Row reduction and Gaussian elimination

Since the textbook, *Linear Algebra Done Right* [sic], gives absolutely no specific information about the fundamentally important notion of Gaussian elimination but cites it as a means for doing computations with examples, it is necessary to post some more specific information about this concept. The excerpt on the following pages is taken from Petersen, *Linear Algebra*. Unfortunately, Petersen does not mention the term "Gaussian elimination" so we need to explain what this means in the context of his book: The goal of the procedure is to find all solutions for a system of m linear equations expressible in the form $AX = 0$, where A is an $m \times n$ matrix of coefficients and X is an $n \times 1$ matrix of unknowns. This is equivalent to finding the kernel of the linear transformation sending X to AX, and Gaussian elimination is another name for (1) putting the matrix A into row reduced echelon form as in Petersen, (2) reading off a basis for the kernel as in the last three and first five lines of pages 68 and 69. In particular, if we define the pivot coordinates to be those which correspond to the leading entries of the nonzero rows (the columns $j(1)$ etc. in the passage from the book), then the solutions are given by allowing the non – pivot coordinates to vary independently and using the equations from the row reduced echelon form to compute the pivot coordinates. For example, if we have the 3×5 row reduced echelon matrix with entries

$$\begin{matrix} 1 & 0 & 0 & 2 & 3 \\ 0 & 0 & 1 & 4 & 5 \end{matrix}$$

then the pivot columns are columns 1 and 3, the space of solutions is 3 – dimensional, and a basis for the space of solutions is given by the following three vectors:

$$(0, 1, 0, 0, 0)$$

$$(-2, 0, -4, 1, 0)$$

$$(-3, 0, -5, 0, 1)$$

(9) If $x_1, ..., x_k$ are linearly dependent, then $L(x_1), ..., L(x_k)$ are linearly dependent.
(10) If $L(x_1), ..., L(x_k)$ are linearly independent, then $x_1, ..., x_k$ are linearly independent.
(11) Let $A \in \mathrm{Mat}_{n\times m}(\mathbb{F})$ and assume that $y_1, ..., y_m \in V$

$$\begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} A$$

where $x_1, ..., x_n$ form a basis for V.
  (a) Show that $y_1, ..., y_m$ span V if and only if A has rank n. Conclude that $m \geq n$.
  (b) Show that $y_1, ..., y_m$ are linearly independent if and only if $\ker(A) = \{0\}$. Conclude that $m \leq n$.
  (c) Show that $y_1, ..., y_m$ form a basis for V if and only if A is invertible. Conclude that $m = n$.

13. Row Reduction

In this section we give a brief and rigorous outline of the standard procedures involved in solving systems of linear equations. The goal in the context of what we have already learned is to find a way of computing the image and kernel of a linear map that is represented by a matrix. Along the way we shall reprove that the dimension is well-defined as well as the dimension formula for linear maps.

The usual way of writing n equations with m variables is

$$\begin{array}{ccc} a_{11}x_1 + \cdots + a_{1m}x_m & = & b_1 \\ \vdots & \vdots & \vdots \\ a_{n1}x_1 + \cdots + a_{nm}x_m & = & b_n \end{array}$$

where the variables are $x_1, ..., x_m$. The goal is to understand for which choices of constants a_{ij} and b_i such systems can be solved and then list all the solutions. To conform to our already specified notation we change the system so that it looks like

$$\begin{array}{ccc} \alpha_{11}\xi_1 + \cdots + \alpha_{1m}\xi_m & = & \beta_1 \\ \vdots & \vdots & \vdots \\ \alpha_{n1}\xi_1 + \cdots + \alpha_{nm}\xi_m & = & \beta_n \end{array}$$

In matrix form this becomes

$$\begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nm} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$$

and can be abbreviated to

$$Ax = b.$$

As such we can easily use the more abstract language of linear algebra to address some general points.

Proposition 3. *Let $L : V \to W$ be a linear map.*
(1) *$L(x) = b$ can be solved if and only if $b \in \mathrm{im}(L)$.*
(2) *If $L(x_0) = b$ and $x \in \ker(L)$, then $L(x + x_0) = b$.*
(3) *If $L(x_0) = b$ and $L(x_1) = b$, then $x_0 - x_1 \in \ker(L)$.*

Therefore, we can find all solutions to $L(x) = b$ provided we can find the kernel $\ker(L)$ and just one solution x_0. Note that the kernel consists of the solutions to what we call the *homogeneous system*: $L(x) = 0$.

With this behind us we are now ready to address the issue of how to make the necessary calculations that allow us to find a solution to

$$\begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nm} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$$

The usual method is through *elementary row operations*. To keep things more conceptual think of the actual linear equations

$$\begin{array}{ccc} \alpha_{11}\xi_1 + \cdots + \alpha_{1m}\xi_m & = & \beta_1 \\ \vdots & \vdots & \vdots \\ \alpha_{n1}\xi_1 + \cdots + \alpha_{nm}\xi_m & = & \beta_n \end{array}$$

and observe that we can perform the following three operations without changing the solutions to the equations:

(1) Interchanging equations (or rows).
(2) Adding a multiple of an equation (or row) to a different equation (or row).
(3) Multiplying an equation (or row) by a nonzero number.

