A.1 Basic Algebra

A.1.1 Fields

In doing coding theory it is advantageous for our alphabet to have a certain amount of mathematical structure. We are familiar at the bit level with boolean addition (EXCLUSIVE OR) and multiplication (AND) within the set $\{0, 1\}$:

$+$	0	1
0	0	1
1	1	0

$\times$	0	1
0	0	0
1	0	1

We wish to give other alphabets, particularly finite ones, a workable arithmetic. The objects we study (and of which the set $\{0, 1\}$ together with the above operations is an example) are called fields. A field is basically a set that possesses an arithmetic having (most of) the properties that we expect — the ability to add, multiply, subtract, and divide subject to the usual laws of commutativity, associativity, and distributivity. The typical examples are the field of rational numbers (usually denoted $\mathbb{Q}$), the field of real numbers $\mathbb{R}$, and the field of complex numbers $\mathbb{C}$; however as just mentioned not all examples are so familiar. The integers do *not* constitute a field because in general it is not possible to divide one integer by another and have the result still be an integer.

field

A *field* is, by definition, a set F, say, equipped with two operations, $+$ (addition) and $\cdot$ (multiplication), which satisfy the following seven usual arithmetic axioms:

(1) (Closure) For each a and b in F, the sum $a + b$ and the product $a \cdot b$ are well-defined members of F.

(2) (Commutativity) For all a and b in F, $a+b = b+a$ and $a \cdot b = b \cdot a$.

(3) (Associativity) For all a, b, and c in F, $(a + b) + c = a + (b + c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

(4) (Distributivity) For all a, b, and c in F, $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(a + b) \cdot c = a \cdot c + b \cdot c$.

(5) (Existence of identity elements) There are distinct elements 0 and 1 of F such that, for all a in F, $a+0 = 0+a = a$ and $a \cdot 1 = 1 \cdot a = a$.

(6) (Existence of additive inverses) For each a of F there is an element $-a$ of F such that $a + (-a) = (-a) + a = 0$.

(7) (Existence of multiplicative inverses) For each a of F that does not equal 0, there is an element a^{-1} of F such that $a \cdot (a^{-1}) = (a^{-1}) \cdot a = 1$.

It should be emphasized that these common arithmetic assumptions are the only ones we make. In particular we make no flat assumptions about operations

called "subtraction" or "division". These operations are best thought of as the "undoing" respectively of addition and multiplication and, when desired, can be defined using the known operations and their properties. Thus subtraction can be defined by $a - b = a + (-b)$ using (6), and division defined by $a/b = a \cdot (b^{-1})$ using (7) (provided b is not 0).

Other familiar arithmetic properties that are not assumed as axioms either must be proven from the assumptions or may be false in certain fields. For instance, it is not assumed but can be proven that always in a field $(-1)\cdot a = -a$. (Try it!) A related, familiar result which can be proven for all fields F is that, given a and b in F, there is always a unique solution x in F to the equation $a+x = b$. On the other hand the properties of positive and/or negative numbers familiar from working in the rational field $\mathbb{Q}$ and the real field $\mathbb{R}$ do not have a place in the general theory of fields. Indeed there is no concept at all of "negative" or "positive" number for the complex field $\mathbb{C}$ or the field $\{0, 1\}$ discussed above.

The only thing keeping the integers $\mathbb{Z}$ from being a field is the axiom (7) concerning multiplicative inverses. Axioms (1)-(6) are valid for $\mathbb{Z}$, but (7) fails miserably; indeed 1 and -1 are the *only* integers that possess multiplicative inverses that are also integers. The integers do satisfy two axioms weaker than (7) but still useful.

($7'$) (Cancellation) If a is not 0 and $a \cdot b = a \cdot c$, then $b = c$.

($7''$) (No Zero Divisors) If $a \cdot b = 0$, then $a = 0$ or $b = 0$.

Axiom ($7'$) is a direct consequence of (7), because multiplying both sides of $a \cdot b = a \cdot c$ by a^{-1} leaves $b = c$. However (7) is not a consequence of ($7'$) as ($7'$) is true in $\mathbb{Z}$ while (7) is not. Similarly axiom ($7''$) is a consequence of (7). If one of a or b is not zero, then multiplying the lefthand side of $a \cdot b = 0$ by its inverse reveals the other as equal to 0. Again ($7''$) is true in $\mathbb{Z}$ while (7) is not, so that (7) is not a consequence of ($7''$).

