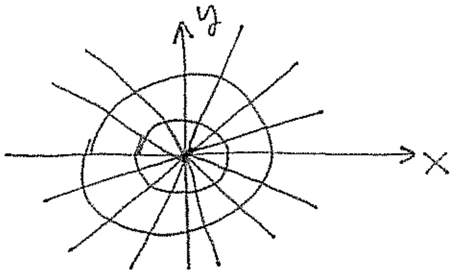
Example 2.6.4. . Cylindrical Coordinates

$$T : \mathbb{R}^3_{xyz} \longrightarrow \mathbb{R}^3_{\theta r z}$$

$$T(x,y,z) = (\theta, r, z) \quad \text{where} \quad \underbrace{\theta = \tan^{-1}(y/x)}_{\text{(Sometimes)}}, \quad r = \sqrt{x^2+y^2}$$

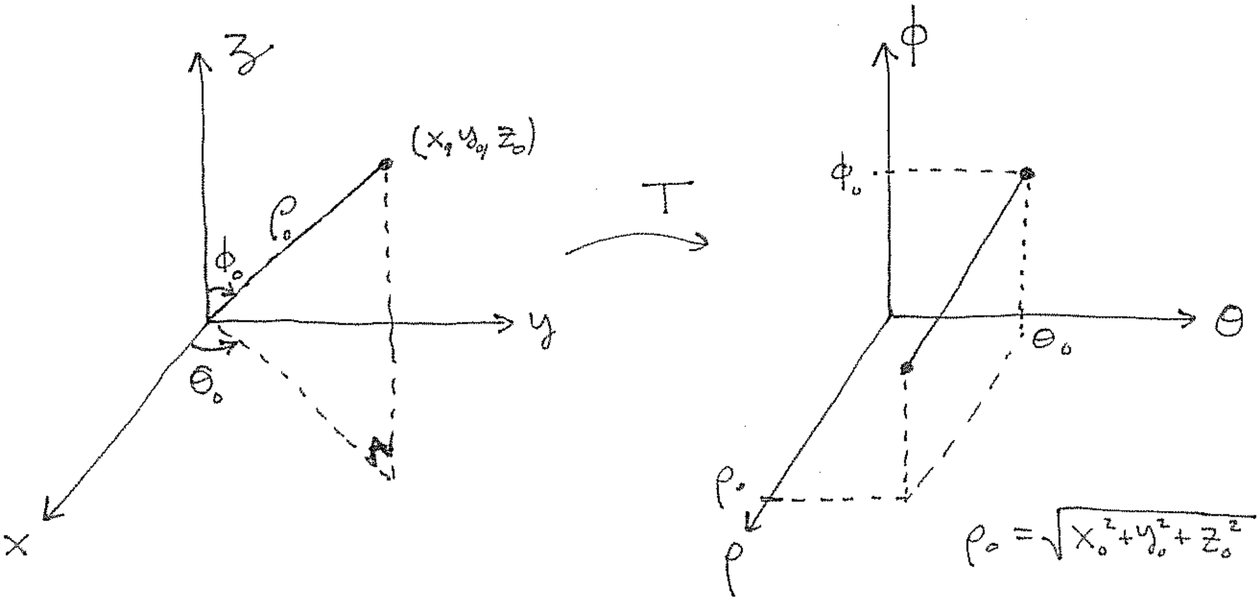


Example 2.6.5. *Consider* $\mathbb{R}^2$ *with the usual* x, y *coordinates,* **polar coordinates** r, θ *are given by the* **polar radius** $r = \sqrt{x^2+y^2}$ *and* **polar angle** $\theta = \tan^{-1}(y/x)$. *These are inverted to give* $x = r\cos(\theta)$ *and* $y = r\sin(\theta)$. *Notice that* θ *is not well defined along* $x = 0$ *if we take the given formula as the definition. Even so the angle at the origin is not well-defined no matter how you massage the equations. Polar coordinates are curvelinear coordinates, setting* $\theta = 0$ *yields a ray along the postive x-axis whereas setting* $r = 0$ *just yields the origin.*



note: in calculus II we allow $r<0$ for convenience of polar graphing.

Example 2.6.6. *Consider* $\mathbb{R}^3$ *with the usual* x, y, z *coordinates,* **spherical coordinates** ρ, θ, ϕ *are given by* **spherical radius** $\rho = \sqrt{x^2+y^2+z^2}$, **polar angle** $\theta = \tan^{-1}(y/x)$ *and* **azimuthial angle** $\phi = \cos^{-1}(z/\sqrt{x^2+y^2+z^2})$. *These are inverted to give* $x = \rho\cos(\theta)\sin(\phi)$ *and* $y = \rho\sin(\theta)\sin(\phi)$ *and* $z = \rho\cos(\phi)$. *Even so the angles can't be well-defined everywhere. The function of inverse tangent can never return a polar angle in quadrants* **II** *or* **III** *because range*$(\tan^{-1}) = (-\pi/2, \pi/2)$. *In order to find angles in the quadrants with* $x < 0$ *we have to adjust the equations by hand as we are taught in trigonmetry. Spherical coordinates are also curvelinear, there is no coordinate axis for the spherical radius and the angles have rays rather than lines for their coordinate axes.*



quick notation

$$T : \underbrace{\mathbb{R}^3_{xyz}}_{\text{tell us use xyz variables for domain}} \longrightarrow \underbrace{\mathbb{R}^3_{\rho\theta\phi}}_{\text{tells us use } \rho\theta\phi \text{ for variables in range}}$$