Using these operations one can put the system in *row echelon form*. This is most easily done by considering the *augmented matrix*, where the variables have disappeared

$$\left[\begin{array}{ccc|c} \alpha_{11} & \cdots & \alpha_{1m} & \beta_1 \\ \vdots & \ddots & \vdots & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nm} & \beta_n \end{array}\right]$$

and then performing the above operations, now on rows, until it takes the special form where

(1) The first nonzero entry in each row is normalized to be 1. This is also called the *leading* 1 for the row.
(2) The leading 1s appear in echelon form, i.e., as we move down along the rows the leading 1s will appear farther to the right.

The method by which we put a matrix into row echelon form is called *Gauss elimination*. Having put the system into this simple form one can then solve it by starting from the last row or equation.

When doing the process on A itself we denote the resulting row echelon matrix by A_{ref}. There are many ways of doing row reductions so as to come up with a row echelon form for A and it is quite likely that one ends up with different echelon forms. To see why consider

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

This matrix is clearly in row echelon form. However we can subtract the second row from the first row to obtain a new matrix which is still in row echelon form:

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

It is now possible to use the last row to arrive at

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The important information about A_{ref} is the placement of the leading 1 in each row and this placement will always be the same for any row echelon form. To get a unique row echelon form we need to reduce the matrix using *Gauss-Jordan elimination*. This process is what we just performed on the above matrix A. The idea is to first arrive at some row echelon form A_{ref} and then starting with the second row eliminate all entries above the leading 1, this is then repeated with row three, etc. In this way we end up with a matrix that is still in row echelon form, but also has the property that all entries below and above the leading 1 in each row are zero. We say that such a matrix is in *reduced row echelon form*. If we start with a matrix A, then the resulting reduced row echelon form is denoted A_{rref}. For example, if we have

$$A_{\text{ref}} = \begin{bmatrix} 0 & 1 & 4 & 1 & 0 & 3 & -1 \\ 0 & 0 & 0 & 1 & -2 & 5 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

then we can reduce further to get a new reduced row echelon form

$$A_{\text{rref}} = \begin{bmatrix} 0 & 1 & 4 & 0 & 2 & -2 & 0 \\ 0 & 0 & 0 & 1 & -2 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The row echelon form and reduced row echelon form of a matrix can more abstractly be characterized as follows. Suppose that we have an $n \times m$ matrix $A = \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}$, where $x_1, ..., x_m \in \mathbb{F}^n$ correspond to the columns of A. Let $e_1, ..., e_n \in \mathbb{F}^n$ be the canonical basis. The matrix is in row echelon form if we can find $1 \leq j_1 < \cdots < j_k \leq m$, where $k \leq n$, such that

$$x_{j_s} = e_s + \sum_{i<s} \alpha_{ij_s} e_i$$

for $s = 1, ..., k$. For all other indices j we have

$$\begin{aligned} x_j &= 0, \text{ if } j < j_1, \\ x_j &\in \operatorname{span}\{e_1, ..., e_s\}, \text{ if } j_s < j < j_{s+1}, \\ x_j &\in \operatorname{span}\{e_1, ..., e_k\}, \text{ if } j_k < j. \end{aligned}$$

Moreover, the matrix is in reduced row echelon form if in addition we assume that

$$x_{j_s} = e_s.$$

Below we shall prove that the reduced row echelon form of a matrix is unique, but before doing so it is convenient to reinterpret the row operations as matrix multiplication.

Let $A \in \mathrm{Mat}_{n\times m}(\mathbb{F})$ be the matrix we wish to row reduce. The row operations we have described can be accomplished by multiplying A by certain invertible $n \times n$ matrices on the left. These matrices are called *elementary matrices*. The define these matrices we use the standard basis matrices E_{kl} where the kl entry is 1 while all other entries are 0. The matrix product $E_{kl}A$ is a matrix whose k^{th} row is the l^{th} row of A and all other rows vanish.

(1) Interchanging rows k and l: This can be accomplished by the matrix multiplication $I_{kl}A$, where

$$\begin{aligned} I_{kl} &= E_{kl} + E_{lk} + \sum_{i \neq k,l} E_{ii} \\ &= E_{kl} + E_{lk} + 1_{\mathbb{F}^n} - E_{kk} - E_{ll} \end{aligned}$$

or in other words the ij entries α_{ij} in I_{kl} satisfy $\alpha_{kl} = \alpha_{lk} = 1$, $\alpha_{ii} = 1$ if $i \neq k, l$, and $\alpha_{ij} = 0$ otherwise. Note that $I_{kl} = I_{lk}$ and $I_{kl}I_{lk} = 1_{\mathbb{F}^n}$. Thus I_{kl} is invertible.

