Systems of Equations

(4/12/19)

Alex Nita

1 Definitions and Notation

Definition 1.1 The generic **system of m linear equations in n unkowns**, over the field of numbers $F = \mathbb{R}$ or $\mathbb{C}$,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

may be condensed, using matrix and vector notation, to

$$A\mathbf{x} = \mathbf{b}$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in F^n = \mathbb{R}^n \quad \text{or} \quad \mathbb{C}^n$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in M_{m,n}(\mathbb{R}) \quad \text{or} \quad M_{m,n}(\mathbb{C}),$$

$$\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in F^m = \mathbb{R}^m \quad \text{or} \quad \mathbb{C}^m$$

A is called the **coefficient matrix** of the system, and its components $a_{ij} \in F = \mathbb{R}$ or $\mathbb{C}$ the **coefficients** of the system. The system is called **homogeneous** if $\mathbf{b} = \mathbf{0}$, and **nonhomogeneous** otherwise. A **solution** to the equation is a particular vector $\mathbf{s} \in \mathbb{R}^n$ *satisfying* or *solving* the equation, $A\mathbf{s} = \mathbf{b}$. The **solution set**, the set of all solutions, is denoted (by me)

$$S_{A,\mathbf{b}} \stackrel{\text{def}}{=} \{\mathbf{s} \in F^n \mid A\mathbf{s} = \mathbf{b}\}$$

■

Definition 1.2 To study matrices in further detail, we will need to perform **elementary row and column operations** on them. If $A \in M_{m,n}(F)$, any one of the following operations on the rows of A is called an elementary row operation

- **(type 1)** Interchanging any two rows (resp. columns) of A.
- **(type 2)** Multiplying any row (resp. column) of A by a nonzero scalar.
- **(type 3)** Adding any scalar multiple of a row (resp. column) of A to another row/column.

Suppose $E \in M_n(F)$ obtained by performing an elementary operation on I_n. Then E is called a **type 1 elementary matrix**, or a **type 2 elementary matrix** if the operation is type 2, or else a **type 3 elementary matrix** if the operation is type 3.

Observation 1.3 If E is an elementary, then left-multiplication of A by E produces the the corresponding row-operation on A. ■

Example 1.4 For example, suppose

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \in M_{3,2}(\mathbb{R})$$

and we want to swap rows 2 and 3. Then we left-multiply by the corresponding type 1 elementary matrix in $M_3(\mathbb{R})$ gotten from I_3 by swapping *its* rows 2 and 3:

$$EA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \\ a_{21} & a_{22} \end{pmatrix} \qquad \begin{pmatrix} \vec{A}_1 \\ \vec{A}_2 \\ \vec{A}_3 \end{pmatrix} \mapsto \begin{pmatrix} \vec{A}_1 \\ \vec{A}_3 \\ \vec{A}_2 \end{pmatrix}$$

Or suppose we performed a type 2 row-operation on A, say scaling row 3 by 2. Then we left-multiply by $E \in M_3(\mathbb{R})$ gotten from I_3 by scaling row 3 by 2:

$$EA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ 2a_{31} & 2a_{32} \end{pmatrix} \qquad \begin{pmatrix} \vec{A}_1 \\ \vec{A}_2 \\ \vec{A}_3 \end{pmatrix} \mapsto \begin{pmatrix} \vec{A}_1 \\ \vec{A}_2 \\ 2\vec{A}_3 \end{pmatrix}$$

Or, finally, suppose we wanted to perform the type 3 row-operation of adding 2 of row 2 to row 1. We first perform it on I_3 and then left-multiply by the result:

$$EA = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} = \begin{pmatrix} a_{11} + 2a_{21} & a_{12} + 2a_{22} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \qquad \begin{pmatrix} \vec{A}_1 \\ \vec{A}_2 \\ \vec{A}_3 \end{pmatrix} \mapsto \begin{pmatrix} \vec{A}_1 + 2\vec{A}_2 \\ \vec{A}_2 \\ \vec{A}_3 \end{pmatrix}$$

■

Definition 1.5 We also have **column operations** of the same 3 types: swapping two columns, scaling a column, and adding a multiple of a column to another column. Accordingly, we have 3 **elementary matrices** for column operations, with the difference that we right-multiply $A \in M_{m,n}(F)$ by $E \in M_m(F)$.

Example 1.6 Let us take A from the above and swap its first and second columns: we do this by right-multiplying A by E which is I_2 with *its* two columns swapped:

$$AE = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a_{12} & a_{11} \\ a_{22} & a_{21} \\ a_{32} & a_{31} \end{pmatrix} \qquad (\mathbf{a}_1\ \mathbf{a}_2) \mapsto (\mathbf{a}_2\ \mathbf{a}_1)$$

Or suppose we scaled column 1 by -3:

$$AE = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} -3 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -3a_{11} & a_{12} \\ -3a_{21} & a_{22} \\ -3a_{31} & a_{32} \end{pmatrix} \qquad (\mathbf{a}_1\ \mathbf{a}_2) \mapsto (-3\mathbf{a}_1\ \mathbf{a}_2)$$

Or, finally, say we add -2 of column 1 to column 2:

$$AE = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} - 2a_{11} \\ a_{21} & a_{22} - 2a_{21} \\ a_{31} & a_{32} - 2a_{31} \end{pmatrix} \qquad (\mathbf{a}_1\ \mathbf{a}_2) \mapsto (\mathbf{a}_1\ \mathbf{a}_2 - 2\mathbf{a}_1)$$

We need row operations to perform row-reduction:

Definition 1.7 (Row-Reduced Echelon Form) A matrix $A \in M_{m,n}(F)$ is said to be in **reduced row echelon form** if, counting from the topmost row to the bottom-most,

(1) Any nonzero row precedes any zero rows (if any).

(2) The first nonzero entry a_{ij} in each nonzero row is the only nonzero entry in its column $\mathbf{a}_j$.

