Linear Algebra Notes

Stanley Li

May 7, 2025

Contents

1 Getting Started

1.1 Parametric representations of lines

Example 1.1. Suppose that L_1 and L_2 are lines in the plane, that the x-intercepts of L_1 and L_2 are 5 and -1, respectively, and that the respective y-intercepts are 5 and 1. Find the point of intersection of L_1 and L_2.

Solution 1.1.1. Pick two points on L_1, i.e. $(5, 0)$ and $(0, 5)$. Let $\mathbf{a} = \langle 5,\ 0\rangle$ and $\mathbf{b} = \langle 0,\ 5\rangle$. Now, L_1 can be represented by the following:

$$\begin{aligned} L_1 &= \{(\mathbf{a} - \mathbf{b})t + \mathbf{b} \mid t \in \mathbb{R}\} \\ &= \left\{ \begin{bmatrix} 5t \\ -5t + 5 \end{bmatrix} \middle| \, t \in \mathbb{R} \right\} \end{aligned}$$

The $(\mathbf{a} - \mathbf{b})t$ represents L_1 parameterized through the origin, so we apply a translation of $\mathbf{a}$ or $\mathbf{b}$. Similarly, let $\mathbf{c} = \langle -1, 0\rangle$ and $\mathbf{d} = \langle 0, 1\rangle$.

$$\begin{aligned} L_2 &= \{(\mathbf{c} - \mathbf{d})s + \mathbf{d} \mid s \in \mathbb{R}\} \\ &= \left\{ \begin{bmatrix} -s \\ -s + 1 \end{bmatrix} \middle| \, s \in \mathbb{R} \right\} \end{aligned}$$

The point of intersection is where $L_{1_x} = L_{2_x}$ and $L_{1_y} = L_{2_y}$, allowing us to define a system of equations.

$$5t = -s \qquad\qquad -5t + 5 = -s + 1$$

$$\begin{aligned} s &= -2 \\ x &= -s = 2 \\ y &= -s + 1 = 3 \end{aligned}$$

The point of intersection is (2, 3). ■

1.2 Linear dependence

Definition 1.1 (Linear combination)**.** Let $V = \{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_k}\}$ where $\mathbf{v_i} \in \mathbb{R}^n$. A *linear combination* of V is defined to be $c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_k\mathbf{v_k}$ where $c_i \in \mathbb{R}$.

Definition 1.2 (Linearly dependent)**.** Let $V = \{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_k}\}$ where $\mathbf{v_i} \in \mathbb{R}^n$. V is *linearly dependent* if and only if $c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_k\mathbf{v_k} = \mathbf{0}$ where $c_i \in \mathbb{R}$ and there exists at least one c_i such that $c_i \neq 0$. In other words, V is *linearly independent* if and only if $c_1, c_2, \dots, c_k = 0$ is the only solution.

Definition 1.3 (Span)**.** Given a set S of vectors, the *span* of S, denoted $\text{span}(S)$, is the set of all linear combinations of S.

Example 1.2. Let $S = \left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \right\}$. Is S linearly dependent?

Solution 1.2.1.

$$\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} c_1 + \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} c_2 + \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} c_3 = 0$$

$$\begin{aligned} c_1 + 2c_2 - c_3 &= 0 \\ -c_1 + c_2 &= 0 \\ 2c_1 + 3c_2 + 2c_3 &= 0 \end{aligned}$$

Solving this system, we find that the only solution is $c_1 = c_2 = c_3 = 0$. Thus, S is linearly independent. ■

We can go further by saying that the span of S is $\mathbb{R}^3$. This can be proven by setting the linear combination of S to an arbitrary 3-dimensional vector and isolating for the scalars c_1, c_2, and c_3, which tells us that any given vector in $\mathbb{R}^3$ can be represented in a specific linear combination of S. This can be thought of intuitively as well: if S is linearly independent, each vector of S introduces new directionality; there is no vector in S that can be defined as a linear combination of the other vectors.

1.3 Linear subspaces

Definition 1.4 (Linear subspace)**.** A set of vectors $V \subseteq \mathbb{R}^n$ is defined to be a linear/vector *subspace* of $\mathbb{R}^n$ if and only if it contains $\mathbf{0}$, it is closed under scalar multiplication, and it is closed under addition:

$$\begin{aligned} \mathbf{0} &\in V \\ \forall\ c \in \mathbb{R},\ c\mathbf{v} &\in V \\ \forall\ \mathbf{v_1}, \mathbf{v_2} \in V, \mathbf{v_1} + \mathbf{v_2} &\in V \end{aligned}$$

Theorem 1.1. *Let $V = \{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_n}\}$. It holds true that $\text{span}(V)$ is a subspace of $\mathbb{R}^n$.*

Proof. If $\text{span}(V)$ is a valid linear subspace, it must contain the zero vector, it must be closed under scalar multiplication, and it must be closed under addition.

