14
Matrices

In this chapter, we discuss basic definitions and results concerning matrices. We shall start out with a very general point of view, discussing matrices whose entries lie in an arbitrary ring R. Then we shall specialize to the case where the entries lie in a field F, where much more can be said.

One of the main goals of this chapter is to discuss "Gaussian elimination," which is an algorithm that allows us to efficiently compute bases for the image and kernel of an F-linear map.

In discussing the complexity of algorithms for matrices over a ring R, we shall treat a ring R as an "abstract data type," so that the running times of algorithms will be stated in terms of the number of arithmetic operations in R. If R is a finite ring, such as $\mathbb{Z}_m$, we can immediately translate this into a running time on a RAM (in later chapters, we will discuss other finite rings and efficient algorithms for doing arithmetic in them).

If R is, say, the field of rational numbers, a complete running time analysis would require an additional analysis of the sizes of the numbers that appear in the execution of the algorithm. We shall not attempt such an analysis here—however, we note that all the algorithms discussed in this chapter do in fact run in polynomial time when $R = \mathbb{Q}$, assuming we represent rational numbers as fractions in lowest terms. Another possible approach for dealing with rational numbers is to use floating point approximations. While this approach eliminates the size problem, it creates many new problems because of round-off errors. We shall not address any of these issues here.

14.1 Basic definitions and properties

Throughout this section, R denotes a ring.

For positive integers m and n, an $m \times n$ **matrix** A over a ring R is a rectangular array

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

where each entry a_{ij} in the array is an element of R; the element a_{ij} is called the (i, j) **entry** of A, which we denote by $A(i, j)$. For $i = 1, \ldots, m$, the i**th row** of A is

$$(a_{i1}, \ldots, a_{in}),$$

which we denote by $\mathrm{Row}_i(A)$, and for $j = 1, \ldots, n$, the j**th column** of A is

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix},$$

which we denote by $\mathrm{Col}_j(A)$. We regard a row of A as a $1 \times n$ matrix, and a column of A as an $m \times 1$ matrix.

The set of all $m \times n$ matrices over R is denoted by $R^{m \times n}$. Elements of $R^{1 \times n}$ are called **row vectors (of dimension** n**)** and elements of $R^{m \times 1}$ are called **column vectors (of dimension** m**)**. Elements of $R^{n \times n}$ are called **square matrices (of dimension** n**)**. We do not make a distinction between $R^{1 \times n}$ and $R^{\times n}$; that is, we view standard n-tuples as row vectors.

We can define the familiar operations of **matrix addition** and **scalar multiplication**:

- If $A, B \in R^{m \times n}$, then $A + B$ is the $m \times n$ matrix whose (i, j) entry is $A(i, j) + B(i, j)$.
- If $c \in R$ and $A \in R^{m \times n}$, then cA is the $m \times n$ matrix whose (i, j) entry is $cA(i, j)$.

The $m \times n$ **zero matrix** is the $m \times n$ matrix, all of whose entries are 0_R; we denote this matrix by $0_R^{m \times n}$ (or just 0, when the context is clear).

Theorem 14.1. *With addition and scalar multiplication as defined above, $R^{m \times n}$ is an R-module. The matrix $0_R^{m \times n}$ is the additive identity, and the additive inverse of a matrix $A \in R^{m \times n}$ is the $m \times n$ matrix whose (i, j) entry is $-A(i, j)$.*

Proof. To prove this, one first verifies that matrix addition is associative and commutative, which follows from the associativity and commutativity of addition in R. The claims made about the additive identity and additive inverses are also easily verified. These observations establish that $R^{m\times n}$ is an abelian group. One also has to check that all of the properties in Definition 13.1 hold. We leave this to the reader. □

We can also define the familiar operation of **matrix multiplication**:

- If $A \in R^{m\times n}$ and $B \in R^{n\times p}$, then AB is the $m \times p$ matrix whose (i, k) entry is
$$\sum_{j=1}^{n} A(i,j)B(j,k).$$

The $n \times n$ **identity matrix** is the matrix $I \in R^{n\times n}$, where $I(i,i) := 1_R$ and $I(i,j) := 0_R$ for $i \neq j$. That is, I has 1_R's on the diagonal that runs from the upper left corner to the lower right corner, and 0_R's everywhere else.