(3) The first nonzero entry in each nonzero row is 1 and it occurs in a column to the right of the first nonzero entry in the preceding row. ■

Example 1.8 The following matrices are **not** in reduced echelon form. The first fails (2) and (3)

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

The second fails (3),

$$\begin{pmatrix} 0 & 1 & 0 & 2 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

The third fails (3),

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

■

Example 1.9 A matrix that *is* in reduced row echelon form is:

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$ ■

Notation 1.10 *Let $\mathcal{E}_n$ denote the* ***set of all elementary*** *$n \times n$* ***matrices****.* ■

Lemma 1.11 *Every elementary matrix is invertible, with inverse an elementary matrix of the same type, so that $\mathcal{E}_m$ is a* <u>*subset*</u> *of* $\mathrm{GL}(n, F)$ *(though not a subgroup). The same applies to column-operation elementary matrices.*

Proof: Any row-swapping type-1 matrix is its own inverse. Any type-2 matrix scaling the ith row by $c \neq 0$ has inverse the type-2 matrix scaling row i by c^{-1}. Any type-3 matrix, adding k times row i to row j has as inverse the type-3 matrix addint $-k$ times row i to row j. ■

Remark 1.12 The product of elementary matrices is generally no longer an elementary matrix. However, the set of all finite products of type I and type III matrices in $\mathcal{E}$, the so-called **transvections**, *do* form a subgroup of $\mathrm{GL}(n, F)$, the **special linear group**,

$$\mathrm{SL}_n^{\pm}(F) = \{A \in \mathrm{GL}(n, F) \mid \det A = \pm 1\}$$

See Lang's two books *Algebra* and $SL_2(\mathbb{R})$ (this is the title of the book). ■

1.1 Interlude on Equivalence Relations and Equivalence Classes

In modern math, a distinction is made between objects which are ***equal*** and objects which are ***equivalent***. Two mathematical objects (invariably sets of some sort) are considered *equal*, by cultural fiat, exactly when they are *indistinguishable*, meaning they *possess the same members* (they are sets, after all), and are accordingly treated as *one and the same object*. Theorems and propositions sometimes assert such equalities, as for example Proposition 3.5 in the 'Matrices and Linear Transformations' notes, which says $\mathrm{Col}(A) = R(A)$.

Equivalent objects, on the other hand, are not equal, they are *the same but different*: this is the idea of *similarity among different individuals*—a thorny idea, since 'similarity' tends to refer to the same properties possessed by *different individual objects*. Modernity has largely discarded the Platonic idea that *the same properties* literally sit in *different objects*. For example, speaking informally, different objects seem to sometimes possess 'the same color' or 'the same shape,' but how this is *literally* possible is unclear, since that 'sameness' proceeds from an *observational judgement* made by us, and is therefore, like all observations, only *approximate*—at least in terms of comparable observables. It seems to be a question of *accuracy* in a certain specific sense. Philosophical skepticism has put us in doubt as to the soundness of knowledge of things *unverified*. Statements of fact need somehow to be *verifiable*. **Measurement**, of course, is the traditional way to assuage the skeptic, who likes to *compare* and *contrast* things before accepting a statement of fact as *true*. Sameness in difference is acceptable, approximately, whenever that sameness is measurable, meaning *comparable* to some *standard* of measurement—the **unit of measurement**. Two different material objects have 'the same' length if, upon comparison with a yardstick, their dimensions provide pretty much *the same readings* off the yardstick.

Because modern mathematics is in line with this way of thinking, it, too, takes comparison as the answer to such conundrums—with the important difference that math deals in *symbolic representation* rather than observational data, so the issue surfaces indirectly, in the somewhat cryptic language of 'partitions of sets,' 'equivalence relations' and 'equivalence classes.' But there is an easy way to understand this, if you look at some examples from *geometry*, which is the symbolic representation of space and spatial relations:

- **Congruence** between shapes, e.g. triangles, is a type of equivalence. Two different triangles can be *congruent*, which is a way of saying 'the same' in respect of their 'measures,' e.g. their side lengths and angles.
- **Similarity** between shapes is another type of equivalence, but this time in respect only of their *proportions*, such as, with triangles, the the *ratios* of the measures of their side lengths.

Definition 1.13 The first step in formalizing this in terms of sets and relations comes with the idea of an **equivalence relation**, $\sim$, on a set X. This is a relation of 'sameness' in the formal sense that:

1. Each element a of a set X is 'similar' or *equivalent* to itself, so $a \sim a$. This is called the **reflexivity** property of the equivalence relation.

2. If an element a in X is equivalent to another element b in X, so that $a \sim b$, then b is also equivalent to a, $b \sim a$. This is the **symmetry** property of $\sim$.

3. If $a \sim b$ and $b \sim c$, then $a \sim c$, which is the **transitivity** property of $\sim$.

The set X is supposed to represent the 'generic collection,' applicable to any given scientific task which is well defined enough to isolate a *class of objects* for its investigation. An equivalence relation $\sim$ on X thus captures, through a clever use of only 3 properties, what we *want* of our formalized concept of 'sameness' among *different* elements. We have transferred the problem of uniting different individuals under one 'similar' property into the more practical idea of *a special type of relationship between* individuals situated inside a definite container. Once we put the matter like this, we simply carve up the set X, the container, into tidy **equivalence classes** or sub-containers of 'similar' objects. The equivalence class is usually **represented** by one member, a, all other elements in its class being equivalent to *it*. The equivalence class of a is denoted

$$\boxed{[a] \stackrel{\text{def}}{=} \{x \in X \mid x \sim a\}}$$

The *set of all equivalence classes* of $\sim$ on X is denoted

$$\boxed{X/\sim \stackrel{\text{def}}{=} \big\{[a] \mid a \in X\big\}}$$

and is called the **quotient set on** X **by** $\sim$. Of course, if $a \sim b$, then $[a] = [b]$, by transitivity, and if $a \nsim b$, then $[a] \cap [b] = \varnothing$, the **empty set** with no elements. This is how X gets *partitioned* into equivalence classes

$$X = \bigsqcup_{[a] \in X/\sim} [a]$$

the *square* cup denoting *disjoint union* (meaning the union $X = \bigcup_{[a]\in X/\sim}[a]$ for which $[a] \neq [b]$ implies $[a] \cap [b] = \varnothing$, i.e. $[a]$ and $[b]$ are disjoint).