(2) Multiplying row l by $\alpha \in \mathbb{F}$ and adding it to row $k \neq l$. This can be accomplished via $R_{kl}(\alpha)A$, where

$$R_{kl}(\alpha) = 1_{\mathbb{F}^n} + \alpha E_{kl}$$

or in other words the ij entries α_{ij} in $R_{kl}(\alpha)$ look like $\alpha_{ii} = 1$, $\alpha_{kl} = \alpha$, and $\alpha_{ij} = 0$ otherwise. This time we note that $R_{kl}(\alpha)R_{kl}(-\alpha) = 1_{\mathbb{F}^n}$.

(3) Multiplying row k by $\alpha \in \mathbb{F} - \{0\}$. This can be accomplished by $M_k(\alpha)A$, where

$$\begin{aligned} M_k(\alpha) &= \alpha E_{kk} + \sum_{i \neq k} E_{ii} \\ &= 1_{\mathbb{F}^n} + (\alpha - 1)E_{kk} \end{aligned}$$

or in other words the ij entries α_{ij} of $M_k(\alpha)$ are $\alpha_{kk} = \alpha$, $\alpha_{ii} = 1$ if $i \neq k$, and $\alpha_{ij} = 0$ otherwise. Clearly $M_k(\alpha)M_k(\alpha^{-1}) = 1_{\mathbb{F}^n}$.

Performing row reductions on A is now the same as doing a matrix multiplication PA, where $P \in \mathrm{Mat}_{n\times n}(\mathbb{F})$ is a product of the elementary matrices. Note that such P are invertible and that P^{-1} is also a product of elementary matrices. The elementary 2×2 matrices look like.

$$\begin{aligned} I_{12} &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \\ R_{12}(\alpha) &= \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix}, \\ R_{21}(\alpha) &= \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix}, \\ M_1(\alpha) &= \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix}, \\ M_2(\alpha) &= \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix}. \end{aligned}$$

If we multiply these matrices onto A from the left we obtain the desired operations:

$$I_{12}A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} \alpha_{21} & \alpha_{22} \\ \alpha_{11} & \alpha_{12} \end{bmatrix}$$

$$R_{12}(\alpha) A = \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} \alpha_{11} + \alpha\alpha_{21} & \alpha_{12} + \alpha\alpha_{22} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}$$

$$R_{21}(\alpha) A = \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha\alpha_{11} + \alpha_{21} & \alpha\alpha_{12} + \alpha_{22} \end{bmatrix}$$

$$M_1(\alpha) A = \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} \alpha\alpha_{11} & \alpha\alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}$$

$$M_2(\alpha) A = \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha\alpha_{21} & \alpha\alpha_{22} \end{bmatrix}$$

We can now move on to the important result mentioned above.

THEOREM 8. (Uniqueness of Reduced Row Echelon Form) *The reduced row echelon form of an $n \times m$ matrix is unique.*

PROOF. Let $A \in \mathrm{Mat}_{n\times m}(\mathbb{F})$ and assume that we have two reduced row echelon forms

$$\begin{aligned} PA &= \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}, \\ QA &= \begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix}, \end{aligned}$$

where $P, Q \in \mathrm{Mat}_{n\times n}(\mathbb{F})$ are invertible. In particular, we have that

$$R \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix} = \begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix}$$

where $R \in \mathrm{Mat}_{n\times n}(\mathbb{F})$ is invertible. We shall show that $x_i = y_i$, $i = 1, ..., m$ by induction on n.

First observe that if $A = 0$, then there is nothing to prove. If $A \neq 0$, then both of the reduced row echelon forms have to be nontrivial. Then we have that

$$\begin{aligned} x_{i_1} &= e_1, \\ x_i &= 0 \text{ for } i < i_1 \end{aligned}$$

and

$$\begin{aligned} y_{j_1} &= e_1, \\ y_i &= 0 \text{ for } i < j_1. \end{aligned}$$

The relationship $Rx_i = y_i$ shows that $y_i = 0$ if $x_i = 0$. Thus $j_1 \geq i_1$. Similarly the relationship $y_i = R^{-1}x_i$ shows that $x_i = 0$ if $y_i = 0$. Hence also $j_1 \leq i_1$. Thus $i_1 = j_1$ and $x_{i_1} = e_1 = y_{j_1}$. This implies that $Re_1 = e_1$ and $R^{-1}e_1 = e_1$. In other words

$$R = \begin{bmatrix} 1 & 0 \\ 0 & R' \end{bmatrix}$$

where $R' \in \mathrm{Mat}_{(n-1)\times(n-1)}(\mathbb{F})$ is invertible. In the special case where $n = 1$, we are finished as we have shown that $R = [1]$ in that case. This anchors our induction. We can now assume that the induction hypothesis is that all $(n-1) \times m$ matrices have unique reduced row echelon forms.