Theorem 14.2.

(i) Matrix multiplication is associative; that is, $A(BC) = (AB)C$ for all $A \in R^{m\times n}$, $B \in R^{n\times p}$, and $C \in R^{p\times q}$.

(ii) Matrix multiplication distributes over matrix addition; that is, $A(C + D) = AC+AD$ and $(A+B)C = AC+BC$ for all $A, B \in R^{m\times n}$ and $C, D \in R^{n\times p}$.

(iii) The $n \times n$ identity matrix $I \in R^{n\times n}$ acts as a multiplicative identity; that is, $AI = A$ and $IB = B$ for all $A \in R^{m\times n}$ and $B \in R^{n\times m}$; in particular, $CI = C = IC$ for all $C \in R^{n\times n}$.

(iv) Scalar multiplication and matrix multiplication associate; that is, $c(AB) = (cA)B = A(cB)$ for all $c \in R$, $A \in R^{m\times n}$, and $B \in R^{n\times p}$.

Proof. All of these are trivial, except for (i), which requires just a bit of computation to show that the (i, ℓ) entry of both $A(BC)$ and $(AB)C$ is equal to (as the reader may verify)
$$\sum_{\substack{1\le j\le n\\ 1\le k\le p}} A(i,j)B(j,k)C(k,\ell). \quad \square$$

Note that while matrix addition is commutative, matrix multiplication in general is not. Indeed, Theorems 14.1 and 14.2 imply that $R^{n\times n}$ satisfies all the properties of a ring except for commutativity of multiplication.

Some simple but useful facts to keep in mind are the following:

- If $A \in R^{m\times n}$ and $B \in R^{n\times p}$, then the ith row of AB is equal to vB, where $v = \mathrm{Row}_i(A)$; also, the kth column of AB is equal to Aw, where $w = \mathrm{Col}_k(B)$.

- If $A \in R^{m \times n}$ and $v = (c_1, \ldots, c_m) \in R^{1 \times m}$, then
$$vA = \sum_{i=1}^{m} c_i \operatorname{Row}_i(A).$$
In words: vA is a linear combination of the rows of A, with coefficients taken from the corresponding entries of v.
- If $A \in R^{m \times n}$ and
$$w = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix} \in R^{n \times 1},$$
then
$$Aw = \sum_{j=1}^{n} d_j \operatorname{Col}_j(A).$$
In words: Aw is a linear combination of the columns of A, with coefficients taken from the corresponding entries of w.

If $A \in R^{m \times n}$, the **transpose** of A, denoted by $A^{\top}$, is defined to be the $n \times m$ matrix whose (j, i) entry is $A(i, j)$.

Theorem 14.3. *If $A, B \in R^{m \times n}$, $C \in R^{n \times p}$, and $c \in R$, then:*

(i) $(A + B)^{\top} = A^{\top} + B^{\top}$;

(ii) $(cA)^{\top} = cA^{\top}$;

(iii) $(A^{\top})^{\top} = A$;

(iv) $(AC)^{\top} = C^{\top} A^{\top}$.

Proof. Exercise. □

If A_i is an $n_i \times n_{i+1}$ matrix, for $i = 1, \ldots, k$, then by associativity of matrix multiplication, we may write the product matrix $A_1 \cdots A_k$, which is an $n_1 \times n_{k+1}$ matrix, without any ambiguity.