Example 1.14 (**Integers Mod** p) On the set of all integers $\mathbb{Z}$ we have the equivalence relation of **congruence mod** 2, which is defined in terms of *divisibility by* 2,

$$n \equiv m \mod 2 \quad \stackrel{\text{def}}{\iff} \quad n - m = 2k, \quad \text{for some } k \in \mathbb{Z}$$

For instance, $3 \equiv -9 \mod 2$, because $3 - (-9) = 12 = 2 \cdot 6$. But $2 \equiv 6 \mod 2$ as well, whereas $2 \not\equiv 3$, because $2 - 3 = -1$, which is not a multiple of 2. This shows that the **integers may be partitioned into two equivalence classes**, the **evens** and the **odds**,

$$\mathbb{Z} = [0] \sqcup [1]$$

The quotient set is called the **ring of integers modulo** 2, and has special notation,

$$\mathbb{Z}_2 \quad \text{or} \quad \mathbb{Z}/2\mathbb{Z} \stackrel{\text{def}}{=} \big\{[0],\ [1]\big\}$$

The idea can be repeated for numbers other than 2. In general, for any $p \in \mathbb{N}$, **congruence mod** p is the equivalence relation on $\mathbb{Z}$ given by

$$n \equiv m \mod p \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad n - m = 2p, \quad \text{for some } k \in \mathbb{Z}$$

and the quotient set is the ring of all integers mod p,

$$\mathbb{Z}_p \quad \text{or} \quad \mathbb{Z}/p\mathbb{Z} \stackrel{\text{def}}{=} \big\{[0],\ [1], \dots, [p-1]\big\} \qquad \blacksquare$$

Example 1.15 (**Projective Sphere and Projective Space**) Unit vectors in $\mathbb{R}^3$ constitute the unit sphere $S^2 = \{\mathbf{x} \in \mathbb{R}^3 \mid \|\mathbf{x}\| = 1\}$. Let us put an equivalence relation on S^2, in terms of *negation*, calling $-\mathbf{x}$ the **antipodal point** of $\mathbf{x} \in S^2$,

$$\mathbf{x} \sim \mathbf{y} \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad \mathbf{y} = -\mathbf{x}$$

whose equivalence classes are pairs of antipodal points,

$$[\mathbf{x}] = \{\mathbf{x}, -\mathbf{x}\}$$

The quotient set $S^2/\sim\, = \{[\mathbf{x}]\,|\,\mathbf{x} \in S^2\}$ is called the **projective sphere**. It is similar (in fact *homeomorphic*) to **projective space**, which is a quotient space of $\mathbb{R}^3\backslash\{\mathbf{0}\}$ by the *scaling* (now by *any* real number c) equivalence relation,

$$\mathbf{x} \sim \mathbf{y} \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad \mathbf{y} = c\mathbf{x} \ \text{ for some } c \in \mathbb{R}$$

denoted

$$\mathbb{RP}^2 \quad \text{or} \quad \mathbb{P}^2(\mathbb{R}) \stackrel{\text{def}}{=} \big\{[\mathbf{x}] \mid \mathbf{x} \in \mathbb{R}^3\backslash\{\mathbf{0}\}\big\} \qquad \blacksquare$$

Theorem 1.16 *Let X be a set and $\mathcal{P} = \{A_i \subseteq X \mid i \in I\}$ a partition of X, that is, a collection of pairwise disjoint subsets whose union is X,*

$$X = \bigsqcup_{i \in I} A_i$$

meaning $X = \bigcup_{i \in I} A_i$ and $A_i \cap A_j = \varnothing$ if $i \neq j$. Then there is exactly one equivalence relation on X from which this partition is derived.

Proof: Define a binary relation $\sim$ on X by setting $x \sim y$ if $x, y \in A_i$ for some $A_i \in \mathcal{P}$. Now, $\sim$ is obviously symmetric, reflexive and transitive, and the relation applies to all elements of X, since $\bigsqcup_{i \in I} = X$ and all elements of $\mathcal{P}$ are pairwise disjoint. Now, suppose there were two equivalence relations $\sim_1$ and $\sim_2$ on X with the above properties. Then, for all $a \in X$ let $[a]$ and $[a]'$ be the equivalence classes determined by a relative to $\sim_1$ and $\sim_2$, respectively. Well, then $[a]$ and $[a]'$, being members of $\mathcal{P}$, must equal the unique element $A_i \in \mathcal{P}$ containing a (by the above paragraph), i.e. $\{x \in A_i \,|\, x \sim_1 a\} = [a] = A_i = [a]' = \{x \in A \,|\, x \sim_2 a\}$, which implies $\sim_1 = \sim_2$. Thus, $\sim$ is the unique equivalence relation from which $\mathcal{P}$ is derived. $\blacksquare$

2 Equivalent Systems of Linear Equations

We now apply the idea of equivalence to systems of linear equations. There are two ways to do this, in terms of solution sets, and in terms of row-reduction to reduced echelon form. It turns out that the two equivalences define the same equivalence relation, and therefore the same equivalence classes, on the set of augmented matrices $(A|\mathbf{b})$ in $M_{m,n+1}(F)$ corresponding to systems $A\mathbf{x} = \mathbf{b}$, but we need some of the theory developed so far to demonstrate this. The upshot is, since row-reduction to reduced echelon form is an equivalence relation on $M_{m,n+1}(\mathbb{R})$, with

$$(A \mid \mathbf{b}) \sim \text{rref}(A \mid \mathbf{b})$$

and therefore

$$[(A \mid \mathbf{b})] = [\text{rref}(A \mid \mathbf{b})]$$

in $M_{m,n+1}(F)/\sim$, we'll have, by the culminating theorem of this section,

$$S_{(A \mid \mathbf{b})} = S_{\text{rref}(A \mid \mathbf{b})}$$

It is for this reason that row-reduction is so fundamental: row-reduction is *exactly* the method to *bring forth* the solution set in an easy-to-see form. For $S_{\text{rref}(A|\mathbf{b})}$ is easily determined by parametrizing the free variables and solving for the dependent variables, as we have done many times in class.