If we define $x_i', y_i' \in \mathbb{F}^{n-1}$ as the last $n-1$ entries in x_i and y_i, i.e.,

$$\begin{aligned} x_i &= \begin{bmatrix} \xi_{1i} \\ x_i' \end{bmatrix}, \\ y_i &= \begin{bmatrix} \upsilon_{1i} \\ y_i' \end{bmatrix}, \end{aligned}$$

then we see that $\begin{bmatrix} x_1' & \cdots & x_m' \end{bmatrix}$ and $\begin{bmatrix} y_1' & \cdots & y_m' \end{bmatrix}$ are still in reduced row echelon form. Moreover, the relationship

$$\begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix} = R \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}$$

now implies that

$$\begin{aligned} \begin{bmatrix} \upsilon_{11} & \cdots & \upsilon_{1m} \\ y_1' & \cdots & y_m' \end{bmatrix} &= \begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix} \\ &= R \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & R' \end{bmatrix} \begin{bmatrix} \xi_{11} & \cdots & \xi_{1m} \\ x_1' & \cdots & x_m' \end{bmatrix} \\ &= \begin{bmatrix} \xi_{11} & \cdots & \xi_{1m} \\ R'x_1' & \cdots & R'x_m' \end{bmatrix} \end{aligned}$$

Thus

$$R' \begin{bmatrix} x_1' & \cdots & x_m' \end{bmatrix} = \begin{bmatrix} y_1' & \cdots & y_m' \end{bmatrix}.$$

The induction hypothesis now implies that $x_i' = y_i'$. This combined with

$$\begin{aligned} \begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix} &= \begin{bmatrix} \upsilon_{11} & \cdots & \upsilon_{1m} \\ y_1' & \cdots & y_m' \end{bmatrix} \\ &= \begin{bmatrix} \xi_{11} & \cdots & \xi_{1m} \\ R'x_1' & \cdots & R'x_m' \end{bmatrix} \\ &= \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix} \end{aligned}$$

shows that $x_i = y_i$ for all $i = 1, ..., m$. □

We are now ready to explain how the reduced row echelon form can be used to identify the kernel and image of a matrix. Along the way we shall reprove some of our earlier results. Suppose that $A \in \mathrm{Mat}_{n\times m}(\mathbb{F})$ and

$$\begin{aligned} PA &= A_{\text{rref}} \\ &= \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}, \end{aligned}$$

where we can find $1 \le j_1 < \cdots < j_k \le m$, such that

$$\begin{aligned} x_{j_s} &= e_s \text{ for } i = 1, ..., k \\ x_j &= 0, \text{ if } j < j_1, \\ x_j &\in \operatorname{span}\{e_1, ..., e_s\}, \text{ if } j_s < j < j_{s+1}, \\ x_j &\in \operatorname{span}\{e_1, ..., e_k\}, \text{ if } j_k < j. \end{aligned}$$

Finally let $i_1 < \cdots < i_{m-k}$ be the indices complementary to $j_1, .., j_k$, i.e.,

$$\{1, ..., m\} = \{j_1, .., j_k\} \cup \{i_1, ..., i_{m-k}\}.$$

We are first going to study the kernel of A. Since P is invertible we see that $Ax = 0$ if and only if $A_{\text{rref}}x = 0$. Thus we need only study the equation $A_{\text{rref}}x = 0$. If we let $x = (\xi_1, ..., \xi_m)$, then the nature of the equations $A_{\text{rref}}x = 0$ will tell us that

$(\xi_1, ..., \xi_m)$ are uniquely determined by $\xi_{i_1}, ..., \xi_{i_{m-k}}$. To see why this is we note that if we have $A_{\text{rref}} = [\alpha_{ij}]$, then the reduced row echelon form tells us that

$$\begin{aligned} \xi_{j_1} + \alpha_{1i_1}\xi_{i_1} + \cdots + \alpha_{1i_{m-k}}\xi_{i_{m-k}} &= 0, \\ &\vdots \\ \xi_{j_k} + \alpha_{ki_1}\xi_{i_1} + \cdots + \alpha_{ki_{m-k}}\xi_{i_{m-k}} &= 0, \end{aligned}$$

Thus $\xi_{j_1}, ..., \xi_{j_k}$ have explicit formulas in terms of $\xi_{i_1}, ..., \xi_{i_{m-k}}$. We actually get a bit more information: If we take $(\alpha_1, ..., \alpha_{m-k}) \in \mathbb{F}^{m-k}$ and construct the unique solution $x = (\xi_1, ..., \xi_m)$ such that $\xi_{i_1} = \alpha_1$, ..., $\xi_{i_{m-k}} = \alpha_{m-k}$ then we have actually constructed a map

$$\begin{aligned} \mathbb{F}^{m-k} &\to \ker(A_{\text{rref}}) \\ (\alpha_1, ..., \alpha_{m-k}) &\to (\xi_1, ..., \xi_m). \end{aligned}$$

We have just seen that this map is onto. The construction also gives us explicit formulas for $\xi_{j_1}, ..., \xi_{j_k}$ that are linear in $\xi_{i_1} = \alpha_1$, ..., $\xi_{i_{m-k}} = \alpha_{m-k}$. Thus the map is linear. Finally if $(\xi_1, ..., \xi_m) = 0$, then we clearly also have $(\alpha_1, ..., \alpha_{m-k}) = 0$, so the map is one-to-one. All in all it is a linear isomorphism.