For an $n \times n$ matrix A, and a positive integer k, we write A^k to denote the product $A \cdots A$, where there are k terms in the product. Note that $A^1 = A$. We may extend this notation to $k = 0$, defining A^0 to be the $n \times n$ identity matrix. One may readily verify the usual rules of exponent arithmetic: for all non-negative integers k, ℓ, we have
$$(A^{\ell})^k = A^{k\ell} = (A^k)^{\ell} \text{ and } A^k A^{\ell} = A^{k+\ell}.$$

It is easy also to see that part (iv) of Theorem 14.3 implies that for all non-negative integers k, we have

$$(A^k)^\top = (A^\top)^k.$$

Algorithmic issues

For computational purposes, matrices are represented in the obvious way as arrays of elements of R. As remarked at the beginning of this chapter, we shall treat R as an "abstract data type," and not worry about how elements of R are actually represented; in discussing the complexity of algorithms, we shall simply count "operations in R," by which we mean additions, subtractions, and multiplications; we shall sometimes also include equality testing and computing multiplicative inverses as "operations in R." In any real implementation, there will be other costs, such as incrementing counters, and so on, which we may safely ignore, as long as their number is at most proportional to the number of operations in R.

The following statements are easy to verify:

- We can multiply an $m \times n$ matrix by a scalar using mn operations in R.
- We can add two $m \times n$ matrices using mn operations in R.
- We can compute the product of an $m \times n$ matrix and an $n \times p$ matrix using $O(mnp)$ operations in R.

It is also easy to see that given an $n \times n$ matrix A, and a non-negative integer e, we can adapt the repeated squaring algorithm discussed in §3.4 so as to compute A^e using $O(\text{len}(e))$ multiplications of $n \times n$ matrices, and hence $O(\text{len}(e)n^3)$ operations in R.

EXERCISE 14.1. Let $A \in R^{m\times n}$. Show that if $vA = 0_R^{1\times n}$ for all $v \in R^{1\times m}$, then $A = 0_R^{m\times n}$.

14.2 Matrices and linear maps

Let R be a ring.

For positive integers m and n, consider the R-modules $R^{1\times m}$ and $R^{1\times n}$. If A is an $m \times n$ matrix over R, then the map

$$\begin{aligned} \lambda_A : \quad R^{1\times m} &\to R^{1\times n} \\ v &\mapsto vA \end{aligned}$$

is easily seen to be an R-linear map—this follows immediately from parts (ii) and (iv) of Theorem 14.2. We call λ_A the **linear map corresponding to** A.

If $v = (c_1, \ldots, c_m) \in R^{1\times m}$, then

$$\lambda_A(v) = vA = \sum_{i=1}^{m} c_i \operatorname{Row}_i(A).$$

From this, it is clear that

- the image of λ_A is the submodule of $R^{1\times n}$ spanned by $\{\operatorname{Row}_i(A)\}_{i=1}^m$; in particular, λ_A is surjective if and only if $\{\operatorname{Row}_i(A)\}_{i=1}^m$ spans $R^{1\times n}$;
- λ_A is injective if and only if $\{\operatorname{Row}_i(A)\}_{i=1}^m$ is linearly independent.

There is a close connection between matrix multiplication and composition of corresponding linear maps. Specifically, let $A \in R^{m\times n}$ and $B \in R^{n\times p}$, and consider the corresponding linear maps $\lambda_A : R^{1\times m} \to R^{1\times n}$ and $\lambda_B : R^{1\times n} \to R^{1\times p}$. Then we have

$$\lambda_B \circ \lambda_A = \lambda_{AB}. \tag{14.1}$$

This follows immediately from the associativity of matrix multiplication.

We have seen how vector/matrix multiplication defines a linear map. Conversely, we shall now see that the action of any R-linear map can be viewed as a vector/matrix multiplication, provided the R-modules involved have bases (which will always be the case for finite dimensional vector spaces).

Let M be an R-module, and suppose that $\mathcal{S} = \{\alpha_i\}_{i=1}^m$ is a basis for M, where $m > 0$. As we know (see Theorem 13.14), every element $\alpha \in M$ can be written uniquely as $c_1\alpha_1 + \cdots + c_m\alpha_m$, where the c_i's are in R. Let us define

$$\operatorname{Vec}_{\mathcal{S}}(\alpha) := (c_1, \ldots, c_m) \in R^{1\times m}.$$

We call $\operatorname{Vec}_{\mathcal{S}}(\alpha)$ the **coordinate vector of α relative to $\mathcal{S}$**. The function

$$\operatorname{Vec}_{\mathcal{S}} : M \to R^{1\times m}$$

is an R-module isomorphism (it is the inverse of the isomorphism ε in Theorem 13.14).