By definition, the span of V is the set of all linear combinations of V:

$$\text{span}(V) = \{c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_n\mathbf{v_n}\ \forall\ c_i \in \mathbb{R}\}$$

1. Inclusion of zero vector:

$$\text{Let } c_1, c_2, \ldots, c_n = 0.$$
$$\implies c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_n\mathbf{v_n} = \mathbf{0}$$
$$\implies \mathbf{0} \in \text{span}(V)$$

2. Closure under scalar multiplication:

$$\text{Let } \mathbf{a} \in \text{span}(V).$$
$$\iff c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_n\mathbf{v_n} = \mathbf{a}$$
$$d(c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_n\mathbf{v_n}) = d\mathbf{a} \text{ where } d \in \mathbb{R}$$
$$\implies dc_1\mathbf{v_1} + dc_2\mathbf{v_2} + \cdots + dc_n\mathbf{v_n} = d\mathbf{a}$$

dc_i is just a scalar, meaning $d\mathbf{a}$ is another linear combination of V:

$$d\mathbf{a} \in \text{span}(V)$$

3. Closure under addition:

Let $\mathbf{a}, \mathbf{b} \in \text{span}(V)$.

$$\begin{aligned} c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_n\mathbf{v_n} &= \mathbf{a} \\ d_1\mathbf{v_1} + d_2\mathbf{v_2} + \cdots + d_n\mathbf{v_n} &= \mathbf{b} \\ (c_1 + d_1)\mathbf{v_1} + (c_2 + d_2)\mathbf{v_2} + \cdots + (c_n + d_n)\mathbf{v_n} &= \mathbf{a} + \mathbf{b} \end{aligned}$$

Again, $(c_i + d_i)$ is just a scalar, meaning $\mathbf{a} + \mathbf{b}$ is another linear combination of V:

$$\mathbf{a} + \mathbf{b} \in \text{span}(V)$$

■

Definition 1.5 (Basis)**.** Let $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ be linearly independent. It follows that S is a *basis* for the subspace $V = \text{span}(S)$.

Lemma 1.1. *Any vector in a subspace V is the result of a unique linear combination of some basis for V.*

Proof. Let $S = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ be a basis for the subspace V. Suppose $\mathbf{a} \in V$.

$$\mathbf{a} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_n\mathbf{v_n} \text{ where } c_i \in \mathbb{R}$$

Assume that $\mathbf{a}$ can be represented by another linear combination of S.

$$\begin{aligned} \mathbf{a} &= d_1\mathbf{v_1} + d_2\mathbf{v_2} + \cdots + d_n\mathbf{v_n} \text{ for some } d_j \neq c_j \\ \mathbf{0} &= (c_1 - d_1)\mathbf{v_1} + (c_2 - d_2)\mathbf{v_2} + \cdots + (c_n - d_n)\mathbf{v_n} \end{aligned}$$

S is linearly independent so the scalar $(c_i - d_i)$ must be zero for $1 \leq i \leq n$. This implies that $c_i = d_i$ which contradicts the statement that some $d_j \neq c_j$. This further implies that $\mathbf{a}$ cannot be represented by more than one linear combination of S. ■

1.4 Dot product

Definition 1.6 (Dot product)**.** The *dot product* of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, denoted $\mathbf{a} \cdot \mathbf{b}$, is defined to be $\mathbf{a} \cdot \mathbf{b} = \mathbf{a_1 b_1} + \mathbf{a_2 b_2} + \cdots + \mathbf{a_n b_n}$.

Lemma 1.2. $\mathbf{a} \cdot \mathbf{a} = \|\mathbf{a}\|^2$.

1.4.1 Properties

Dot products are commutative ($\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$), distributive ($\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$), and associative with scalars ($c(\mathbf{a} \cdot \mathbf{b}) = (c\mathbf{a}) \cdot \mathbf{b} = \mathbf{a} \cdot (c\mathbf{b})$). These properties can be easily proven with the definition of the dot product.

1.4.2 Geometric representation

The dot product of two vectors can also be represented in relation to the angle between them, θ, by the following identity: $\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \cos(\theta)$.

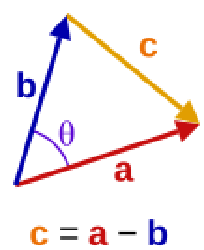

Figure 1: The angle θ in the triangle constructed by $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{a} - \mathbf{b}$

Proof. Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ such that $\mathbf{a}, \mathbf{b} \neq 0$. Using the Law of Cosines,

$$\begin{aligned}
\|\mathbf{a} - \mathbf{b}\|^2 &= \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\|\mathbf{a}\|\|\mathbf{b}\| \cos(\theta) \\
\|\mathbf{a}\|^2 - 2(\mathbf{a} \cdot \mathbf{b}) + \|\mathbf{b}\|^2 &= \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\|\mathbf{a}\|\|\mathbf{b}\| \cos(\theta) \\
-2(\mathbf{a} \cdot \mathbf{b}) &= -2\|\mathbf{a}\|\|\mathbf{b}\| \cos(\theta) \\
\mathbf{a} \cdot \mathbf{b} &= \|\mathbf{a}\|\|\mathbf{b}\| \cos(\theta)
\end{aligned}$$

■

1.4.3 Interpretation

The geometric representation makes it easy to recognize that $\mathbf{a} \cdot \mathbf{b}$ is maximized when $\theta = 0°$, which is when $\mathbf{a}$ and $\mathbf{b}$ are collinear in the same direction. On the other hand, $\mathbf{a} \cdot \mathbf{b}$ is minimized when $\theta = 180°$, which is when $\mathbf{a}$ and $\mathbf{b}$ are collinear in opposite directions. When $\mathbf{a}$ and $\mathbf{b}$ are perpendicular to each other (i.e. $\theta = 90°$), $\mathbf{a} \cdot \mathbf{b} = 0$. This means that we can interpret the dot product as a measure of collinearity.

Definition 1.7 (Orthogonal)**.** Vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ are *orthogonal* to each other if and only if $\mathbf{a} \cdot \mathbf{b} = 0$.