Example 2.6.7. *Consider $\mathbb{R}^n$ with the usual Cartesian coordinates $x = (x_1, x_2, \dots, x_n)$. If $p \in \mathbb{R}^n$ then we can write*

$$p = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n = [e_1|e_2|\cdots|e_n][x_1, x_2, \dots, x_n]^T$$

Let T be an orthogonal transformation and define a rotated basis f_i by $[f_1|\cdots|f_n] = [e_1|\cdots|e_n]R = R$ where $R \in SO(n)$. Since $R^TR = I$ it follows that $R^{-1} = R^T$ and so $[e_1|\cdots|e_n] = [f_1|\cdots|f_n]R^T$. Note that $p = [f_1|\cdots|f_n]R^Tp$. However, the y-coordinates will satisfy $p = [f_1|\cdots|f_n]y$ where $y = [y_1, y_2, \dots, y_n]^T$. We deduce,

$$y = R^T x.$$

We find that if we set up a rotated coordinate system where the new basis is formed by rotating the standard basis by R then the new coordinates relate to the old coordinates by the inverse rotation $R^T = R^{-1}$.

Let me break down the example in the $n = 2$ case.

Example 2.6.8. *Let $\{e_1, e_2\}$ be the standard basis for $\mathbb{R}^2$. In invite the reader to check that $R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \in SO(2)$. If our calculation is correct in the previous example the new coordinate axes should be obtained from the standard basis by the inverse transformation.*

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos\theta + y\sin\theta \\ -x\sin\theta + y\cos\theta \end{bmatrix}$$

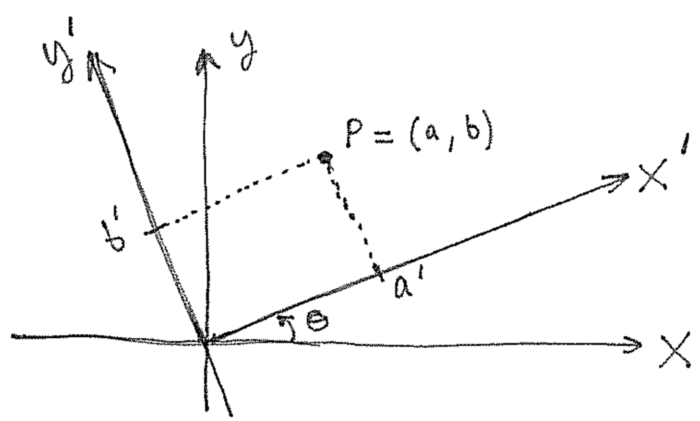
The inverse transformations to give x, y in terms of x', y' are similar

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x'\cos\theta - y'\sin\theta \\ x'\sin\theta + y'\cos\theta \end{bmatrix}$$

Let's find the equations of the primed coordinate axes.

1. *The y' axis has equation $x' = 0$ hence $x = -y'\sin(\theta)$ and $y = y'\cos(\theta)$ which yields $y = -\cot(\theta)x$ for $y' \neq 0$*
2. *Likewise, the x' axis has equation $y' = 0$ hence $x = x'\cos(\theta)$ and $y = x'\sin(\theta)$ which yields $y = \tan(\theta)x$ for $x' \neq 0$.*

Therefore the new primed axes are perpendicular to one another and are apparently rotated by angle θ in the clockwise direction as illustrated below.



rotated coordinates
Could also
draw two
planes to
See transformation.

2.7 orthogonal complements

Perhaps you've seen part of this Theorem before:

Proposition 2.7.1. *Pythagorean Theorem in n-dimensions*

If $x, y \in \mathbb{R}^{n\times 1}$ are orthogonal vectors then $||x||^2+||y||^2 = ||x+y||^2$. Moreover, if $x_1, x_2, \dots x_k$ are orthogonal then

$$||x_1||^2 + ||x_2||^2 + \cdots + ||x_k||^2 = ||x_1 + x_2 + \cdots + x_k||^2$$

The notation $W \leq V$ is meant to read "W is a **subspace** of V". A subspace is a subset of a vector space which is again a vector space with respect to the operations of V

Proposition 2.7.2. Existence of Orthonormal Basis

If $W \leq \mathbb{R}^{n\times 1}$ then there exists an orthonormal basis of W

The proof of the proposition above relies on an algorthim called **Gram-Schmidt orthogonalization**. That algorithm allows you to take any set of linearly indepedent vectors and replace it with a new set of vectors which are pairwise orthogonal.