Moreover, row-reduction can be achieved by matrix multiplication by invertible *elementary matrices*, a particularly simple and transparent process. Let us explain.

Definition 2.1 Two systems $A\mathbf{x} = \mathbf{b}$ and $B\mathbf{x} = \mathbf{d}$, where $A, B \in M_{m,n}(F)$ and $\mathbf{b}, \mathbf{d} \in F^m$, are said to be **equivalent systems** if their solution sets are equal,

$$\begin{aligned} A\mathbf{x} = \mathbf{b} \text{ is equivalent to } B\mathbf{x} = \mathbf{d} \quad &\overset{\text{def}}{\iff} \quad (A \mid \mathbf{b}) \sim (B \mid \mathbf{d}) \\ &\overset{\text{def}}{\iff} \quad S_{(A \mid \mathbf{b})} = S_{(B \mid \mathbf{d})} \end{aligned}$$

■

Note that we identify the *system* $A\mathbf{x} = \mathbf{b}$ with its augmented matrix $(A|\mathbf{b})$ in $M_{m,n}(F)$ in each case. That way, everything may be easily rephrased in terms of matrices alone. This is, in any case, the idea behind row-reducing the augmented matrix instead of the system with the $\mathbf{x}$-variables in it.

Proposition 2.2 *The relation* $\sim$ *defined above, in terms of solution sets, is an* ***equivalence relation*** *on* $M_{m,n+1}(F)$*, which we'll denote call* **solution equivalence** *and denote*

$$\sim_S$$

Proof: (1) $A\mathbf{x} = \mathbf{b}$ is obviously equivalent itself in terms of solution sets, so reflexivity $(A|\mathbf{b}) \sim_S (A|\mathbf{b})$ is clear. (2) If $(A|\mathbf{b}) \sim_S (B|\mathbf{d})$, then the systems have the same solution set, $S_{(A \mid \mathbf{b})} = S_{(B \mid \mathbf{d})}$, and since equality is symmetric, $S_{(B \mid \mathbf{d})} = S_{(A \mid \mathbf{b})}$, we immediately conclude that $(B \mid \mathbf{d}) \sim_S (A \mid \mathbf{b})$. (3) If $(A \mid \mathbf{b}) \sim_S (B \mid \mathbf{d})$ and $(B \mid \mathbf{d}) \sim_S (C \mid \mathbf{f})$, then $S_{(A \mid \mathbf{b})} = S_{(B \mid \mathbf{d})}$ and $S_{(B \mid \mathbf{d})} = S_{(C \mid \mathbf{f})}$, so since equality is transitive, we conclude $S_{(A \mid \mathbf{b})} = S_{(C \mid \mathbf{f})}$, so that $(A \mid \mathbf{b}) \sim_S (C \mid \mathbf{f})$. ■

The main idea of the proof is that $\sim_S$ is defined in terms of $=$ on solution sets, and $=$ satisfies all three conditions. We simply verifify these.

Definition 2.3 Two systems $A\mathbf{x} = \mathbf{b}$ and $B\mathbf{x} = \mathbf{d}$ will be called **row equivalent** if there is a sequence $E_1, \dots, E_k \in \mathcal{E}_m$ of elementary matrices such that $(B \mid \mathbf{d}) = E_k \cdots E_2 E_1 (A \mid \mathbf{b})$,

$$\begin{aligned} A\mathbf{x} = \mathbf{b} \text{ is equivalent to } B\mathbf{x} = \mathbf{d} \quad &\stackrel{\text{def}}{\Longleftrightarrow} \quad (A \mid \mathbf{b}) \sim (B \mid \mathbf{d}) \\ &\stackrel{\text{def}}{\Longleftrightarrow} \quad (B \mid \mathbf{d}) = E_k \cdots E_2 E_1 (A \mid \mathbf{b}) \end{aligned}$$

■

Proposition 2.4 *The relation $\sim$ defined above, in terms of row operations, is an* ***equivalence relation*** *on $M_{m,n+1}(F)$, which we'll call* **row equivalence** *and denote*

$$\sim_R$$

Proof: (1) Since we include the identity matrix I_n in $\mathcal{E}_m$ as the trivial row operation which does nothing. Clearly $(A \mid \mathbf{b}) = I_n (A \mid \mathbf{b})$, which shows that $(A \mid \mathbf{b}) \sim_R (A \mid \mathbf{b})$, establishing reflexivity. (2) As to symmetry, this depends on the invertibility of elementary matrices, Lemma 1.11, combined with Theorem 4.6, which says that products of invertible matrices are invertible, and $(AB)^{-1} = B^{-1} A^{-1}$:

$$\begin{aligned} (A \mid \mathbf{b}) \sim_R (B \mid \mathbf{d}) \quad &\Longleftrightarrow \quad (B \mid \mathbf{d}) = E_k E_{k-1} \cdots E_2 E_1 (A \mid \mathbf{b}) \ \text{ for some } E_1, \dots, E_k \in \mathcal{E}_m \\ &\Longleftrightarrow \quad (A \mid \mathbf{b}) = (E_k E_{k-1} \cdots E_2 E_1)^{-1} (B \mid \mathbf{d}) \\ &\phantom{\Longleftrightarrow \quad (A \mid \mathbf{b})} = E_1^{-1} E_2^{-1} \cdots E_k^{-1} (B \mid \mathbf{d}) \\ &\Longleftrightarrow \quad (B \mid \mathbf{d}) \sim_R (A \mid \mathbf{b}) \end{aligned}$$