1.4.4 Cauchy-Schwarz Inequality

Lemma 1.3. *For some scalar* $c \geq 0$ *and some vector* $\mathbf{a}$*,* $c\|\mathbf{a}\| = \|c\mathbf{a}\|$*.*

Proof. Let $c \geq 0, c \in \mathbb{R}$. Let $\mathbf{a} \in \mathbb{R}^n$.

$$\begin{aligned} c\|\mathbf{a}\| &= c\sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} \\ &= \sqrt{c^2(a_1^2 + a_2^2 + \cdots + a_n^2)} \\ &= \sqrt{(ca_1)^2 + (ca_2)^2 + \cdots + (ca_n)^2} \\ &= \|c\mathbf{a}\| \end{aligned}$$

■

Theorem 1.2 (Cauchy-Schwarz inequality)**.** *If* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*, then* $\|\mathbf{x}\|\|\mathbf{y}\| \geq |\mathbf{x} \cdot \mathbf{y}|$*. Furthermore,* $\|\mathbf{x}\|\|\mathbf{y}\| = |\mathbf{x} \cdot \mathbf{y}|$ *if and only if* $\mathbf{y} = c\mathbf{x}$ *for some* $c \in \mathbb{R}$*.*

Proof. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ be nonzero. Let $p(t) = \|t\mathbf{y} - \mathbf{x}\|^2$. Note that $p(t) \geq 0$ for all $t \in \mathbb{R}$.

$$\begin{aligned} p(t) &= (t\mathbf{y} - \mathbf{x}) \cdot (t\mathbf{y} - \mathbf{x}) \\ &= t^2\|\mathbf{y}\|^2 - 2t(\mathbf{x} \cdot \mathbf{y}) + \|\mathbf{x}\|^2 \end{aligned}$$

Let $a = \|\mathbf{y}\|^2$. Let $b = 2(\mathbf{x} \cdot \mathbf{y})$. Let $c = \|\mathbf{x}\|^2$. Note that $a \neq 0$ because $\mathbf{y} \neq \mathbf{0}$.

$$\begin{aligned} p(t) &= at^2 - bt + c \geq 0 \\ p\left(\frac{b}{2a}\right) &= a\left(\frac{b}{2a}\right)^2 - b\left(\frac{b}{2a}\right) + c \\ &= \frac{b^2}{4a} - \frac{b^2}{2a} + c \\ &= -\frac{b^2}{4a} + c \geq 0 \\ \implies c &\geq \frac{b^2}{4a} \end{aligned}$$

Substituting back for a, b, and c:

$$\begin{aligned} \|\mathbf{x}\|^2 &\geq \frac{4(\mathbf{x} \cdot \mathbf{y})^2}{4\|\mathbf{y}\|^2} \\ \|\mathbf{x}\|^2\|\mathbf{y}\|^2 &\geq (\mathbf{x} \cdot \mathbf{y})^2 \\ \|\mathbf{x}\|\|\mathbf{y}\| &\geq |\mathbf{x} \cdot \mathbf{y}| \end{aligned}$$

Consider the case of the equality:

$$\begin{aligned} \|\mathbf{x}\|\|\mathbf{y}\| &= |\mathbf{x} \cdot \mathbf{y}| \\ \|\mathbf{x}\|\|\mathbf{y}\| &= |\|\mathbf{x}\|\|\mathbf{y}\| \cos(\theta)| \end{aligned}$$

This can be true if and only if $\theta = 0°$ or $\theta = 180°$, meaning $\mathbf{x}$ and $\mathbf{y}$ must be collinear with each other.

While we initially assumed that $\mathbf{x}$ and $\mathbf{y}$ were nonzero, it is easy to see that the inequality still holds when $\mathbf{x} = \mathbf{0}$ or $\mathbf{y} = \mathbf{0}$. ■

Theorem 1.3 (Triangle inequality). *If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. This inequality becomes an equality if and only if $\mathbf{y} = c\mathbf{x}$ for some $c \in \mathbb{R}$ such that $c \geq 0$. In other words, the sum of the lengths of two sides of a triangle is always greater than the length of its third side.*

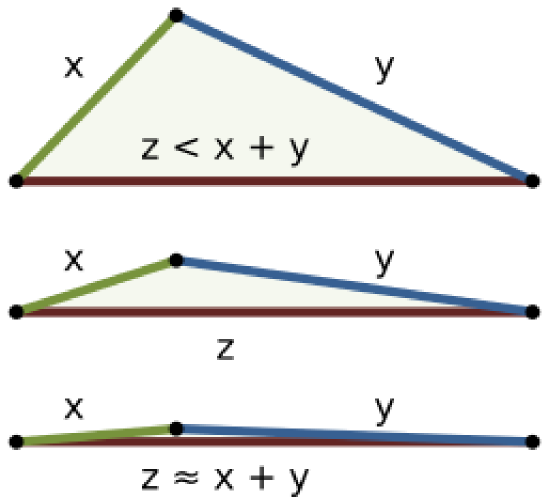

Figure 2: As two sides of a triangle get closer to being collinear, the sum of their lengths gets closer to the length of the third side.