This leads us to the following result.

THEOREM 9. (Uniqueness of Dimension) *Let* $A \in \text{Mat}_{n\times m}(\mathbb{F})$, *if* $n < m$, *then* $\ker(A) \neq \{0\}$. *Consequently* $\mathbb{F}^n$ *and* $\mathbb{F}^m$ *are not isomorphic.*

PROOF. Using the above notation we have $k \leq n < m$. Thus $m - k > 0$. From what we just saw this implies $\ker(A) = \ker(A_{\text{rref}}) \neq \{0\}$. In particular it is not possible for A to be invertible. This shows that $\mathbb{F}^n$ and $\mathbb{F}^m$ cannot be isomorphic. □

Having now shown that the dimension of a vector space is well-defined we can then establish the dimension formula. Part of the proof of this theorem is to identify a basis for the image of a matrix. Note that this proof does not depend on the result that subspaces of finite dimensional vector spaces are finite dimensional. In fact for the subspaces under consideration, namely, the kernel and image, it is part of the proof to show that they are finite dimensional.

THEOREM 10. (The Dimension Formula) *Let* $A \in \text{Mat}_{n\times m}(\mathbb{F})$, *then*

$$m = \dim(\ker(A)) + \dim(\text{im}(A)).$$

PROOF. We use the above notation. We just saw that $\dim(\ker(A)) = m - k$, so it remains to check why $\dim(\text{im}(A)) = k$?

If

$$A = \begin{bmatrix} y_1 & \cdots & y_m \end{bmatrix},$$

then we have $y_i = P^{-1}x_i$, where

$$A_{\text{rref}} = \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}.$$

We know that each

$$x_j \in \text{span}\{e_1, ..., e_k\} = \text{span}\{x_{j_1}, ..., x_{j_k}\},$$

thus we have that

$$y_j \in \text{span}\{y_{j_1}, ..., y_{j_k}\}.$$

Moreover, as P is invertible we see that $y_{j_1},...,y_{j_k}$ must be linearly independent as $e_1,...,e_k$ are linearly independent. This proves that $y_{j_1},...,y_{j_k}$ form a basis for $\operatorname{im}(A)$. □

COROLLARY 9. (Subspace Theorem) *Let $M \subset \mathbb{F}^n$ be a subspace. Then M is finite dimensional and* $\dim(M) \leq n$.

PROOF. Recall from "Subspaces" that every subspace $M \subset \mathbb{F}^n$ has a complement. This means that we can construct a projection as in "Linear Maps and Subspaces" that has M as kernel. This means that M is the kernel for some $A \in \operatorname{Mat}_{n\times n}(\mathbb{F})$. Thus the previous theorem implies the claim. □

It might help to see an example of how the above constructions work.

EXAMPLE 43. *Suppose that we have a 4×7 matrix*

$$A = \begin{bmatrix} 0 & 1 & 4 & 1 & 0 & 3 & -1 \\ 0 & 0 & 0 & 1 & -2 & 5 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Then

$$A_{\text{rref}} = \begin{bmatrix} 0 & 1 & 4 & 0 & 2 & -2 & 0 \\ 0 & 0 & 0 & 1 & -2 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus $j_1 = 2$, $j_2 = 4$, and $j_3 = 7$. The complementary indices are $i_1 = 1$, $i_2 = 3$, $i_3 = 5$, and $i_4 = 6$. Hence

$$\operatorname{im}(A) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ 1 \\ 1 \end{bmatrix} \right\}$$

and

$$\ker(A) = \left\{ \begin{bmatrix} \xi_1 \\ -4\xi_3 - 2\xi_5 + 2\xi_6 \\ \xi_3 \\ 2\xi_5 - 5\xi_6 \\ \xi_5 \\ \xi_6 \\ 0 \end{bmatrix} : \xi_1, \xi_3, \xi_5, \xi_6 \in \mathbb{F} \right\}.$$

Our method for finding a basis for the image of a matrix leads us to a different proof of the rank theorem. The *column rank* of a matrix is simply the dimension of the image, in other words, the maximal number of linearly independent column vectors. Similarly the *row rank* is the maximal number of linearly independent rows. In other words, the row rank is the dimension of the image of the transposed matrix.

THEOREM 11. (The Rank Theorem) *Any $n \times m$ matrix has the property that the row rank is equal to the column rank.*

Proof. We just saw that the column rank for A and A_{rref} are the same and equal to k with the above notation. Because of the row operations we use, it is clear that the rows of A_{rref} are linear combinations of the rows of A. As the process can be reversed the rows of A are also linear combinations of the rows A_{rref}. Hence A and A_{rref} also have the same row rank. Now A_{rref} has k linearly independent rows and must therefore have row rank k. □

Using the rank theorem together with the dimension formula leads to an interesting corollary.