Let N be another R-module, and suppose that $\mathcal{T} = \{\beta_j\}_{j=1}^n$ is a basis for N, where $n > 0$. Just as in the previous paragraph, every element $\beta \in N$ has a unique coordinate vector $\operatorname{Vec}_{\mathcal{T}}(\beta) \in R^{1\times n}$ relative to $\mathcal{T}$.

Now let $\rho : M \to N$ be an arbitrary R-linear map. Our goal is to define a matrix $A \in R^{m\times n}$ with the following property:

$$\operatorname{Vec}_{\mathcal{T}}(\rho(\alpha)) = \operatorname{Vec}_{\mathcal{S}}(\alpha)A \ \text{ for all } \alpha \in M. \tag{14.2}$$

In words: if we multiply the coordinate vector of α on the right by A, we get the coordinate vector of $\rho(\alpha)$.

Constructing such a matrix A is easy: we define A to be the matrix whose ith row, for $i = 1, \dots, m$, is the coordinate vector of $\rho(\alpha_i)$ relative to $\mathcal{T}$. That is,

$$\mathrm{Row}_i(A) = \mathrm{Vec}_{\mathcal{T}}(\rho(\alpha_i)) \text{ for } i = 1, \dots, m.$$

Then for an arbitrary $\alpha \in M$, if $(c_1, \dots, c_m)$ is the coordinate vector of α relative to $\mathcal{S}$, we have

$$\rho(\alpha) = \rho\Big(\sum_i c_i\alpha_i\Big) = \sum_i c_i\rho(\alpha_i)$$

and so

$$\mathrm{Vec}_{\mathcal{T}}(\rho(\alpha)) = \sum_i c_i \,\mathrm{Vec}_{\mathcal{T}}(\rho(\alpha_i)) = \sum_i c_i \,\mathrm{Row}_i(A) = \mathrm{Vec}_{\mathcal{S}}(\alpha)A.$$

Furthermore, A is the only matrix satisfying (14.2). Indeed, if A' also satisfies (14.2), then subtracting, we obtain

$$\mathrm{Vec}_{\mathcal{S}}(\alpha)(A - A') = 0_R^{1\times n}$$

for all $\alpha \in M$. Since the map $\mathrm{Vec}_{\mathcal{S}} : M \to R^{1\times m}$ is surjective, this means that $v(A - A')$ is zero for all $v \in R^{1\times m}$, and from this, it is clear (see Exercise 14.1) that $A - A'$ is the zero matrix, and so $A = A'$.

We call the unique matrix A satisfying (14.2) the **matrix of ρ relative to $\mathcal{S}$ and $\mathcal{T}$**, and denote it by $\mathrm{Mat}_{\mathcal{S},\mathcal{T}}(\rho)$.

Recall that $\mathrm{Hom}_R(M, N)$ is the R-module consisting of all R-linear maps from M to N (see Theorem 13.12). We can view $\mathrm{Mat}_{\mathcal{S},\mathcal{T}}$ as a function mapping elements of $\mathrm{Hom}_R(M, N)$ to elements of $R^{m\times n}$.

Theorem 14.4. *The function* $\mathrm{Mat}_{\mathcal{S},\mathcal{T}} : \mathrm{Hom}_R(M, N) \to R^{m\times n}$ *is an R-module isomorphism. In particular, for every $A \in R^{m\times n}$, the pre-image of A under* $\mathrm{Mat}_{\mathcal{S},\mathcal{T}}$ *is* $\mathrm{Vec}_{\mathcal{T}}^{-1} \circ \lambda_A \circ \mathrm{Vec}_{\mathcal{S}}$, *where $\lambda_A : R^{1\times m} \to R^{1\times n}$ is the linear map corresponding to A.*