Proof. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. By the Cauchy-Schwarz inequality (Theorem 1.2),

$$\|\mathbf{x}\|\|\mathbf{y}\| \geq |\mathbf{x} \cdot \mathbf{y}| \geq \mathbf{x} \cdot \mathbf{y}$$

Note that $|\mathbf{x} \cdot \mathbf{y}| = \mathbf{x} \cdot \mathbf{y}$ if and only if $\mathbf{x} \cdot \mathbf{y} \geq 0$. Else, $|\mathbf{x} \cdot \mathbf{y}| > \mathbf{x} \cdot \mathbf{y}$. Additionally, $|\mathbf{x} \cdot \mathbf{y}| = \|\mathbf{x}\|\|\mathbf{y}\|$ if and only if $\mathbf{y} = c\mathbf{x}$ for some $c \in \mathbb{R}$, as stated in the Cauchy-Schwarz inequality. So, if $c \geq 0$ and $\mathbf{y} = c\mathbf{x}$, then $\mathbf{x} \cdot \mathbf{y} = |\mathbf{x} \cdot \mathbf{y}| = \|\mathbf{x}\|\|\mathbf{y}\|$.

Now, consider the following:

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2(\mathbf{x} \cdot \mathbf{y}) + \|\mathbf{y}\|^2$$

Since $\|\mathbf{x}\|\|\mathbf{y}\| \geq \mathbf{x} \cdot \mathbf{y}$, we can create the inequality $\|\mathbf{x} + \mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2$. For the reason previously mentioned, this inequality becomes an equality if and only if $\mathbf{y} = c\mathbf{x}$ where $c \geq 0$.

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &\leq \ |\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 \\ \|\mathbf{x} + \mathbf{y}\|^2 &\leq (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 \\ \|\mathbf{x} + \mathbf{y}\| &\leq \|\mathbf{x}\| + \|\mathbf{y}\| \end{aligned}$$

■

1.5 Cross product

Definition 1.8 (Cross product)**.** The *cross product* of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$, denoted $\mathbf{a} \times \mathbf{b}$, is defined to be

$$\begin{aligned}\mathbf{a} \times \mathbf{b} &= \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \times \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \\ &= \begin{bmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{bmatrix}\end{aligned}$$

Remark. A way to easily remember how to calculate the cross product is by writing it as a determinant (covered in Section ??):

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

1.5.1 Properties

Cross products are anticommutative ($\mathbf{a}\times\mathbf{b} = -(\mathbf{b}\times\mathbf{a})$), distributive ($\mathbf{a}\times(\mathbf{b}+\mathbf{c}) = (\mathbf{a}\times\mathbf{b})+(\mathbf{a}\times\mathbf{c})$), and associative with scalars ($c(\mathbf{a} \times \mathbf{b}) = (c\mathbf{a}) \times \mathbf{b} = \mathbf{a} \times (c\mathbf{b})$). They are not associative themselves. These properties can be proven from the definition of the cross product.

1.5.2 Geometric representation

Theorem 1.4. *If $\mathbf{a} \times \mathbf{b} = \mathbf{c}$, then $\mathbf{c}$ is orthogonal to both $\mathbf{a}$ and $\mathbf{b}$.*

Proof. Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$.

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{bmatrix} = \mathbf{c}$$

$$\begin{aligned}\mathbf{a} \cdot \mathbf{c} &= a_1(a_2b_3 - a_3b_2) + a_2(a_3b_1 - a_1b_3) + a_3(a_1b_2 - a_2b_1) \\ &= a_1a_2b_3 - a_1a_3b_2 + a_2a_3b_1 - a_1a_2b_3 + a_1a_3b_2 - a_2a_3b_1 \\ &= 0 \\ \mathbf{b} \cdot \mathbf{c} &= b_1(a_2b_3 - a_3b_2) + b_2(a_3b_1 - a_1b_3) + b_3(a_1b_2 - a_2b_1) \\ &= a_2b_1b_3 - a_3b_1b_2 + a_3b_1b_2 - a_1b_2b_3 + a_1b_2b_3 - a_2b_1b_3 \\ &= 0\end{aligned}$$

∎

The direction of the normal vector obtained from the cross product $\mathbf{a} \times \mathbf{b}$ can be determined using the right-hand rule: point your index finger in the direction of $\mathbf{a}$, then point your middle finger in the direction of $\mathbf{b}$, then stick out your thumb. Your thumb represents the direction of the normal vector.

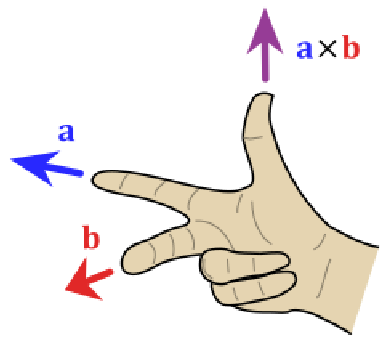

Figure 3: The right-hand rule

Similar to the dot product, the cross product of two vectors can also be represented in relation to the angle between them, θ, by the following identity: $\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\|\|\mathbf{b}\| \sin(\theta)$.