In fact axioms ($7'$) and ($7''$) are equivalent in the following sense: if the set R has operations $+$ and $\cdot$ that satisfy (1) through (6), then either both axioms ($7'$) and ($7''$) hold or neither does. To see that ($7'$) implies ($7''$), apply ($7'$) to $a \cdot b = a \cdot 0$. On the other hand, to see that ($7''$) implies ($7'$), apply ($7''$) to $a \cdot (b - c) = 0$.

We are interested mainly in *finite fields*, those fields with a finite number of elements of which $\{0, 1\}$ is our only example so far. The most familiar way of giving a reasonable arithmetic to a finite set is to do modular arithmetic in the integers. For a fixed positive integer n, called the *modulus* we give the set $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ an arithmetic by first performing the usual integer addition or multiplication and then reducing the result modulo n back into the set $\mathbb{Z}_n$ by subtracting off multiples of n. This is "clock arithmetic" when $n = 12$ (or, these days, when $n = 24$).

finite fields

modulus

The question arises as to whether $\mathbb{Z}_n$ is a field. The field $\{0, 1\}$ already mentioned several times is nothing other than the integers mod 2, $\mathbb{Z}_2$. It is not difficult to check that $\mathbb{Z}_n$ with modular arithmetic satisfies axioms (1) through

(6). On the other hand, the answer as to whether $\mathbb{Z}_n$ satisfies (7) or the weaker (7′) and (7″) depends upon the particular value of the modulus n. For instance, all are true when $n = 2$. For $n = 6$ we have $2 \cdot 3 = 0 \pmod 6$ (whence $2 \cdot 3 = 2 \cdot 0 \pmod 6$); yet neither 2 nor 3 equals 0 in the integers mod6. Therefore each of (7), (7′), and (7″) is false in $\mathbb{Z}_6$.

Although the arithmetics of $\mathbb{Z}$ and $\mathbb{Z}_6$ do not make them into fields, the structures clearly are of interest. A set F equipped with an addition and multiplication that satisfy (1) through (6) we shall call a *commutative ring*. ("Commutative" because the multiplication satisfies the commutative law.) A *ring* satisfies each of (1) through (6) with the possible exception of the commutativity of multiplication. If the commutative ring F additionally satisfies the equivalent axioms (7′) and (7″), then it is called an *integral domain* (in honor of the integers!). Clearly all fields and all integral domains are commutative rings. As (7) implies (7′) and (7″), every field is also an integral domain while the integers provide the prime example of an integral domain that is not a field. $\mathbb{Z}_6$ is an example of a commutative ring that is not an integral domain and so certainly not a field.

commutative ring

ring

integral domain

An element of a ring that has an inverse, as in (7), is called a *unit*; so fields are exactly those commutative rings in which every nonzero element is a unit.

unit

(A.1.1) LEMMA. *Let n be an integer larger than 1. The following are equivalent:*

(1) *n is a prime;*
(2) *$\mathbb{Z}_n$ is an integral domain;*
(3) *$\mathbb{Z}_n$ is a field.*

PROOF. *(1) implies (2)*: Assume n is a prime, and that $a \cdot b = 0$ in $\mathbb{Z}_n$. Then the integer ab is a multiple of n. As n is prime, it divides either a or b; hence either a or b is 0 in $\mathbb{Z}_n$. This verifies axiom (7″).

(2) implies (1): As with our example of $\mathbb{Z}_6$, if n is not prime, then each factorization $ab = n$ in $\mathbb{Z}$ with $1 < a, b < n$ gives rise to an equality $a \cdot b = 0$ in $\mathbb{Z}_n$ with neither a nor b equal to 0. Thus if n is not a prime, then $\mathbb{Z}_n$ does not satisfy (7″) and so is not an integral domain.

(3) implies (2) as axiom (7) implies axioms (7′) and (7″).