Corollary 10. *Let $A \in \mathrm{Mat}_{n\times n}(\mathbb{F})$. Then*

$$\dim(\ker(A)) = \dim\left(\ker\left(A^t\right)\right),$$

where $A^t \in \mathrm{Mat}_{n\times n}(\mathbb{F})$ is the transpose of A.

We are now going to clarify what type of matrices P occur when we do the row reduction to obtain $PA = A_{\mathrm{rref}}$. If we have an $n \times n$ matrix A with trivial kernel, then it must follow that $A_{\mathrm{rref}} = 1_{\mathbb{F}^n}$. Therefore, if we perform Gauss-Jordan elimination on the augmented matrix

$$A|1_{\mathbb{F}^n},$$

then we end up with an answer that looks like

$$1_{\mathbb{F}^n}|B.$$

The matrix B evidently satisfies $AB = 1_{\mathbb{F}^n}$. To be sure that this is the inverse we must also check that $BA = 1_{\mathbb{F}^n}$. However, we know that A has an inverse A^{-1}. If we multiply the equation $AB = 1_{\mathbb{F}^n}$ by A^{-1} on the left we obtain $B = A^{-1}$. This settles the uncertainty.

The space of all invertible $n \times n$ matrices is called the *general linear group* and is denoted by:

$$Gl_n(\mathbb{F}) = \left\{A \in \mathrm{Mat}_{n\times n}(\mathbb{F}) : \exists\ A^{-1} \in \mathrm{Mat}_{n\times n}(\mathbb{F}) \text{ with } AA^{-1} = A^{-1}A = 1_{\mathbb{F}^n}\right\}.$$

This space is a so called *group*. This means that we have a set G and a product operation $G \times G \to G$ denoted by $(g,h) \to gh$. This product operation must satisfy

(1) Associativity: $(g_1g_2)\,g_3 = g_1\,(g_2g_3)$.
(2) Existence of a unit $e \in G$ such that $eg = ge = g$.
(3) Existence of inverses: For each $g \in G$ there is $g^{-1} \in G$ such that $gg^{-1} = g^{-1}g = e$.

If we use matrix multiplication in $Gl_n(\mathbb{F})$ and $1_{\mathbb{F}^n}$ as the unit, then it is clear that $Gl_n(\mathbb{F})$ is a group. Note that we don't assume that the product operation in a group is commutative, and indeed it isn't commutative in $Gl_n(\mathbb{F})$ unless $n = 1$.

If a possibly infinite subset $S \subset G$ of a group has the property that any element in G can be written as a product of elements in S, then we say that S *generates* G.

We can now prove

Theorem 12. *The general linear group $Gl_n(\mathbb{F})$ is generated by the elementary matrices I_{kl}, $R_{kl}(\alpha)$, and $M_k(\alpha)$.*

Proof. We already observed that I_{kl}, $R_{kl}(\alpha)$, and $M_k(\alpha)$ are invertible and hence form a subset in $Gl_n(\mathbb{F})$. Let $A \in Gl_n(\mathbb{F})$, then we know that also $A^{-1} \in Gl_n(\mathbb{F})$. Now observe that we can find $P \in Gl_n(\mathbb{F})$ as a product of elementary matrices such that $PA^{-1} = 1_{\mathbb{F}^n}$. This was the content of the Gauss-Jordan elimination process for finding the inverse of a matrix. This means that $P = A$ and hence A is a product of elementary matrices. □

As a corollary we have:

Corollary 11. *Let $A \in \mathrm{Mat}_{n\times n}(\mathbb{F})$, then it is possible to find $P \in Gl_n(\mathbb{F})$ such that PA is upper triangular:*

$$PA = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ 0 & \beta_{22} & \cdots & \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_{nn} \end{bmatrix}$$

Moreover

$$\ker(A) = \ker(PA)$$

and $\ker(A) \neq \{0\}$ *if and only if the product of the diagonal elements in PA is zero:*

$$\beta_{11}\beta_{22}\cdots\beta_{nn} = 0.$$

We are now ready to see how the process of calculating A_{rref} using row operations can be interpreted as a change of basis in the image space.

Two matrices $A, B \in \mathrm{Mat}_{n\times m}(\mathbb{F})$ are said to be *row equivalent* if we can find $P \in Gl_n(\mathbb{F})$ such that $A = PB$. Thus row equivalent matrices are the matrices that can be obtained from each other via row operations. We can also think of row equivalent matrices as being different matrix representations of the same linear map with respect to different bases in $\mathbb{F}^n$. To see this consider a linear map $L : \mathbb{F}^m \to \mathbb{F}^n$ that has matrix representation A with respect to the standard bases. If we perform a change of basis in $\mathbb{F}^n$ from the standard basis $f_1, ..., f_n$ to a basis $y_1, ..., y_n$ such that

$$\begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} f_1 & \cdots & f_n \end{bmatrix} P,$$

i.e., the columns of P are regarded as a new basis for $\mathbb{F}^n$, then $B = P^{-1}A$ is simply the matrix representation for $L : \mathbb{F}^m \to \mathbb{F}^n$ when we have changed the basis in $\mathbb{F}^n$ according to P. This information can be encoded in the diagram

$$\begin{array}{ccc} \mathbb{F}^m & \xrightarrow{A} & \mathbb{F}^n \\ \downarrow 1_{\mathbb{F}^m} & & \downarrow 1_{\mathbb{F}^n} \\ \mathbb{F}^m & \xrightarrow{L} & \mathbb{F}^n \\ \uparrow 1_{\mathbb{F}^m} & & \uparrow P \\ \mathbb{F}^m & \xrightarrow{B} & \mathbb{F}^n \end{array}$$