Proof. Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$.

$$\begin{aligned}
\mathbf{a} \times \mathbf{b} &= \begin{bmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{bmatrix} \\
\|\mathbf{a} \times \mathbf{b}\|^2 &= (a_2b_3 - a_3b_2)^2 + (a_3b_1 - a_1b_3)^2 + (a_1b_2 - a_2b_1)^2 \\
&= a_2^2b_3^2 - 2a_2a_3b_2b_3 + a_3^2b_2^2 + a_3^2b_1^2 - 2a_1a_3b_1b_3 + a_1^2b_3^2 + a_1^2b_2^2 - 2a_1a_2b_1b_2 + a_2^2b_1^2 \\
&= a_1^2(b_3^2 + b_2^2) + a_2^2(b_3^2 + b_1^2) + a_3^2(b_2^2 + b_1^2) - 2(a_2a_3b_2b_3 + a_1a_3b_1b_3 + a_1a_2b_1b_2)
\end{aligned}$$

$$\begin{aligned}
\|\mathbf{a}\|\|\mathbf{b}\| \cos(\theta) &= a_1b_1 + a_2b_2 + a_3b_3 \\
\|\mathbf{a}\|^2\|\mathbf{b}\|^2 \cos^2(\theta) &= (a_1b_1 + a_2b_2 + a_3b_3)^2 \\
&= a_1^2b_1^2 + a_1a_2b_1b_2 + a_1a_3b_1b_3 + a_1a_2b_1b_2 + a_2^2b_2^2 + a_2a_3b_2b_3 + a_1a_3b_1b_3 \\
&\quad + a_2a_3b_2b_3 + a_3^2b_3^2 \\
&= a_1^2b_1^2 + a_2^2b_2^2 + a_3^2b_3^2 + 2(a_1a_2b_1b_2 + a_1a_3b_1b_3 + a_2a_3b_2b_3)
\end{aligned}$$

$$\begin{aligned}
\|\mathbf{a} \times \mathbf{b}\|^2 + \|\mathbf{a}\|^2\|\mathbf{b}\|^2 \cos(\theta) &= a_1^2(b_3^2 + b_2^2) + a_2^2(b_3^2 + b_1^2) + a_3^2(b_2^2 + b_1^2) + a_1^2b_1^2 + a_2^2b_2^2 + a_3^2b_3^2 \\
&= a_1^2(b_1^2 + b_2^2 + b_3^2) + a_2^2(b_1^2 + b_2^2 + b_3^2) + a_3^2(b_1^2 + b_2^2 + b_3^2) \\
&= (b_1^2 + b_2^2 + b_3^2)(a_1^2 + a_2^2 + a_3^2) \\
&= (b_1^2 + b_2^2 + b_3^2)(a_1^2 + a_2^2 + a_3^2) \\
&= \|\mathbf{b}\|^2\|\mathbf{a}\|^2
\end{aligned}$$

$$\begin{aligned}
\|\mathbf{a}\times\mathbf{b}\|^2 + \|\mathbf{a}\|^2\|\mathbf{b}\|^2\cos(\theta) &= \|\mathbf{a}\|^2\|\mathbf{b}\|^2 \\
\|\mathbf{a}\times\mathbf{b}\|^2 &= \|\mathbf{a}\|^2\|\mathbf{b}\|^2(1-\cos^2(\theta)) \\
\|\mathbf{a}\times\mathbf{b}\|^2 &= \|\mathbf{a}\|^2\|\mathbf{b}\|^2\sin^2(\theta) \\
\|\mathbf{a}\times\mathbf{b}\| &= \|\mathbf{a}\|\|\mathbf{b}\|\sin^2(\theta)
\end{aligned}$$

■

1.5.3 Interpretation

From the geometric representation of the cross product, it is evident that $\|\mathbf{a}\times\mathbf{b}\|$ is maximized when $\theta = 90°$ and $\|\mathbf{a}\times\mathbf{b}\| = 0$ when $\theta = 0°$. This means that the cross product can be interpreted to be a measure of perpendicularity; its magnitude is maximized when **a** and **b** are perpendicular and minimized when they are collinear.

Theorem 1.5. *If* $\mathbf{a},\mathbf{b}\in\mathbb{R}^3$, *then* $\|\mathbf{a}\times\mathbf{b}\|$ *is equal to the area of the parallelogram formed with sides* **a** *and* **b**.

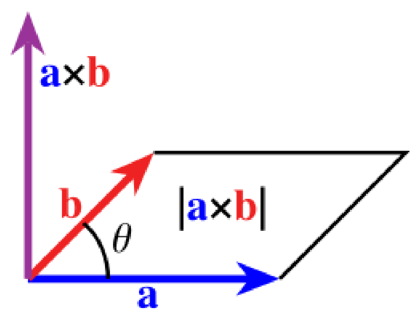

Figure 4: The area of the parallelogram with sides **a** and **b** is $\|\mathbf{a}\times\mathbf{b}\|$.

Proof. Let $\mathbf{a},\mathbf{b}\in\mathbb{R}^3$ form a parallelogram with base $\|\mathbf{b}\|$ and height h. It follows that $h = \|\mathbf{a}\|\sin(\theta)$.

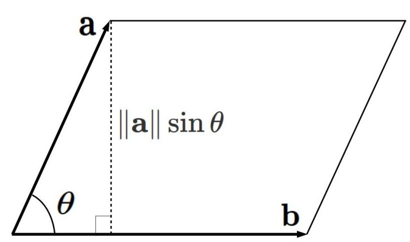

Figure 5: The parallelogram constructed by **a** and **b** with height $h = \|\mathbf{a}\|\sin(\theta)$.

The area of the parallelogram, given by its base times its height, is then $\|\mathbf{a}\|\|\mathbf{b}\|\sin(\theta) = \|\mathbf{a}\times\mathbf{b}\|$. ■

1.6 Planes in $\mathbb{R}^3$

The general equation of a plane is $Ax + By + Cz = D$ where A, B, C, and D are constants.