(2) implies (3): Let $\mathbb{Z}_n^\sharp = \{1, \ldots, n-1\}$, the set of nonzero elements of $\mathbb{Z}_n$. Choose $a \in \mathbb{Z}_n^\sharp$. As (by assumption) $\mathbb{Z}_n$ is an integral domain, for distinct elements $z_1, z_2 \in \mathbb{Z}_n^\sharp$, the products $a \cdot z_1$ and $a \cdot z_2$ are also distinct by (7′). Therefore the set $a\mathbb{Z}_n^\sharp = \{a \cdot z | z \in \mathbb{Z}_n^\sharp\}$ contains $n-1$ distinct members of $\mathbb{Z}_n$. Indeed $0 \notin a\mathbb{Z}_n^\sharp$ by (7″), so $a\mathbb{Z}_n^\sharp$ is a subset of $\mathbb{Z}_n^\sharp$. Thus $a\mathbb{Z}_n^\sharp$ is a subset of $\mathbb{Z}_n^\sharp$ containing the same number of elements as $\mathbb{Z}_n^\sharp$. We conclude that $\mathbb{Z}_n^\sharp = a\mathbb{Z}_n^\sharp$. In particular, $1 \in \mathbb{Z}_n^\sharp = a\mathbb{Z}_n^\sharp$; and there is a z in $\mathbb{Z}_n^\sharp$ with $a \cdot z = 1$. Therefore all the nonzero members of $\mathbb{Z}_n$ have multiplicative inverses in $\mathbb{Z}_n$, and $\mathbb{Z}_n$ is a field. □

(A.1.2) PROBLEM. *Extend the argument of Lemma A.1.1 that (2) implies (3) to prove the more general result that every finite integral domain is in fact a field. (The integers of course provide an example of an infinite integral domain that is not a field.)*

If F is a field and K is a subset of F, K is said to be a *subfield* of F provided that the set K equipped with the addition and multiplication of F is a field in its own right. If this is the case, then we write $K \leq F$ or $F \geq K$. The addition and multiplication of K will be commutative, associative, and distributive as they already are in F; so the crucial requirements are that K be closed under addition and multiplication and contain the additive and multiplicative inverses of all its elements. As examples, the rational field $\mathbb{Q}$ is a subfield of the real field $\mathbb{R}$, which in turn is a subfield of the complex field $\mathbb{C}$. subfield

If K is a subfield of F, then we call F an *extension field* of K. Thus $\mathbb{C}$ is an extension field of $\mathbb{R}$, and both $\mathbb{C}$ and $\mathbb{R}$ are extension fields of $\mathbb{Q}$. As we shall mainly be concerned with finite fields, important examples of subfields for us are provided by the next result. extension field

(A.1.3) LEMMA. *Let F be a finite field, and consider the subset K of F composed of all elements of F that can be written as a sum of 1's:*

$$K = \{1,\ 1+1,\ 1+1+1,\ 1+1+1+1,\ \dots\}.$$

Then K is a subfield $\mathbb{Z}_p$ of F, for some prime p.

PROOF. Notice that by definition K is closed under addition, while an easy application of the distributive law in F shows that K is closed under multiplication.

As F is finite, so is K. Therefore there are distinct positive integers m and n (m larger than n) with the sum of m 1's equal to the sum of n 1's. (Indeed, there are many such pairs m, n.) Equivalently the sum of $m-n$ 1's equals 0 in F and K, $m-n$ a positive integer. Let p be the smallest positive integer such that 0 is a sum of p 1's in F and K. We conclude that K is composed precisely of the p distinct elements

$$1,\ 1+1,\ \dots,\ \sum_{i=1}^{p} 1 = \overbrace{1+\cdots+1}^{p \text{ times}} = 0.$$

The set K is therefore a copy of $\mathbb{Z}_p$. As K is contained in the field F, no two nonzero members of K have product 0; so by Lemma A.1.1 p is a prime, completing the result. □

The prime p of Lemma A.1.3 is called the *characteristic* of the field F, and K is (for obvious reasons) called the *prime subfield* of F. characteristic prime subfield

A.1.2 Vector spaces

The passage from the real line to real, Euclidean, three-dimensional space is the most familiar case of the passage from a field to a vector space over a field. If F is a field and n is any positive integer, we may use the arithmetic structure of F to give the set F^n of n-tuples from F,

$$F^n = \{(a_1, a_2, \ldots, a_n) \,|\, a_i \in F\},$$

additive and multiplicative structures as well. We define "vector addition" of members of F^n via

$$(a_1, a_2, \ldots, a_n) \oplus (b_1, b_2, \ldots, b_n) = (c_1, c_2, \ldots, c_n)$$

where $c_i = a_i + b_i$ (addition in F), for each $i = 1, \ldots, n$. We define "scalar multiplication" of members of F^n by members of F via

$$\alpha \star (a_1, a_2, \ldots, a_n) = (\alpha \cdot a_1, \alpha \cdot a_2, \ldots, \alpha \cdot a_n)$$

where $\alpha \cdot a_i$ is the usual multiplication in the field F. These two operations make F^n into a vector space over F.