(3) If $(A \mid \mathbf{b}) \sim_R (B \mid \mathbf{d})$ and $(B \mid \mathbf{d}) \sim_R (C \mid \mathbf{f})$, then there are elementary matrices $E_1, \dots, E_k$ and $F_1, \dots, F_\ell \in \mathcal{E}_m$ such that

$$\begin{aligned} (B \mid \mathbf{d}) &= E_k \cdots E_1 (A \mid \mathbf{b}) \\ (C \mid \mathbf{f}) &= F_\ell \cdots F_1 (B \mid \mathbf{d}) \end{aligned}$$

and therefore, plugging the first into the second,

$$\begin{aligned} (C \mid \mathbf{f}) &= F_\ell \cdots F_1 (B \mid \mathbf{d}) \\ &= F_\ell \cdots F_1 E_k \cdots E_1 (A \mid \mathbf{b}) \end{aligned}$$

which shows that $(A \mid \mathbf{b}) \sim_R (C \mid \mathbf{f})$, so $\sim_R$ is transitive. ■

Proposition 2.5 *For all $B \in \mathrm{GL}(m, F)$ and all $(A \mid \mathbf{b}) \in M_{m,n+1}(F)$ whe have*

$$\boxed{S_{(BA \mid B\mathbf{b})} = S_{(A \mid \mathbf{b})}}$$

which, applied to $B = E_k \cdots E_1$, shows that

$$\boxed{\sim_R \implies \sim_S}$$

in $M_{m,n+1}(F)$.

Remark 2.6 It will remain to prove the other implication, $\sim_S \implies \sim_R$, which says that solution equivalence implies row equivalence, for then we will have $\sim_R \iff \sim_S$, and be satisfied to know that row-reduction is *precisely* the method by which to solve linear systems. ■

Proof: To show that $S_{(BA\,|\,B\mathbf{b})} = S_{(A\,|\mathbf{b})}$, we show that a solution vector $\mathbf{x}$ lies in the RHS iff it lies in the LHS:

$$\begin{aligned}
\mathbf{x} \in S_{(A\,|\mathbf{b})} \quad &\iff \quad A\mathbf{x} = \mathbf{b} \\
&\iff \quad A\mathbf{x} = \mathbf{b} = I_n\mathbf{b} = (B^{-1}B)\mathbf{b} = B^{-1}(B\mathbf{b}) \\
&\iff \quad BA\mathbf{x} = B\mathbf{b} \quad (\textit{multiply both sides above by } B) \\
&\iff \quad \mathbf{x} \in S_{(BA\,|\,B\mathbf{b})}
\end{aligned}$$

which completes the proof. Since $\mathcal{E}_m \subseteq \mathrm{GL}(m, F)$ by Lemma 1.11, we're done. ■

Corollary 2.7 *For any $(A\,|\,\mathbf{b}) \in M_{m,n+1}(F)$ representing a system $A\mathbf{x} = \mathbf{b}$, we have*

$$\boxed{S_{(A\,|\,\mathbf{b})} = S_{\mathrm{rref}(A\,|\,\mathbf{b})}}$$ ■

OK, so we got half of our target: ***row-equivalent systems must have the same solution sets, and in particular, the row-reduced echelon form of $A\mathbf{x} = \mathbf{b}$ must have the same solution set as the original system***. What about the reverse direction: if $A\mathbf{x} = \mathbf{b}$ and $B\mathbf{x} = \mathbf{d}$ have the same solution sets, that is, ***if $S_{(A\,|\,\mathbf{b})} = S_{(B\,|\mathbf{d})}$, then is it the case that $(A\,|\,\mathbf{b}) \sim_R (B\,|\,\mathbf{d})$, which in particular implies that $\mathrm{rref}(A\,|\,\mathbf{b}) = \mathrm{rref}(B\,|\,\mathbf{d})$?*** Let us begin with an observation about the nature of $S_{(A\,|\mathbf{b})}$. It will simplify matters somewhat.

Definition 2.8 A subspace U of $\mathbb{R}^n$, we recall, is a nonempty subset which must contain $\mathbf{0}$ and be closed under addition and scalar multiplication. ***If we add a nonzero vector $\mathbf{v}$ to U, or translate U by $\mathbf{v}$***, we get what's called an **affine subspace**, or more fully **an affine subspace modeled on the subspace** U,

$$\boxed{\mathbf{v} + U \stackrel{\text{def}}{=} \{\mathbf{v} + \mathbf{u} \,|\, \mathbf{u} \in U\}}$$ ■

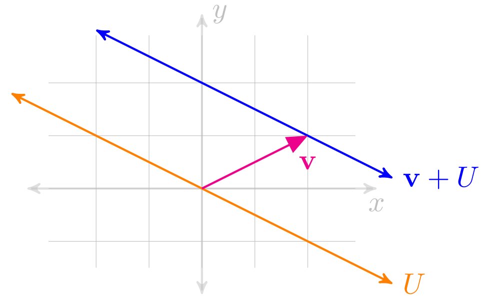
Example 2.9 Consider the case of the line $x + 2y = 0$, which is a subspace U in $\mathbb{R}^2$,

$$U = \mathrm{span}\begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

and let us translate this line by $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, which will give us the line $x + 2y = 4$, an affine subspace of $\mathbb{R}^2$:

$$\mathbf{v} + U = \{\mathbf{v} + \mathbf{u} \,|\, \mathbf{u} \in U\} = \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} + t\begin{pmatrix} -2 \\ 1 \end{pmatrix} \,\middle|\, t \in \mathbb{R} \right\}$$

Here's the picture:



■

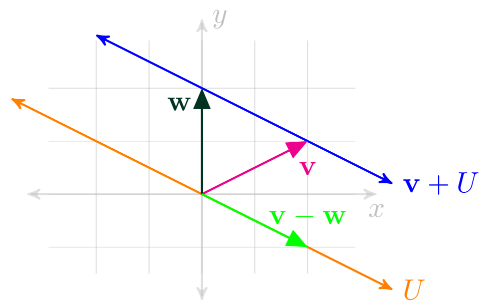
Remark 2.10 The vector $\mathbf{v}$ isn't unique in characterizing the affine subspace $\mathbf{v}+U$, for there are many others. If $\mathbf{w}$ is such another, so that

$$\mathbf{v}+U=\mathbf{w}+U$$

then

$$\mathbf{v}-\mathbf{w}\in U$$

and this is *the only way* this can happen. In the example above, one such $\mathbf{w}$ is $\begin{pmatrix}0\\2\end{pmatrix}$.