Proof. To show that $\mathrm{Mat}_{\mathcal{S},\mathcal{T}}$ is an R-linear map, let $\rho, \rho' \in \mathrm{Hom}_R(M, N)$, and let $c \in R$. Also, let $A := \mathrm{Mat}_{\mathcal{S},\mathcal{T}}(\rho)$ and $A' := \mathrm{Mat}_{\mathcal{S},\mathcal{T}}(\rho')$. Then for all $\alpha \in M$, we have

$$\begin{aligned}\mathrm{Vec}_{\mathcal{T}}((\rho + \rho')(\alpha)) &= \mathrm{Vec}_{\mathcal{T}}(\rho(\alpha) + \rho'(\alpha)) = \mathrm{Vec}_{\mathcal{T}}(\rho(\alpha)) + \mathrm{Vec}_{\mathcal{T}}(\rho'(\alpha))\\ &= \mathrm{Vec}_{\mathcal{S}}(\alpha)A + \mathrm{Vec}_{\mathcal{S}}(\alpha)A' = \mathrm{Vec}_{\mathcal{S}}(\alpha)(A + A').\end{aligned}$$

As this holds for all $\alpha \in M$, and since the matrix of a linear map is uniquely determined, we must have $\mathrm{Mat}_{\mathcal{S},\mathcal{T}}(\rho + \rho') = A + A'$. A similar argument shows that $\mathrm{Mat}_{\mathcal{S},\mathcal{T}}(c\rho) = cA$. This shows that $\mathrm{Mat}_{\mathcal{S},\mathcal{T}}$ is an R-linear map.

To show that the map $\mathrm{Mat}_{\mathcal{S},\mathcal{T}}$ is injective, it suffices to show that its kernel is trivial. If ρ is in the kernel of this map, then setting $A := 0_R^{m\times n}$ in (14.2), we see that $\mathrm{Vec}_{\mathcal{T}}(\rho(\alpha))$ is zero for all $\alpha \in M$. But since the map $\mathrm{Vec}_{\mathcal{T}} : N \to R^{1\times n}$ is injective, this implies $\rho(\alpha)$ is zero for all $\alpha \in M$. Thus, ρ must be the zero map.

To show surjectivity, we show that every $A \in R^{m\times n}$ has a pre-image under $\mathrm{Mat}_{\mathcal{S},\mathcal{T}}$ as described in the statement of the theorem. So let A be an $m \times n$ matrix, and let $\rho := \mathrm{Vec}_{\mathcal{T}}^{-1} \circ \lambda_A \circ \mathrm{Vec}_{\mathcal{S}}$. Again, since the matrix of a linear map is uniquely determined, it suffices to show that (14.2) holds for this particular A and ρ. For every $\alpha \in M$, we have

$$\begin{aligned}\mathrm{Vec}_{\mathcal{T}}(\rho(\alpha)) &= \mathrm{Vec}_{\mathcal{T}}(\mathrm{Vec}_{\mathcal{T}}^{-1}(\lambda_A(\mathrm{Vec}_{\mathcal{S}}(\alpha)))) = \lambda_A(\mathrm{Vec}_{\mathcal{S}}(\alpha))\\ &= \mathrm{Vec}_{\mathcal{S}}(\alpha)A.\end{aligned}$$

That proves the theorem. □

As a special case of the above, suppose that $M = R^{1\times m}$ and $N = R^{1\times n}$, and $\mathcal{S}$ and $\mathcal{T}$ are the standard bases for M and N (see Example 13.27). In this case, the functions $\mathrm{Vec}_{\mathcal{S}}$ and $\mathrm{Vec}_{\mathcal{T}}$ are the identity maps, and the previous theorem implies that the function

$$\begin{aligned}\Lambda : \quad R^{m\times n} &\to \mathrm{Hom}_R(R^{1\times m}, R^{1\times n})\\ A &\mapsto \lambda_A\end{aligned}$$

is the inverse of the function $\mathrm{Mat}_{\mathcal{S},\mathcal{T}} : \mathrm{Hom}_R(R^{1\times m}, R^{1\times n}) \to R^{m\times n}$. Thus, the function Λ is also an R-module isomorphism.