When we consider abstract matrices rather than systems of equations we could equally well have performed column operations. This is accomplished by multiplying the elementary matrices on the right rather than the left. We can see explicitly what happens in the 2×2 case:

$$AI_{12} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \alpha_{12} & \alpha_{11} \\ \alpha_{22} & \alpha_{21} \end{bmatrix}$$

$$AR_{12}(\alpha) = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha\alpha_{11} + \alpha_{12} \\ \alpha_{21} & \alpha\alpha_{21} + \alpha_{22} \end{bmatrix}$$

$$AR_{21}(\alpha) = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} = \begin{bmatrix} \alpha_{11} + \alpha\alpha_{12} & \alpha_{12} \\ \alpha_{21} + \alpha\alpha_{22} & \alpha_{22} \end{bmatrix}$$

$$AM_1(\alpha) = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \alpha\alpha_{11} & \alpha_{12} \\ \alpha\alpha_{21} & \alpha_{22} \end{bmatrix}$$

$$AM_2(\alpha) = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha\alpha_{12} \\ \alpha_{21} & \alpha\alpha_{22} \end{bmatrix}$$

The only important and slightly confusing thing to be aware of is that, while $R_{kl}(\alpha)$ as a row operation multiplies row l by α and then adds it to row k, it now multiplies column k by α and adds it to column l as a column operation. This is because AE_{kl} is the matric whose l^{th} column is the k^{th} column of A and whose other columns vanish.

Two matrices $A, B \in \mathrm{Mat}_{n\times m}(\mathbb{F})$ are said to be *column equivalent* if $A = BQ$ for some $Q \in Gl_m(\mathbb{F})$. According to the above interpretation this corresponds to a change of basis in the domain space $\mathbb{F}^m$.

More generally we say that $A, B \in \mathrm{Mat}_{n\times m}(\mathbb{F})$ are *equivalent* if $A = PBQ$, where $P \in Gl_n(\mathbb{F})$ and $Q \in Gl_m(\mathbb{F})$. The diagram for the change of basis then looks like

$$\begin{array}{ccc} \mathbb{F}^m & \xrightarrow{A} & \mathbb{F}^n \\ \downarrow 1_{\mathbb{F}^m} & & \downarrow 1_{\mathbb{F}^n} \\ \mathbb{F}^m & \xrightarrow{L} & \mathbb{F}^n \\ \uparrow Q^{-1} & & \uparrow P \\ \mathbb{F}^m & \xrightarrow{B} & \mathbb{F}^n \end{array}$$

In this way we see that two matrices are equivalent if and only if they are matrix representations for the same linear map. Recall from the previous section that any linear map between finite dimensional spaces always has a matrix representation of the form

$$\begin{bmatrix} 1 & \cdots & & 0 & & 0 \\ & \ddots & & & & \\ 0 & \cdots & 1 & \vdots & & \vdots \\ & & & 0 & & 0 \\ \vdots & & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

where there are k ones in the diagonal if the linear map has rank k. This implies

Corollary 12. (Characterization of Equivalent Matrices) $A, B \in \mathrm{Mat}_{n\times m}(\mathbb{F})$ *are equivalent if and only if they have the same rank. Moreover any matrix of rank* k *is equivalent to a matrix that has* k *ones on the diagonal and zeros elsewhere.*

13.1. Exercises.

(1) Find bases for kernel and image for the following matrices.

(a) $\begin{bmatrix} 1 & 3 & 5 & 1 \\ 2 & 0 & 6 & 0 \\ 0 & 1 & 7 & 2 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 \\ 0 & 3 \\ 1 & 4 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

(d) $\begin{bmatrix} \alpha_{11} & 0 & \cdots & 0 \\ \alpha_{21} & \alpha_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{bmatrix}$ In this case it will be necessary to discuss whether or not $\alpha_{ii} = 0$ for each $i = 1, ..., n.$

(2) Find A^{-1} for each of the following matrices.

(a) $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

(3) Let $A \in \text{Mat}_{n\times m}(\mathbb{F})$. Show that we can find $P \in Gl_n(\mathbb{F})$ that is a product of matrices of the types I_{ij} and $R_{ij}(\alpha)$ such that PA is upper triangular.

(4) Let $A = \text{Mat}_{n\times n}(\mathbb{F})$. We say that A has an LU decomposition if $A = LU$, where L is lower triangular with 1s on the diagonal and U is upper triangular. Show that A has an LU decomposition if all the *leading principal minors* are invertible. The leading principal $k\times k$ minor is the $k \times k$ submatrix gotten from A by eliminating the last $n-k$ rows and columns. Hint: Do Gauss elimination using only $R_{ij}(\alpha)$.