Matter of fact, ***one way to characterize affine sets is in terms of an equivalence relation on*** $\mathbb{R}^n$:

$$\boxed{\mathbf{v}\sim\mathbf{w}\quad\overset{\text{def}}{\iff}\quad\mathbf{v}-\mathbf{w}\in U}$$

Then, $\mathbf{v}+U$ is the equivalence class of $\mathbf{v}$ under this equivalence relation:

$$\boxed{[\mathbf{v}]=\{\mathbf{w}\in\mathbb{R}^n\mid\mathbf{v}\sim\mathbf{w}\}=\{\mathbf{w}\in\mathbb{R}^n\mid\mathbf{v}-\mathbf{w}\in U\}=\mathbf{v}+U}$$ ■

Exercise 2.11 Show that the relation defined in the remark above,

$$\mathbf{v}\sim\mathbf{w}\quad\overset{\text{def}}{\iff}\quad\mathbf{v}-\mathbf{w}\in U$$

is an equivalence relation, and that the equivalence classes $[\mathbf{v}]$ are precisely the affine subspaces $\mathbf{v}+U$,

$$[\mathbf{v}]=\mathbf{v}+U$$ ■

Proposition 2.12 *Let* $(A\,|\,\mathbf{b}) \in M_{m,n+1}(F)$ *be the augmented matrix representing the system* $A\mathbf{x} = \mathbf{b}$*. Then the solution set* $S_{(A\,|\,\mathbf{b})}$ *is an affine subspace modeled on the null space of* A*,*

$$\boxed{S_{(A\,|\,\mathbf{b})} = \mathbf{s} + N(A)}$$

where $\mathbf{s} \in S_{(A\,|\,\mathbf{b})}$ *is a particular solution. By the previous exercise, 2.11, the solution set is also an equivalence class of vectors in* $\mathbb{R}^n$*, represented by* $\mathbf{s}$*,*

$$\boxed{S_{(A\,|\,\mathbf{b})} = [\,\mathbf{s}\,] = \{\mathbf{v} \in \mathbb{R}^n \,|\, \mathbf{s} - \mathbf{v} \in N(A)\}}$$

since $\mathbf{s} \sim \mathbf{v} \overset{\text{def}}{\iff} \mathbf{s} - \mathbf{v} \in N(A)$*.*

This, incidentally, is why we worried only about homogeneous systems at first, because those treat $N(A)$. After we come understand $N(A)$, we translate it gladly by $\mathbf{s}$ to get all other nonhomogeneous solutions $\mathbf{s} + N(A)$.

Proof: To show the equality of sets $S_{(A\,|\,\mathbf{b})} = \mathbf{s} + N(A)$ we show a double inclusion, that is we show $S_{(A\,|\,\mathbf{b})}$ is contained in $\mathbf{s} + N(A)$, and vice-versa. To show $S_{(A\,|\,\mathbf{b})}$ is contained in $\mathbf{s} + N(A)$, we take a solution vector $\mathbf{x}$ in $S_{(A\,|\,\mathbf{b})}$ and proceed somehow to demonstrate that it is in $\mathbf{s} + N(A)$:

Step 1: $S_{(A\,|\,\mathbf{b})} \subseteq \mathbf{s} + N(A)$. Suppose $\mathbf{x} \in S_{(A\,|\,\mathbf{b})}$. Then $A\mathbf{x} = \mathbf{b}$, so since $\mathbf{s} \in S_{(A\,|\,\mathbf{b})}$, too,

$$A(\mathbf{x} - \mathbf{s}) = A\mathbf{x} - A\mathbf{s} = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

we see that $\mathbf{x} - \mathbf{s} \in N(A)$ is a null vector. Since this is true for all $\mathbf{x} \in S_{(A\,|\,\mathbf{b})}$, we see that

$$S_{(A\,|\,\mathbf{b})} - \mathbf{s} \subseteq N(A)$$

and therefore, by adding $\mathbf{s}$ to each set,

$$S_{(A\,|\,\mathbf{b})} \subseteq \mathbf{s} + N(A)$$

Step 2: $S_{(A\,|\,\mathbf{b})} \supseteq \mathbf{s} + N(A)$. If $\mathbf{s} + \mathbf{x} \in \mathbf{s} + N(A)$, then by definition $\mathbf{s} \in S_{(A\,|\,\mathbf{b})}$ and $\mathbf{x} \in N(A)$, so

$$A(\mathbf{s} + \mathbf{x}) = A\mathbf{s} + A\mathbf{x} = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

is also a solution, i.e. $\mathbf{s} + \mathbf{x} \in S_{(A\,|\,\mathbf{b})}$. ■

We now take our first strong step towards showing that $\sim_S$ implies $\sim_R$, keeping in mind the proposition we just laid down.

Proposition 2.13 $(A\,|\,\mathbf{b}) \sim_S (B\,|\,\mathbf{d}) \implies N(A) = N(B)$*, i.e.*

$$\boxed{S_{(A\,|\,\mathbf{b})} = S_{(B\,|\,\mathbf{d})} \implies N(A) = N(B)}$$

Proof: Suppose $(A\,|\,\mathbf{b}) \sim_S (B\,|\,\mathbf{d})$, or $S_{(A\,|\,\mathbf{b})} = S_{(B\,|\,\mathbf{d})}$. By the previous proposition this means

$$\mathbf{s} + N(A) = \mathbf{t} + N(B)$$

for some particular solutions, $\mathbf{s}$ of $A\mathbf{x} = \mathbf{b}$ and $\mathbf{t}$ of $B\mathbf{x} = \mathbf{d}$. But then

$$\mathbf{s} = \mathbf{s} + \mathbf{0} \in \mathbf{s} + N(A) = \mathbf{t} + N(B) = S_{(B\,|\,\mathbf{d})}$$

which means $\mathbf{s}$ is a solution of $B\mathbf{x} = \mathbf{d}$, too. Similarly, $\mathbf{t}$ is also a solution of $A\mathbf{x} = \mathbf{b}$. Because of this,

$$\mathbf{s}, \mathbf{t} \in \mathbf{s} + N(A) = S_{(A\,|\,\mathbf{b})} = S_{(B\,|\,\mathbf{d})} = \mathbf{t} + N(B)$$

and therefore, by Exercise 2.11,

$$\mathbf{t} - \mathbf{s} \in N(A) \cap N(B)$$

From this it finally follows that

$$N(A) = -\mathbf{s} + \big(\mathbf{s} + N(A)\big) = -\mathbf{s} + \big(\mathbf{t} + N(B)\big) = \big(\mathbf{t} - \mathbf{s}\big) + N(B) = N(B) \qquad \blacksquare$$