vector space

Given a field F, a *vector space V over F* is, by definition, a set V (whose members are called the *vectors* of V) equipped with two operations $\oplus$ (vector addition) and $\star$ (scalar multiplication), satisfying the following:

(1) (Closure) For each $\mathbf{v}$ and $\mathbf{w}$ in V, $\mathbf{v} \oplus \mathbf{w}$ is a well-defined member of V. For each α in F and $\mathbf{v}$ in V, $\alpha \star \mathbf{v}$ is a well-defined member of V.

(2) (Commutativity) For each $\mathbf{v}$ and $\mathbf{w}$ in V, $\mathbf{v} \oplus \mathbf{w} = \mathbf{w} \oplus \mathbf{v}$.

(3) (Associativity) For each $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V, $(\mathbf{u} \oplus \mathbf{v}) \oplus \mathbf{w} = \mathbf{u} \oplus (\mathbf{v} \oplus \mathbf{w}$. For each α, β in F and $\mathbf{v}$ in V, $(\alpha \cdot \beta) \star \mathbf{v} = \alpha \star (\beta \star \mathbf{v})$.

(4) (Distributivity) For each α, β in F and $\mathbf{v}, \mathbf{w}$ in V, $(\alpha + \beta) \star \mathbf{v} = (\alpha \star \mathbf{v}) \oplus (\beta \star \mathbf{v})$ and $\alpha \star (\mathbf{v} \oplus \mathbf{w}) = (\alpha \star \mathbf{v}) \oplus (\alpha \star \mathbf{w})$.

(5) (Existence of vector identity) There is a vector $\mathbf{0}$ of V such that, for each $\mathbf{v}$ of V, $\mathbf{v} \oplus \mathbf{0} = \mathbf{0} \oplus \mathbf{v} = \mathbf{v}$.

(6) (Existence of vector inverses) For each $\mathbf{v}$ of V there is a vector $-\mathbf{v}$ of V such that $\mathbf{v} \oplus (-\mathbf{v}) = (-\mathbf{v}) \oplus \mathbf{v} = \mathbf{0}$.

(7) (Scalar identity properties) For each $\mathbf{v}$ of V, $1 \star \mathbf{v} = \mathbf{v}$ and $0 \star \mathbf{v} = \mathbf{0}$.

F-space

For brevity, we sometimes say that V is an *F-vector space* or even an *F-space.* Note that scalar multiplication $\star$ is not multiplication of one vector by another but multiplication of a vector in V by a member of the field F. (F is usually called the *scalar field* of the vector space V, and its members are *scalars.*)

scalar field
scalars

The set F^n with the operations defined above is now easily seen to be a vector space over F. The similarity between the above axioms and those of Section A.1.1 explains the fact that F may be thought of as a vector space over itself. (After all, the distinction between F and F^1 is merely a pair of parentheses.) Many examples of vector spaces do not resemble the space of n-tuples at all. For instance, the set of all continuous and differentiable functions on the real line is a vector space over the real numbers.

Most of the vector spaces we shall study will naturally sit inside vector spaces F^n (because the spaces F^n are the natural universes for the codes we study). A subset W of the vector space V over F is a *subspace* of V if the operations of V give W the structure of a vector space over F in its own right. In this case we shall write $W \leq V$ or $V \geq W$. As most of the axioms (2)-(7) will have already been checked within V, the main force of this definition is in the assumption that W is closed as in (1). In fact, the subset W of V will be a subspace of V if and only if, for all α in F and all $\mathbf{v}, \mathbf{w}$ in W, $\alpha \star \mathbf{v}$ is in W and $\mathbf{v} \oplus \mathbf{w}$ is in W. Thus V itself and $\{\mathbf{0}\}$ are rather trivial subspaces of V. More typical is the subspace of F^n composed of all vectors $(a_1, a_2, \ldots, a_n)$ with $a_1 + a_2 + \cdots + a_n = 0$. subspace

(A.1.4) PROBLEM. *Prove that the nonempty subset W of the F-vector space V is a subspace if and only if $\alpha\mathbf{v} + \mathbf{w} \in W$, for all $\mathbf{v}, \mathbf{w} \in W$ and $\alpha \in F$.*