Finding the general equation of a plane. To start, we use the fact that $\mathbf{n} \cdot (\mathbf{x} - \mathbf{x_0}) = 0$ where $\mathbf{n}$ is the vector normal to the plane, $\mathbf{x}$ is an arbitrary vector specifying a point on the plane, and $\mathbf{x_0}$ is a vector specifying a known point on the plane. Notice that the vector $\mathbf{x} - \mathbf{x_0}$ lies on the plane so its dot product with the normal vector will be zero.

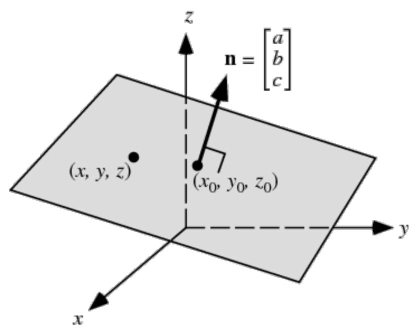

Figure 6: A plane with points (x, y, z) and (x_0, y_0, z_0) and normal vector $\mathbf{n}$.

$$\begin{aligned}
\mathbf{n} \cdot (\mathbf{x} - \mathbf{x_0}) &= 0 \\
\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} \cdot \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} &= 0 \\
n_1 x - n_1 x_0 + n_2 y - n_2 y_0 + n_3 z - n_3 z_0 &= 0 \\
n_1 x + n_2 y + n_3 z &= n_1 x_0 + n_2 y_0 + n_3 z_0
\end{aligned}$$

From this result, we can see that $A = n_1$, $B = n_2$, $C = n_3$, and $D = n_1 x_0 + n_2 y_0 + n_3 z_0$. Note that if we keep the normal vector the same but change the point $\mathbf{x_0}$ (i.e. creating a parallel plane), then the only value that changes is D. This means that changing D will only shift the plane while keeping it parallel.

Finding the minimum distance between a point and a plane. Suppose $\mathbf{x_1} \in \mathbb{R}^3$ specifies a point that is not necessarily on the plane $Ax + By + Cz = D$. Let d be the minimum distance from the plane to $\mathbf{x_1}$. This implies that $\mathbf{d}$ is normal to the plane. Let $\mathbf{x_0}$ specify a point on the plane. Let $\mathbf{f} = \mathbf{x_1} - \mathbf{x_0}$. Let θ be the angle between d and $\mathbf{f}$. Let $\mathbf{n}$ be a normal vector to the plane.

$$\begin{aligned}
\cos(\theta) &= \frac{d}{\|\mathbf{f}\|} \\
\|\mathbf{f}\| \cos(\theta) &= d \\
\frac{\|\mathbf{n}\| \|\mathbf{f}\| \cos(\theta)}{\|\mathbf{n}\|} &= d \\
\frac{\mathbf{n} \cdot \mathbf{f}}{\|\mathbf{n}\|} &= d \\
\frac{\mathbf{n} \cdot (\mathbf{x_1} - \mathbf{x_0})}{\|\mathbf{n}\|} &= d \\
\frac{Ax_1 + By_1 + Cz_1 - D}{\sqrt{A^2 + B^2 + C^2}} &= d
\end{aligned}$$

2 $\mathbf{Ax} = \mathbf{b}$ and the Four Subspaces

Definition 2.1 (Matrix multiplication)**.** If $\mathbf{A}$ is a matrix of size $m \times n$ and $\mathbf{B}$ is a matrix of size $n \times p$, then

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix}$$
$$= \begin{bmatrix} a_{11}b_{11} + \cdots + a_{1n}b_{n1} & a_{11}b_{12} + \cdots + a_{1n}b_{n2} & \cdots & a_{11}b_{1p} + \cdots + a_{1n}b_{np} \\ a_{21}b_{11} + \cdots + a_{2n}b_{n1} & a_{21}b_{12} + \cdots + a_{2n}b_{n2} & \cdots & a_{21}b_{1p} + \cdots + a_{2n}b_{np} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{11} + \cdots + a_{mn}b_{n1} & a_{m1}b_{12} + \cdots + a_{mn}b_{n2} & \cdots & a_{m1}b_{1p} + \cdots + a_{mn}b_{np} \end{bmatrix}$$

This is like taking the dot product of the rows of $\mathbf{A}$ with the columns of $\mathbf{B}$. Notice that the result of $\mathbf{AB}$ is a matrix of size $m \times p$. This definition is extended to matrix-vector multiplication by treating a n-dimensional vector as a matrix of size $n \times 1$.

In the case of matrix-vector multiplication, an equivalent and commonly more useful computation is taking the linear combination of the column vectors of $\mathbf{A}$ where the scalars are the corresponding components in $\mathbf{x}$; if $\mathbf{A} = \begin{bmatrix} \mathbf{v_1} & \mathbf{v_2} & \cdots & \mathbf{v_n} \end{bmatrix}$, then $\mathbf{Ax} = x_1\mathbf{v_1} + x_2\mathbf{v_2} + \cdots + x_n\mathbf{v_n}$.