To summarize, we see that an R-linear map ρ from M to N, together with particular bases for M and N, uniquely determine a matrix A such that the action of multiplication on the right by A implements the action of ρ with respect to the given bases. There may be many bases for M and N to choose from, and different choices will in general lead to different matrices. Also, note that in general, a basis may be indexed by an arbitrary finite set; however, in defining coordinate vectors and matrices of linear maps, the index set must be ordered in some way. In any case, from a computational perspective, the matrix A gives us an efficient way to compute the map ρ, assuming elements of M and N are represented as coordinate vectors with respect to the given bases.

We have taken a "row-centric" point of view. Of course, if one prefers, by simply transposing everything, one can equally well take a "column-centric" point of view, where the action of ρ corresponds to multiplication of a column vector on the left by a matrix.

Example 14.1. Consider the quotient ring $E = R[X]/(f)$, where $f \in R[X]$ with $\deg(f) = \ell > 0$ and $\mathrm{lc}(f) \in R^*$. Let $\xi := [X]_f \in E$. As an R-module, E has a basis $\mathcal{S} := \{\xi^{i-1}\}_{i=1}^{\ell}$ (see Example 13.30). Let $\rho : E \to E$ be the ξ-multiplication map, which sends $\alpha \in E$ to $\xi\alpha \in E$. This is an R-linear map. If

$f = c_0 + c_1X + \cdots + c_{\ell-1}X^{\ell-1} + c_\ell X^\ell$, then the matrix of ρ relative to $\mathcal{S}$ is the $\ell \times \ell$ matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \\ -c_0/c_\ell & -c_1/c_\ell & -c_2/c_\ell & \cdots & -c_{\ell-1}/c_\ell \end{pmatrix},$$

where for $i = 1, \ldots, \ell - 1$, the ith row of A contains a 1 in position $i + 1$, and is zero everywhere else. The matrix A is called the **companion matrix of** f. □

Example 14.2. Let $x_1, \ldots, x_k \in R$. Let $R[X]_{<k}$ be the set of polynomials $g \in R[X]$ with $\deg(g) < k$, which is an R-module with a basis $\mathcal{S} := \{X^{i-1}\}_{i=1}^k$ (see Example 13.29). The multi-point evaluation map

$$\begin{aligned} \rho : \quad R[X]_{<k} &\to R^{1\times k} \\ g &\mapsto (g(x_1), \ldots, g(x_k)) \end{aligned}$$

is an R-linear map. Let $\mathcal{T}$ be the standard basis for $R^{1\times k}$. Then the matrix of ρ relative to $\mathcal{S}$ and $\mathcal{T}$ is the $k \times k$ matrix

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_k \\ x_1^2 & x_2^2 & \cdots & x_k^2 \\ \vdots & \vdots & & \vdots \\ x_1^{k-1} & x_2^{k-1} & \cdots & x_k^{k-1} \end{pmatrix}.$$

The matrix A is called a **Vandermonde matrix**. □

EXERCISE 14.2. Let $\sigma : M \to N$ and $\tau : N \to P$ be R-linear maps, and suppose that M, N, and P have bases $\mathcal{S}$, $\mathcal{T}$, and $\mathcal{U}$, respectively. Show that

$$\mathrm{Mat}_{\mathcal{S},\mathcal{U}}(\tau \circ \sigma) = \mathrm{Mat}_{\mathcal{S},\mathcal{T}}(\sigma) \cdot \mathrm{Mat}_{\mathcal{T},\mathcal{U}}(\tau).$$

EXERCISE 14.3. Let V be a vector space over a field F with basis $\mathcal{S} = \{\alpha_i\}_{i=1}^m$. Suppose that U is a subspace of V of dimension $\ell < m$. Show that there exists a matrix $A \in F^{m\times(m-\ell)}$ such that for all $\alpha \in V$, we have $\alpha \in U$ if and only if $\mathrm{Vec}_{\mathcal{S}}(\alpha)A$ is zero. Such a matrix A is called a **parity check matrix** for U.