If W is a subspace of V, then a *cosets* of W in V is a translate of W by some fixed vector. If we translate each vector of W by the vector $\mathbf{v}$, we get the coset $\mathbf{x} + W = \{\mathbf{x} + \mathbf{w} \mid \mathbf{w} \in W\}$. You should convince yourself that if $\mathbf{y} \in \mathbf{x} + W$, then $\mathbf{y} + W = \mathbf{x} + W$; so two cosets are either disjoint or equal. As an example, a typical subspace of dimension 2 in 3-dimensional Euclidean space is a plane through the origin, while a typical coset is a translate of such a subspace and so is a plane that need not be through the origin. cosets

One way of constructing a subspace of the F-vector space V is by taking the *span* $\langle S \rangle$ of a nonempty subset S of V. This is, by definition, the smallest subspace of V that contains S; however this may not be the best way of thinking of $\langle S \rangle$. We usually view $\langle S \rangle$ instead as the subspace composed of all linear combinations of members of S: span

$$\langle S \rangle = \{ \sum_{\mathbf{v} \in S} \alpha_{\mathbf{v}} \mathbf{v} \mid \alpha_{\mathbf{v}} \in F \}.$$

You should convince yourself that these two definitions of $\langle S \rangle$ are equivalent. If $V = \langle S \rangle$, then S is a called a *spanning set* in V. spanning set

A *basis* of the vector space V is a minimal spanning set for V, a set that spans V but no proper subset of it spans V. basis

(A.1.5) THEOREM. *If the vector space V has a finite basis $\mathcal{B}$, then every basis of V contains the same number of vectors as $\mathcal{B}$.*

This theorem will be proven in the following subsection. (The theorem is in fact true without the assumption that $\mathcal{B}$ is finite.) The common size for the bases of V is the *dimension* of V. dimension

linearly dependent

A set Δ of vectors from V is called *linearly dependent* if there is a set of coefficients $\alpha_{\mathbf{v}}$, for $\mathbf{v} \in \Delta$, such that the linear combination $\sum_{\mathbf{v}\in\Delta} \alpha_{\mathbf{v}}\mathbf{v}$ equals $\mathbf{0}$. The equation

$$\sum_{\mathbf{v}\in\Delta} \alpha_{\mathbf{v}}\mathbf{v} = \mathbf{0}$$

linear dependence

is then called a *linear dependence* of Δ.

linearly independent

A subset Δ is *linearly independent* if it is not linearly dependent. The maximal linearly independent subsets of V are precisely the bases of V. (Check!) In particular, every linearly independent subset of V belongs to a basis. (For infinite dimensional spaces, this is the best way to see that a basis exists.)

(A.1.6) Problem.

(a) Let E be an extension field of F. Prove that E is a vector space over F with scalar multiplication induced by the field multiplication of E.

(b) Using (1), show that every finite field has a prime power number of elements.

$GF(q)$
$\mathbb{F}_q$

If q is a power of a prime, we often write $GF(q)$ or $\mathbb{F}_q$ for a field containing q elements.

Remarks on notation

Notice that in vector spaces we have two concepts of "addition" ($+$ in F and $\oplus$ in V) and two of "multiplication" ($\cdot$ in F and $\star$ in V) and that for formal precision we must distinguish between them. (See, for instance, axioms (3) and (4) above.) Often to simplify notation we adopt the usual practice of denoting all forms of addition by $+$ and all forms of multiplication by juxtaposition; so for α, β in F and $\mathbf{v}, \mathbf{w}$ in V we usually write

$$\alpha\beta \text{ for } \alpha \cdot \beta \text{ ; } \mathbf{v} + \mathbf{w} \text{ for } \mathbf{v} \oplus \mathbf{w} \text{ ; and } \alpha\mathbf{v} \text{ for } \alpha \star \mathbf{v}.$$

In doing this we risk ambiguity. To counter this possibility we often adopt other conventions which may already have been noticed. For instance, we usually write vectors in boldface thus: $\mathbf{v}$.

A.1.3 Matrices

Just as we can examine vector spaces over arbitrary fields, so can we define matrices with entries from an arbitrary field. If K is a field, we denote by $K^{m,n}$ the collection of all $m \times n$ matrices with entries from K. Notice that the vector space K^n of row vectors of length n is equal to $K^{1,n}$. The vector space of column vectors of length m is $K^{m,1}$. The usual componentwise addition and subtraction is defined on $K^{m,n}$ and has all the expected properties. Together with scalar multiplication, these give $K^{m,n}$ the structure of a vector space over K of dimension mn.