Proposition 2.14 $(A|\mathbf{b}) \sim_R (B|\mathbf{d}) \implies \text{Row}(A) = \text{Row}(B)$, *i.e.* $\text{span}(\vec{A}_1, \ldots, \vec{A}_m) = \text{span}(\vec{B}_1, \ldots, \vec{B}_m)$.

Proof: Consider each type of elementary matrix/row operation, one at a time:

Type I: Let $E \in \mathcal{E}_m$ be a type I elementary matrix, swapping rows i and j. Then obviously

$$\text{span}(\vec{A}_1, \ldots, \vec{A}_i, \ldots, \vec{A}_j, \ldots, \vec{A}_m) = \text{span}(\vec{A}_1, \ldots, \vec{A}_j, \ldots, \vec{A}_i \ldots, \vec{A}_m)$$

since addition is commutative. Calling the RHS $B = EA$, we see that $\text{span}(\vec{A}_1, \ldots, \vec{A}_m) = \text{span}(\vec{B}_1, \ldots, \vec{B}_m)$.

Type II: Let $E \in \mathcal{E}_m$ be a type II elementary matrix, scaling row i by $c \neq 0$. Then, in terms of rows,

$$B \stackrel{\text{def}}{=} EA = \begin{pmatrix} \vec{A}_1 \\ \vdots \\ c\vec{A}_i \\ \vdots \\ \vec{A}_m \end{pmatrix}$$

so $\text{span}(\vec{B}_1, \ldots, \vec{B}_m) = \text{span}(\vec{A}_1, \ldots, c\vec{A}_i, \ldots, \vec{A}_m) = \text{span}(\vec{A}_1, \ldots, \vec{A}_m)$, where the last equality follows from the fact that

$$\begin{aligned} \mathbf{x} \in \text{span}(\vec{A}_1, \ldots, c\vec{A}_i, \ldots, \vec{A}_m) \quad &\iff \quad \mathbf{x} = \sum_{j \neq i}^{m} a_j \vec{A}_j + a_i(c\vec{A}_i) \\ &\iff \quad \mathbf{x} = \sum_{j \neq i}^{m} a_j \vec{A}_j + (a_i c)\vec{A}_i \\ &\iff \quad \mathbf{x} \in \text{span}(\vec{A}_1, \ldots, \vec{A}_m) \end{aligned}$$

Type III: Let $E \in \mathcal{E}_m$ be a type III elementary matrix, adding $c\vec{A}_i$ to $\vec{A}_j$,

$$B \stackrel{\text{def}}{=} EA = \begin{pmatrix} \vec{A}_1 \\ \vdots \\ \vec{A}_i \\ \vdots \\ c\vec{A}_i + \vec{A}_j \\ \vdots \\ \vec{A}_m \end{pmatrix}$$

To show that $\text{span}(\vec{A}_1, \ldots, \vec{A}_m) = \text{span}(\vec{B}_1, \ldots, \vec{B}_m)$, we use a double inclusion, starting with the LHS: let $\mathbf{x} \in \text{span}(\vec{A}_1, \ldots, \vec{A}_m)$. Then,

$$\begin{aligned} \mathbf{x} &= \sum_{k=1}^{m} a_k \vec{A}_k \\ &= \sum_{k=1}^{m} a_k \vec{A}_k - a_j c\vec{A}_i + a_j c\vec{A}_i \\ &= \sum_{k \neq i,j} a_k \vec{A}_k + (a_i - a_j c)\vec{A}_i + a_j(c\vec{A}_i + \vec{A}_j) \\ &= \sum_{k \neq i,j} a_j \vec{B}_k + (a_i - a_j c)\vec{B}_i + a_j \vec{B}_j \end{aligned}$$

so $\mathbf{x} \in \text{span}(\vec{B}_1, \ldots, \vec{B}_m)$. Conversely, if $\mathbf{x} \in \text{span}(\vec{B}_1, \ldots, \vec{B}_m)$, then

$$\begin{aligned} \mathbf{x} &= \sum_{k=1}^{m} a_k \vec{B}_k \\ &= \sum_{k \neq i,j}^{m} a_k \vec{B}_k + a_i \vec{B}_i + a_j \vec{B}_j \\ &= \sum_{k \neq i,j} a_k \vec{A}_k + a_i \vec{A}_i + a_j(c\vec{A}_i + \vec{A}_j) \\ &= \sum_{k \neq i,j} a_k \vec{A}_k + (a_i + a_j c)\vec{A}_i + a_j \vec{A}_j \end{aligned}$$

which lies in $\text{span}(\vec{A}_1, \ldots, \vec{A}_m)$. ■

Remark 2.15 In Exercise 1.18 of the 'Bases, Coordinates and Representations' notes you proved that the row-space of any matrix A is orthogonal to its null space, and that therefore $\text{Row}(A) \cap N(A) = \{\mathbf{0}\}$. We wish now to say that $N(A)$ *consists entirely* of such orthogonal vectors, and that in fact $\text{Row}(A)$ is this set of orthogonal vectors. ■

Definition 2.16 For any subspace U of $\mathbb{R}^n$, we define its **orthogonal complement** to be all those vectors in $\mathbb{R}^n$ which are orthogonal to U, and denote it

$$\boxed{U^\perp \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{u} \cdot \mathbf{x} = 0 \;\text{ for } \underline{\text{all}} \;\; \mathbf{u} \in U\}}$$ ■