Example 2.7.3. *For the record, the standard basis of $\mathbb{R}^{n\times 1}$ is an orthonormal basis and*

$$v = (v \cdot e_1)e_1 + (v \cdot e_2)e_2 + \cdots + (v \cdot e_n)e_n$$

for any vector v in $\mathbb{R}^{n\times 1}$.

Definition 2.7.4.

Suppose $W_1, W_2 \subseteq \mathbb{R}^{n\times 1}$ then we say W_1 is **orthogonal** to W_2 iff $w_1 \cdot w_2 = 0$ for all $w_1 \in W_1$ and $w_2 \in W_2$. We denote orthogonality by writing $W_1 \perp W_2$.

Definition 2.7.5.

Let V be a vector space and $W_1, W_2 \leq V$. If every $v \in V$ can be written as $v = w_1 + w_2$ for a unique pair of $w_1 \in W_1$ and $w_2 \in W_2$ then we say that V is the **direct sum** of W_1 and W_2. Moreover, we denote the statement "V is a direct sum of W_1 and W_2" by $V = W_1 \oplus W_2$.

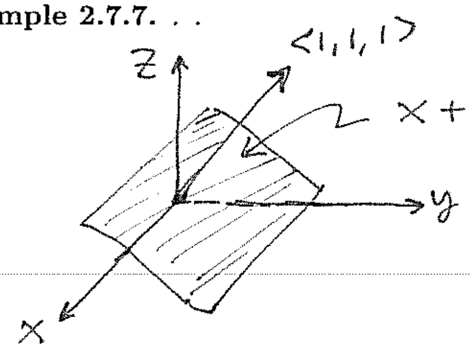
Proposition 2.7.6.

Let $W \leq \mathbb{R}^{n\times 1}$ then

1. $\mathbb{R}^{n\times 1} = W \oplus W^{\perp}$.
2. $dim(W) + dim(W^{\perp}) = n$,
3. $(W^{\perp})^{\perp} = W$,

Basically the cross-product is used in V^3 to select the perpendicular to a plane formed by two vectors. The theorem above tells us that if we wished to choose a perpendicular direction for a 2-dimensional plane inside V^5 then we would have a $5 - 2 = 3$-dimensional orthogonal complement to choose a "normal" for the plane. In other words, the concept of a normal vector to a plane is not so simple in higher dimensions. We could have a particular plane with two different "normal" vectors which were orthogonal!

Example 2.7.7. . . .



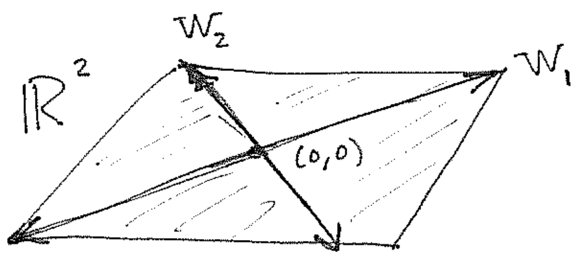
$x+y+z=0$ has normal $\langle 1,1,1\rangle$

$W_1 = \{(x,y,z) \mid x+y+z = 0\}$ = plane

$W_2 = \{t\langle 1,1,1\rangle \mid t \in \mathbb{R}\}$ = normal line.

$$\mathbb{R}^3 = W_1 \oplus W_2$$

Example 2.7.8. . . .



$W_1 = \text{span}\{\langle 1,1\rangle\}$

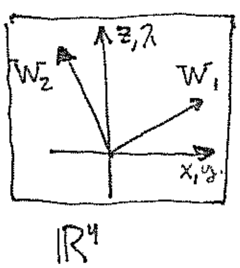
$W_2 = \text{span}\{\langle -1,1\rangle\}$

$$\mathbb{R}^2 = W_1 \oplus W_2$$

Example 2.7.9. . . .

Let $W_1 = \text{span}\{(1,1,1,1), (1,0,0,0)\}$

$\Rightarrow W_1 = \{s(1,1,1,1) + t(1,0,0,0) \mid s,t \in \mathbb{R}\}$



This is a plane in 4-dim'l space. Let's find its orthogonal complement.

$$W_2 = \{v \in \mathbb{R}^4 \mid v \cdot w = 0 \ \forall w \in W_1\}$$

need $v \cdot (1,1,1,1) = 0$ & $v \cdot (1,0,0,0) = 0$

Let $v = (x,y,z,\lambda)$ need $x+y+z+\lambda = 0$ & $x = 0$

Hence $y + z + \lambda = 0$ for $v = (x,y,z,\lambda) \in W_2$.

Thus $\underbrace{W_2 = \{(0,y,z,-y-z) \mid y,z \in \mathbb{R}\}}_{\text{this is normal to } W_1}$, $\underline{W_1 \oplus W_2 = \mathbb{R}^4}$.