Matrix multiplication is also defined by the familiar formula (*i.e.*, entries of the product being the dot product of rows of the first matrix with columns of the second). Matrix multiplication also has all the expected properties — associativity, distributivity over addition, block multiplication. Because the most usual matrix manipulations involve only addition, subtraction, and multiplication, the entries need not always be restricted to a field but might instead be from an integral domain (or even a ring).

You may notice that the set $K^{n,n}$ of square matrices together with the operations of matrix addition and multiplication satisfies all the axioms (1) through (6) with the exception of commutativity of multiplication. Thus $K^{n,n}$ is an example of a noncommutative ring.

If A is an $m \times n$ matrix with entries from the field K, then the *row space* of A, $\mathrm{RS}(A)$, is the subspace of K^n that is spanned by the rows of A. (We shall often look at codes that have been defined as the row space of certain matrices.) Similarly the *column space* of A, $\mathrm{CS}(A)$, is the subspace of $K^{m,1}$ spanned by the columns of A. The *null space* of A, $\mathrm{NS}(A)$, is the space of column vectors $\mathbf{x} \in K^{n,1}$ such that $A\mathbf{x} = \mathbf{0}$. (Notice that the null space can be thought of as the space of all linear dependencies on the set of columns.) The dimension of the row space of A is the *row rank* of A, and the dimension of the column space of A is the *column rank* of A. The dimension of $\mathrm{NS}(A)$ is the *nullity* of A.

row space
column space
null space
row rank
column rank
nullity

More complicated but familiar matrix processes can also be done over arbitrary fields. In particular, Gauss-Jordan elimination is still available. That is, by sequences of elementary row operations on a matrix it is possible to transform the matrix into reduced row echelon form. Several of the standard consequences of Gaussian elimination then become available. In particular we have:

(A.1.7) THEOREM. *Let A be an $m \times n$ matrix with entries from the field K.*

(1) *The row rank of A equals the column rank of A. (This common dimension being called the* rank *of A.)*

(2) *The rank of A plus the nullity of A equals n, the number of columns of A.*

rank

Before proving this theorem, we give a detailed discussion of echelon form and its properties. The *leading entry* of a row is its first nonzero entry, reading from left to right.

The matrix A is said to be in *row echelon form* if it satisfies:

leading entry
row echelon form

(1) the leading entry of each row is to the right of the leading entries of previous rows;

(2) all rows composed entirely of zeros are at the bottom of the matrix.

reduced row echelon form
RREF

The matrix A is said to be in *reduced row echelon form* **RREF**, if it additionally satisfies:

(3) the leading entry of each row equals 1 and is the only nonzero entry in its column.

pivot entries
pivot columns

The various leading entries of the matrix $\mathbf{RREF}(A)$ are also sometimes called the *pivot entries* of $\mathbf{RREF}(A)$ and the columns containing them are the *pivot columns* of $\mathbf{RREF}(A)$ and A. The row rank of $\mathbf{RREF}(A)$ (indeed any matrix in row echelon form) is particularly easy to calculate; it is just the number of nonzero rows. It only takes a few seconds more to realize that this is also equal to the column rank of $\mathbf{RREF}(A)$. We will reduce the proof of Theorem A.1.7 to this special case, where we have just seen that the theorem (or at least its first part) is evident.

Elementary row operations have one of three forms:

(i) subtracting a multiple of one row from another;

(ii) interchanging two rows;

(iii) multiplying a row by a nonzero constant.

The usual elimination techniques then give:

(A.1.8) THEOREM. *Every matrix A with entries from a field can be transformed by a sequence of elementary row operations into a matrix $\mathbf{RREF}(A)$ that is in reduced row echelon form.* □

The verification is routine, but it is important that the matrix entries are from a field. For more general rings the result may not be true. (Imagine what could be done by integer row operations to a matrix with entries from $\mathbf{Z}$ whose first column contained only even integers.)

The notation suggests that $\mathbf{RREF}(A)$ is uniquely determined. This is indeed the case.