Since 'dotting-with-$\mathbf{u}$' is a linear transformation $T_\mathbf{u} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$, any orthogonal vector $\mathbf{x}$ to $\mathbf{u}$ is a null vector of $T_\mathbf{u}$. The null space of $T_\mathbf{u}$ is a subspace, however, and any intersection of subspaces is a subspace. Therefore,

$$U^\perp = \bigcap_{\mathbf{u} \in U} N(T_\mathbf{u})$$

is a subspace of $\mathbb{R}^n$. Moreover, it is a <u>*complement*</u> of U, so that

$$\boxed{\mathbb{R}^n = U \oplus U^\perp}$$

which explanis the name 'orthogonal complement.' To see this, apply the Rank-Nullity Theorem to the **projection** π_U **onto** U,

$$\boxed{\pi_U(\mathbf{x}) \stackrel{\text{def}}{=} \begin{cases} \mathbf{x}, & \text{if } \mathbf{x} \in U \\ \mathbf{0}, & \text{if } \mathbf{x} \in U^\perp \end{cases}}$$

When we study the dot product more deeply, we will be able to show that $N(\pi_U) = U^\perp$. As a result, if we choose a basis β_1 for $U = R(\pi_U)$ and another β_2 for $N(\pi_U) = U^\perp$, we deduce from the Rank-Nullity Theorem that

$$\begin{aligned} \mathbb{R}^n &= R(\pi_U) \oplus N(\pi_U) \\ &= U \oplus U^\perp \end{aligned}$$

Indeed, we have that $\beta \stackrel{\text{def}}{=} \beta_1 \cup \beta_2$ is a basis for $\mathbb{R}^n$. What is the significance of this? **Answer:** We apply the above to $U = N(A)$, and conclude that

Proposition 2.17 *For any matrix $A \in M_{m,n}(F)$, the null space is the orthogonal complement of the row space*

$$\boxed{\text{Row}(A) = N(A)^\perp}$$

and vice-versa

$$\boxed{\text{Row}(A)^\perp = N(A)}$$

By the Rank-Nullity Theorem we conclude that

$$\boxed{\mathbb{R}^n = N(A) \oplus \text{Row}(A)}$$

Proof: First, $\mathbf{x} \in N(A)$ iff $A\mathbf{x} = \mathbf{0}$ iff $\mathbf{x} \in \text{Row}(A)^\perp$, which shows the second equality. The first equality follows from the fact that $U^{\perp\perp} = U$, by an easy exercise. ■

Combining Propositions 2.13 and 2.17 we come to the following conclusion:

$$\boxed{\begin{aligned}(A\,|\,\mathbf{b}) \sim_S (B\,|\,\mathbf{d}) &\overset{\text{def}}{\Longrightarrow} S_{(A\,|\,\mathbf{b})} = S_{(B\,|\,\mathbf{d})}\\ &\overset{\text{Prop 2.13}}{\Longrightarrow} N(A) = N(B)\\ &\overset{\text{Prop 2.17}}{\Longrightarrow} \text{Row}(A) = N(A)^\perp = N(B)^\perp = \text{Row}(B)\end{aligned}}$$

Proposition 2.18 *Let $A, B \in M_{m,n}(F)$. Then*

$$\boxed{N(A) = N(B) \quad\Longrightarrow\quad \text{rref}(A) = \text{rref}(B)}$$

Proof: Proposition 2.14 says that because $(A\,|\,\mathbf{0}) \sim_R \text{rref}(A\,|\,\mathbf{0})$ we must have

$$\boxed{\text{Row}(\text{rref}(A)) = \text{Row}(A) = N(A)^\perp \overset{\text{hyp}}{=} N(B)^\perp = \text{Row}(B) = \text{Row}(\text{rref}(B))}$$

Let $P = E_k \cdots E_1$ such that $PA = \text{rref}(A)$ and let $Q = F_\ell \cdots F_1$ such that $QB = \text{rref}(B)$. The rows and columns of $\text{rref}(A)$ and $\text{rref}(B)$ are uniquely determined by P and Q, respectively, and it remains only to check that these must be the same. But we observe that by Corollary 2.7,

$$\boxed{N(A) = S_{(A\,|\,\mathbf{0})} = S_{\text{rref}(A\,|\,\mathbf{0})} = N(\text{rref}(A))}$$

and similarly $N(B) = N(\text{rref}(B))$, from which we also get

$$\boxed{N(\text{rref}(A)) = N(\text{rref}(B))}$$

But $\text{rref}(A)$ is uniquely determined by its pivot columns, which are the standard basis vectors $\mathbf{e}_1$ through $\mathbf{e}_k$, $k = \text{rank}\,A$, and which determine the non-pivot columns in terms of specific linear combinations of the pivots. These are unaltered by P-multiplication. And these *entirely* determine the behavior of A, giving it a basis for each of $N(A)$ and $\text{Row}(A)$. Should $\text{rref}(A) \neq \text{rref}(B)$, it must be because some column, and therefore some row, were different. But this difference would be detected immediately were it in the location of a pivot column, for then the row spaces would *have* to be different. But if it were to occur in a nonpivot column, the parametrization of the free variables would produce a difference in the null spaces. Either way, we'd run into a contradiction with the above equalities. ■

Corollary 2.19 *The row-reduced echelon form of A is unique.* ■

Corollary 2.20 *$\sim_S$ implies $\sim_R$ on $M_{m,n}(\mathbb{R})$, and therefore we conclude that*

$$\boxed{(A\,|\,\mathbf{b}) \sim_S (B\,|\,\mathbf{d}) \quad\Longleftrightarrow\quad (A\,|\,\mathbf{b}) \sim_R (B\,|\,\mathbf{d})}$$

and therefore

$$\boxed{S_{(A\,|\,\mathbf{b})} = S_{(B\,|\,\mathbf{d})} \quad\Longleftrightarrow\quad \text{rref}(A\,|\,\mathbf{b}) = \text{rref}(B\,|\,\mathbf{d})}$$ ■