(5) Assume that $A = PB$,where $P \in Gl_n(\mathbb{F})$
  (a) Show that $\ker(A) = \ker(B)$.
  (b) Show that if the column vectors $y_{i_1}, ..., y_{i_k}$ of B form a basis for $\text{im}(B)$, then the corresponding column vectors $x_{i_1}, ..., x_{i_k}$ for A form a basis for $\text{im}(A)$.

(6) Let $A \in \text{Mat}_{n\times m}(\mathbb{F})$.
  (a) Show that the $m \times m$ elementary matrices $I_{ij}, R_{ij}(\alpha), M_i(\alpha)$ when multiplied on the right correspond to column operations.
  (b) Show that we can find $Q \in Gl_m(\mathbb{F})$ such that AQ is lower triangular.
  (c) Use this to conclude that $\text{im}(A) = \text{im}(AQ)$ and describe a basis for $\text{im}(A)$.
  (d) Use Q to find a basis for $\ker(A)$ given a basis for $\ker(AQ)$ and describe how you select a basis for $\ker(AQ)$.

(7) Let $A \in \text{Mat}_{n\times n}(\mathbb{F})$ be upper triangular.
  (a) Show that $\dim(\ker(A)) \leq$ number of zero entries on the diagonal.
  (b) Give an example where $\dim(\ker(A)) <$ number of zero entries on the diagonal.

(8) In this exercise you are asked to show some relationships between the elementary matrices.

(a) Show that $M_i(\alpha) = I_{ij} M_j(\alpha) I_{ji}$.
(b) Show that $R_{ij}(\alpha) = M_j(\alpha^{-1}) R_{ij}(1) M_j(\alpha)$.
(c) Show that $I_{ij} = R_{ij}(-1) R_{ji}(1) R_{ij}(-1) M_j(-1)$.
(d) Show that $R_{kl}(\alpha) = I_{ki} I_{lj} R_{ij}(\alpha) I_{jl} I_{ik}$, where in case $i = k$ or $j = k$ we interpret $I_{kk} = I_{ll} = 1_{\mathbb{F}^n}$.

(9) A matrix $A \in Gl_n(\mathbb{F})$ is a *permutation matrix* if $Ae_1 = e_{\sigma(i)}$ for some bijective map (permutation)

$$\sigma : \{1, ..., n\} \to \{1, ..., n\}.$$

(a) Show that

$$A = \sum_{i=1}^{n} E_{\sigma(i)i}$$

(b) Show that A is a permutation matrix if and only if A has exactly one entry in each row and column which is 1 and all other entries are zero.
(c) Show that A is a permutation matrix if and only if it is a product of the elementary matrices I_{ij}.

(10) Assume that we have two fields $\mathbb{F} \subset \mathbb{L}$, such as $\mathbb{R} \subset \mathbb{C}$, and consider $A \in \mathrm{Mat}_{n\times m}(\mathbb{F})$. Let $A_{\mathbb{L}} \in \mathrm{Mat}_{n\times m}(\mathbb{L})$ be the matrix A thought of as an element of $\mathrm{Mat}_{n\times m}(\mathbb{L})$. Show that $\dim_{\mathbb{F}}(\ker(A)) = \dim_{\mathbb{L}}(\ker(A_{\mathbb{L}}))$ and $\dim_{\mathbb{F}}(\mathrm{im}(A)) = \dim_{\mathbb{L}}(\mathrm{im}(A_{\mathbb{L}}))$. Hint: Show that A and $A_{\mathbb{L}}$ have the same reduced row echelon form.

(11) Given $\alpha_{ij} \in \mathbb{F}$ for $i < j$ and $i, j = 1, ..., n$ we wish to solve

$$\frac{\xi_i}{\xi_j} = \alpha_{ij}.$$

(a) Show that this system either has no solutions or infinitely many solutions. Hint: try $n = 2, 3$ first.
(b) Give conditions on α_{ij} that guarantee an infinite number of solutions.
(c) Rearrange this system into a linear system and explain the above results.

14. Dual Spaces*

For a vector space V over $\mathbb{F}$ we define the *dual vector space* $V' = \hom(V, \mathbb{F})$ as the set of linear functions on V. One often sees the notation V^* for V'. However, we have reserved V^* for the conjugate vector space to a complex vector space. When V is finite dimensional we know that V and V' have the same dimension. In this section we shall see how the dual vector space can be used as a substitute for an inner product on V in case V doesn't come with a natural inner product (see chapter 3 for the theory on inner product spaces).

We have a natural dual pairing $V \times V' \to \mathbb{F}$ defined by $(x, f) = f(x)$ for $x \in V$ and $f \in V'$. We are going to think of (x, f) as a sort of inner product between x and f. Using this notation will enable us to make the theory virtually the same as for inner product spaces. Observe that this pairing is linear in both variables. Linearity in the first variable is a consequence of using linear functions in the second