(A.1.9) PROBLEM. *Prove that the matrix $A \in K^{m,n}$, K a field, has a unique row reduced echelon form. (*HINT: *Prove that every vector of* $\mathrm{RS}(A)$ *has its leftmost nonzero entry in a pivot column, then either (i) try to write the rows of a second* **RREF** *as linear combinations of the rows of the first, or (ii) observe that the pivot columns are the leftmost columns that form a basis for* $\mathrm{CS}(A)$*.)*

As expected, each elementary row operation can be accomplished through left multiplication by an appropriate elementary matrix. Let $a\epsilon_{i,j}$ be the matrix that has a in its (i,j)-entry and 0's elsewhere (and write $\epsilon_{i,j}$ for $1\epsilon_{i,j}$), and let I be the identity matrix. Then left multiplication by

(i) $I + a\epsilon_{i,j}$ adds a times row j to row i;

(ii) $I - \epsilon_{i,i} - \epsilon_{j,j} + \epsilon_{i,j} + \epsilon_{j,i}$ interchanges rows i and j;

(iii) $I + (a-1)\epsilon_{i,i}$ multiplies row i by a.

The inverse of $I + a\epsilon_{i,j}$ is $I - a\epsilon_{i,j}$; $I - \epsilon_{i,i} - \epsilon_{j,j} + \epsilon_{i,j} + \epsilon_{j,i}$ is its own inverse; and $I + (a^{-1} - 1)\epsilon_{i,i}$ is the inverse of $I + (a-1)\epsilon_{i,i}$ for nonzero a. Therefore each elementary matrix is invertible. In particular we have $XA = \mathbf{RREF}(A)$, where the invertible matrix X is the product of those elementary matrices that correspond to the elementary row operations that take A to $\mathbf{RREF}(A)$.

(A.1.10) PROBLEM. *Let Y be an invertible $k \times k$ matrix with entries from the field K, and let A be the $k \times 2k$ matrix $\big(Y \,|\, I\big)$, the columns of Y followed by the columns of a $k \times k$ identity matrix. Prove that* $\mathbf{RREF}(A) = \big(I \,|\, Y^{-1}\big)$.

(A.1.11) PROBLEM. *Let Y be a $k \times k$ matrix with entries from the field K. Prove that the following are equivalent:*
(a) *Y is invertible;*
(b) $\mathrm{NS}(Y) = \mathbf{0}$*;*
(c) *Y has rank k;*
(d) $\mathbf{RREF}(Y) = I$.

(A.1.12) PROPOSITION. *Let A be an $m \times n$ matrix with entries from the field K.*
(1) *The column rank of* $\mathbf{RREF}(A)$ *equals the row rank of* $\mathbf{RREF}(A)$.
(2) $\mathrm{RS}(\mathbf{RREF}(A)) = \mathrm{RS}(A)$;
(3) $\mathrm{NS}(\mathbf{RREF}(A)) = \mathrm{NS}(A)$;
(4) $\dim(\mathrm{CS}(\mathbf{RREF}(A)) = \dim(\mathrm{CS}(A))$*, that is,* $\mathbf{RREF}(A)$ *has column rank equal to the column rank of A.*

PROOF. (1) Both numbers are equal to the number of pivot entries in $\mathbf{RREF}(A)$. Each of (2), (3), and (4) can be proven using the fact that there is an invertible matrix X with $XA = \mathbf{RREF}(A)$. For (4) it should be noted that (whether X is invertible or not) we have $\mathrm{CS}(XA) = X\mathrm{CS}(A) = \{\, X\mathbf{a} \,|\, \mathbf{a} \in \mathrm{CS}(A) \,\}$. □

(A.1.13) PROBLEM. *Prove completely parts (2), (3), and (4) of the proposition. Give an example that shows that* $\mathrm{CS}(\mathbf{RREF}(A))$ *and* $\mathrm{CS}(A)$ *need not be equal.*

If Σ is a set of vectors in F^n, then we can easily find a basis for $\langle \Sigma \rangle$ by forming the matrix A whose rows are the members of Σ and then passing to $\mathbf{RREF}(A)$ with its nonzero rows giving the desired basis. This observation is the basis for our proof of Theorem A.1.5: a vector space V with a finite basis $\mathcal{B}$ has all of its bases of size $|\mathcal{B}|$.

PROOF OF THEOREM A.1.5.
Choose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_d\}$ to be a basis for V of smallest size (necessarily finite). Let $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_d, \ldots\}$ be a second basis of V. Note that $|\mathcal{C}| \geq d = |\mathcal{B}|$

by choice. (If we encounter a second basis that is infinite, then for $\mathcal{C}$ we instead choose a finite subset of the basis that has at least $d+1$ members.)