Theorem 2.1 (Associativity of matrix multiplication)**.** *If* $\mathbf{A} \in F^{m\times n}$, $\mathbf{B} \in F^{n\times p}$, *and* $\mathbf{C} \in F^{p\times q}$, *then* $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

Proof. Note that $(\mathbf{AB})_{m\times p}\mathbf{C}_{p\times q}$ is a valid product resulting in an $m\times q$ matrix, and $\mathbf{A}_{m\times n}(\mathbf{BC})_{n\times q}$ is similarly well-defined and also results in an $m \times q$ matrix. To prove that $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$, we just have to show that every entry is equivalent: $((\mathbf{AB})\mathbf{C})_{ij} = (\mathbf{A}(\mathbf{BC}))_{ij}$ for $1 \leq i \leq m$ and $1 \leq j \leq q$.

$$\begin{aligned} LHS &= ((\mathbf{AB})_{m\times p}\mathbf{C}_{p\times q})_{ij} \\ &= \sum_{k=1}^{p} (\mathbf{AB})_{ik}\mathbf{C}_{kj} \end{aligned}$$

This is derived from the definition of matrix multiplication; a resulting entry at the ith row and jth column is the "dot product" between the ith row of the first matrix, $\mathbf{AB}$, and the jth column of the second matrix, $\mathbf{C}$. We will similarly express $(\mathbf{A}_{m\times n}\mathbf{B}_{n\times p})_{ik}$ as a sum.

$$= \sum_{k=1}^{p} (\sum_{l=1}^{n} \mathbf{A}_{il}\mathbf{B}_{lk})\mathbf{C}_{kj}$$

By the distributive property,

$$= \sum_{k=1}^{p} \sum_{l=1}^{n} \mathbf{A}_{il}\mathbf{B}_{lk}\mathbf{C}_{kj}$$

We'll do the same for the right-hand side.

$$\begin{aligned}
RHS &= (\mathbf{A}_{m\times n}(\mathbf{BC})_{n\times q})_{ij} \\
&= \sum_{l=1}^{n} \mathbf{A}_{il}(\mathbf{B}_{n\times p}\mathbf{C}_{p\times q})_{lj} \\
&= \sum_{l=1}^{n} \mathbf{A}_{il}(\sum_{k=1}^{p} \mathbf{B}_{lk}\mathbf{C}_{kj}) \\
&= \sum_{l=1}^{n}\sum_{k=1}^{p} \mathbf{A}_{il}\mathbf{B}_{lk}\mathbf{C}_{kj}
\end{aligned}$$

Since fields are commutative, we can swap the summations.

$$= \sum_{k=1}^{p}\sum_{l=1}^{n} \mathbf{A}_{il}\mathbf{B}_{lk}\mathbf{C}_{kj}$$

$LHS = RHS$. ■

2.1 Elimination with matrices

We can set up a *coefficient matrix* to solve a linear system of equations. Then, in the process called *Gaussian elimination*, we manipulate the rows of the matrix by combining different multiples of each row. The desired result is a matrix in *row echelon form*, which contains *pivot entries* in each column. A pivot entry is a nonzero entry which sits below and to the right of the previous pivot entry. Every entry under the pivot entry must be zero. The first entry of a matrix in row echelon form should be a pivot entry.

Example 2.1. Solve the following linear system:

$$\begin{array}{rcrcrcr}
x & + & 2y & + & z & = & 2 \\
3x & + & 8y & + & z & = & 12 \\
 & & 4y & + & z & = & 2
\end{array}$$

Solution 2.1.1. Let $\mathbf{A}$ be the coefficient matrix. We will use Gaussian elimination to put $\mathbf{A}$ in row echelon form.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} \xrightarrow{\mathbf{r_2}-3\mathbf{r_1}} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix} \xrightarrow{\mathbf{r_3}-2\mathbf{r_2}} \begin{bmatrix} \underline{1} & 2 & 1 \\ 0 & \underline{2} & -2 \\ 0 & 0 & \underline{5} \end{bmatrix} = \mathbf{U}$$

The underlined entries in the $\mathbf{U} = \text{ref}(\mathbf{A})$ are the pivot entries. While this example worked out well (we were able to find $\text{ref}(\mathbf{A})$), the process is not always straightforward. For example, if the first original entry of $\mathbf{A}$ was 0, we would have to swap the first row with a suitable row beneath. Likewise, if we came across a zero in a pivot position in a later step, we could again try to exchange the row with a suitable row beneath. Still, there are cases where a pivot entry cannot be found.

Let's repeat the same process but with an augmented matrix this time. An *augmented matrix* adds the column containing the solutions to the coefficient matrix.

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 3 & 8 & 1 & 12 \\ 0 & 4 & 1 & 2 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 4 & 1 & 2 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 0 & 5 & -10 \end{array}\right]$$

We can rewrite the system of equations:

$$x + 2y + z = 2 \qquad 2y - 2z = 6 \qquad 5z = -10$$

$$\begin{aligned} z &= -2 \\ 2y - 2(-2) = 6 &\implies y = 1 \\ x + 2(1) + (-2) = 2 &\implies x = 2 \end{aligned}$$

■

Let's try to solve the same system using matrix multiplication to show our steps. First of all, suppose $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{B} = \begin{bmatrix} \mathbf{r_1} \\ \vdots \\ \mathbf{r_n} \end{bmatrix}$ where $\mathbf{r_i} \in \mathbb{R}^m$ is a row vector. Then,

$$\begin{aligned} \mathbf{x}^T\mathbf{B} &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \mathbf{r_1} \\ \vdots \\ \mathbf{r_n} \end{bmatrix} \\ &= x_1\mathbf{r_1} + x_2\mathbf{r_2} + \cdots + x_n\mathbf{r_n} \end{aligned}$$

Note that the result is a row vector of size $1 \times m$. This row-matrix multiplication, which yields a linear combination of the rows of a matrix, is analagous to matrix-vector multiplication which yields a linear combination of the columns of a matrix.