For each i, write $\mathbf{c}_i$ as a linear combination of members of $\mathcal{B}$:

$$\mathbf{c}_i = \sum_{j=1}^{d} a_{i,j}\mathbf{b}_j\,;$$

and let the matrix A have (i,j)-entry $a_{i,j}$. As $\mathcal{C}$ is linearly independent, the row rank of A equals $|\mathcal{C}|$. However by Proposition A.1.12 the row rank of A equals the row rank of $\mathbf{RREF}(A)$ which is at most d, the number of columns of A. Therefore $|\mathcal{C}| \leq d$, completing the proof. □

For any matrix A, another advantage to having $R = \mathbf{RREF}(A)$ available is the ease with which its null space (and so that of A) can be calculated. Let the (i,j)-entry of R be $r_{i,j}$, and assume that the pivot entries are $r_{i,p(i)} = 1$, for $i = 1,\ldots,r$, (r being the row rank of A). Set $\mathcal{P} = \{\, p(i) \mid i = 1,\ldots,r\}$, the indices of the pivot columns of R.

For each nonpivot column $k \notin \mathcal{P}$ we construct a null vector $\mathbf{n}_k$ of R with a 1 in position k and 0 in all other nonpivot columns. The j-entry of $\mathbf{n}_k$ is given by:

$$\begin{array}{ll} (\mathbf{n}_k)_j = 1 & \text{if } j = k; \\ (\mathbf{n}_k)_j = 0 & \text{if } j \neq k \text{ and } j \notin \mathcal{P}; \\ (\mathbf{n}_k)_j = -r_{i,k} & \text{if } j = p(i) \in \mathcal{P}. \end{array}$$

This produces $n - r$ linearly independent vectors of $\mathrm{NS}(R)$. It is easy to see that $\mathbf{0}$ is the only null vector of R (and A) that is 0 in all nonpivot columns. Thus $\{\mathbf{n}_k \mid k \notin \mathcal{P}\}$ is a basis of $\mathrm{NS}(R) = \mathrm{NS}(A)$.

(A.1.14) Problem. *Check that each $\mathbf{n}_k$ is indeed a null vector of R, and supply the remaining details of the proof that these vectors form a basis for* $\mathrm{NS}(R)$.

In particular we have just proven that the nullity of A is equal to the number of nonpivot columns in $\mathbf{RREF}(A)$. This together with Proposition A.1.12 allows us to prove Theorem A.1.7 easily.

Proof of Theorem A.1.7.
For part (1), we have:

$$\begin{array}{rcl} \text{row rank of } A & = & \text{row rank of } \mathbf{RREF}(A) \text{ by A.1.12(2)} \\ & = & \text{column rank of } \mathbf{RREF}(A) \text{ by A.1.12(1)} \\ & = & \text{column rank of } A \text{ by A.1.12(4).} \end{array}$$

For part (2), if n is the number of columns in A (and $\mathbf{RREF}(A)$), then

$$\begin{array}{rcl} \text{rank of } A & = & \text{number of pivot columns in } \mathbf{RREF}(A) \text{ by A.1.12(1)} \\ & = & n \text{ minus number of nonpivot columns in } \mathbf{RREF}(A) \\ & = & n \text{ minus the nullity of } A \text{ by the above.} \end{array}$$

Thus both parts of the theorem are proved. □

We are familiar with the fact that division by matrices is a much trickier process than the other three arithmetic operations. In particular some concept

of determinant is usually needed to talk about matrix inverses. Again the usual theory of determinants carries over to square matrices over arbitrary fields (and even rings). The standard formula gives a multiplicative function from $K^{n,n}$ into K. We shall have little need for this and leave the most elementary case as an exercise.

(A.1.15) PROBLEM. *For a* 2×2 *matrix* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *with entries from a commutative ring* R*, we define the* determinant $\det(A) = ad - bc$. determinant

(a) *Prove that if* A *and* B *are both* 2×2 *matrices with entries from* R *then* $\det(AB) = \det(A)\det(B)$.

(b) *Prove that* A *has a matrix inverse with entries from* R *if and only if* $\det(A)$ *has an inverse in* R.

(c) *Prove that when* R *is a field,* A *has a matrix inverse with entries from* R *if and only if* $\det(A) \neq 0$.