Suppose now that $\mathbf{r_i}$ is a 1×3 row vector. Consider this operation:

$$\begin{bmatrix} \mathbf{r_1} \\ \mathbf{r_2} \\ \mathbf{r_3} \end{bmatrix} \to [\mathbf{r_1}].$$

Clearly, we can represent the operation using this row-matrix product:

$$\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{r_1} \\ \mathbf{r_2} \\ \mathbf{r_3} \end{bmatrix} = [\mathbf{r_1}].$$

Now consider this operation:

$$\begin{bmatrix} \mathbf{r_1} \\ \mathbf{r_2} \\ \mathbf{r_3} \end{bmatrix} \to \begin{bmatrix} \mathbf{r_1} \\ \mathbf{r_2} \\ \mathbf{r_3} \end{bmatrix}.$$

Firstly, the matrix we select should have a size of 3×3. The first row of the result is $\mathbf{r_1}$, just like in the previous example. So, it makes sense to make the first row of the matrix $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, just like before. Similarly, to get $\mathbf{r_2}$ in the second row of the result, we can make the second row of the matrix $\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$. Then, we do the same for the third row. The matrix multiplication looks like this:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{r_1} \\ \mathbf{r_2} \\ \mathbf{r_3} \end{bmatrix} = \begin{bmatrix} \mathbf{r_1} \\ \mathbf{r_2} \\ \mathbf{r_3} \end{bmatrix}$$

The matrix we just created is a special matrix called the *identity matrix.* Specifically, this is a 3×3 identity matrix, typically labelled I_3.

Definition 2.2 (Identity matrix)**.** An $n \times n$ identity matrix, denoted I_n, is a square matrix with ones along its diagonal and zeros everywhere else. An identity matrix has the property that $I_m\mathbf{C} = \mathbf{C}$ where $\mathbf{C}$ is a $m \times n$ matrix.

Theorem 2.2. *Suppose* $\mathbf{E} \in F^{m\times m}$ *and* $\mathbf{A} \in F^{m\times n}$*. Let* $\mathbf{r_i} \in F^{1\times n}$ *be the row vector representing the* i*th row of* $\mathbf{A}$ *for* $i = 1, \ldots, m$*. Then, an equivalent "row-based" matrix multiplication can be performed like this:*

$$\mathbf{EA} = \begin{bmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \ddots & \vdots \\ e_{m1} & \cdots & e_{mm} \end{bmatrix} \begin{bmatrix} \mathbf{r_1} \\ \vdots \\ \mathbf{r_m} \end{bmatrix} = \begin{bmatrix} e_{11}\mathbf{r_1} + \cdots + e_{1m}\mathbf{r_m} \\ \vdots \\ e_{m1}\mathbf{r_1} + \cdots + e_{mm}\mathbf{r_m} \end{bmatrix}$$

Proof. Since $\mathbf{r_i}$ is the row vector representing the ith row of $\mathbf{A}$, it has the form $\mathbf{r_i} = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}$.

$$\begin{aligned}
\mathbf{EA} &= \begin{bmatrix} e_{11}a_{11} + \cdots + e_{1m}a_{m1} & \cdots & e_{11}a_{1n} + \cdots + e_{1m}a_{mn} \\ e_{21}a_{11} + \cdots + e_{2m}a_{m1} & \cdots & e_{21}a_{1n} + \cdots + e_{2m}a_{mn} \\ \vdots & \ddots & \vdots \\ e_{m1}a_{11} + \cdots + e_{mm}a_{m1} & \cdots & e_{m1}a_{1n} + \cdots + e_{mm}a_{mn} \end{bmatrix} \\
&= e_{11} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} + \cdots + e_{1m} \begin{bmatrix} a_{m1} & \cdots & a_{mn} \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} + e_{21} \begin{bmatrix} 0 & \cdots & 0 \\ a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \\
&+ \cdots + e_{2m} \begin{bmatrix} 0 & \cdots & 0 \\ a_{m1} & \cdots & a_{mn} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} + \cdots + e_{m1} \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \\ a_{11} & \cdots & a_{1n} \end{bmatrix} + \cdots + e_{mm} \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \\
&= \begin{bmatrix} & e_{11}\mathbf{r_1} & \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} + \cdots + \begin{bmatrix} & e_{1m}\mathbf{r_m} & \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} + \begin{bmatrix} 0 & \cdots & 0 \\ & e_{21}\mathbf{r_1} & \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 & \cdots & 0 \\ & e_{2m}\mathbf{r_m} & \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \\
&+ \cdots + \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \\ & e_{m1}\mathbf{r_1} & \end{bmatrix} + \cdots + \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \\ & e_{mm}\mathbf{r_m} & \end{bmatrix} \\
&= \begin{bmatrix} e_{11}\mathbf{r_1} + \cdots + e_{1m}\mathbf{r_m} \\ e_{21}\mathbf{r_1} + \cdots + e_{2m}\mathbf{r_m} \\ \vdots \\ e_{m1}\mathbf{r_1} + \cdots + e_{mm}\mathbf{r_m} \end{bmatrix}
\end{aligned}$$

■

Solution 2.1.2. Going back to the original problem, let's create an *elementary matrix* to perform the first row operation from Solution 2.1.1 (assigning row 2 to $\mathbf{r_2} - 3\mathbf{r_1}$). Since this operation was done to remove the entry at the 2nd row and 1st column, we'll call this elementary matrix $\mathbf{E_{21}}$:

$$\mathbf{E_{21}A} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix}$$

Repeat the process:

$$\mathbf{E_{32}}(\mathbf{E_{21}A}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{bmatrix}$$