EXERCISE 14.4. Let F be a finite field, and let A be a non-zero $m \times n$ matrix over F. Suppose one chooses a vector $v \in F^{1\times m}$ at random. Show that the probability that vA is the zero vector is at most $1/|F|$.

EXERCISE 14.5. Design and analyze a probabilistic algorithm that takes as input matrices $A, B, C \in \mathbb{Z}_p^{m\times m}$, where p is a prime. The algorithm should run in time $O(m^2 \operatorname{len}(p)^2)$ and should output either "yes" or "no" so that the following holds:

- if $C = AB$, then the algorithm should always output "yes";
- if $C \neq AB$, then the algorithm should output "no" with probability at least 0.999.

14.3 The inverse of a matrix

Let R be a ring.

For a square matrix $A \in R^{n\times n}$, we call a matrix $B \in R^{n\times n}$ an **inverse** of A if $BA = AB = I$, where I is the $n \times n$ identity matrix. It is easy to see that if A has an inverse, then the inverse is unique: if B and C are inverses of A, then

$$B = BI = B(AC) = (BA)C = IC = C.$$

Because the inverse of A is uniquely determined, we denote it by A^{-1}. If A has an inverse, we say that A is **invertible**, or **non-singular**. If A is not invertible, it is sometimes called **singular**. We will use the terms "invertible" and "not invertible." Observe that A is the inverse of A^{-1}; that is, $(A^{-1})^{-1} = A$.

If A and B are invertible $n \times n$ matrices, then so is their product: in fact, it is easily verified that $(AB)^{-1} = B^{-1}A^{-1}$. It follows that if A is an invertible matrix, and k is a non-negative integer, then A^k is invertible with inverse $(A^{-1})^k$, which we also denote by A^{-k}.

It is also easy to see that A is invertible if and only if the transposed matrix $A^\top$ is invertible, in which case $(A^\top)^{-1} = (A^{-1})^\top$. Indeed, $AB = I = BA$ holds if and only if $B^\top A^\top = I = A^\top B^\top$.

We now develop a connection between invertible matrices and R-module isomorphisms. Recall from the previous section the R-module isomorphism

$$\begin{aligned} \Lambda : \quad R^{n\times n} &\to \operatorname{Hom}_R(R^{1\times n}, R^{1\times n}) \\ A &\mapsto \lambda_A, \end{aligned}$$

where for each $A \in R^{n\times n}$, λ_A is the corresponding R-linear map

$$\begin{aligned} \lambda_A : \quad R^{1\times n} &\to R^{1\times n} \\ v &\mapsto vA. \end{aligned}$$

Evidently, λ_I is the identity map.

Theorem 14.5. *Let $A \in R^{n\times n}$, and let $\lambda_A : R^{1\times n} \to R^{1\times n}$ be the corresponding R-linear map. Then A is invertible if and only if λ_A is bijective, in which case $\lambda_{A^{-1}} = \lambda_A^{-1}$.*

Proof. Suppose A is invertible, and that B is its inverse. We have $AB = BA = I$, and hence $\lambda_{AB} = \lambda_{BA} = \lambda_I$, from which it follows (see (14.1)) that $\lambda_B \circ \lambda_A = \lambda_A \circ \lambda_B = \lambda_I$. Since λ_I is the identity map, this implies λ_A is bijective.

Suppose λ_A is bijective. We know that the inverse map λ_A^{-1} is also an R-linear map, and since the mapping Λ above is surjective, we have $\lambda_A^{-1} = \lambda_B$ for some $B \in R^{n \times n}$. Therefore, we have $\lambda_B \circ \lambda_A = \lambda_A \circ \lambda_B = \lambda_I$, and hence (again, see (14.1)) $\lambda_{AB} = \lambda_{BA} = \lambda_I$. Since the mapping Λ is injective, it follows that $AB = BA = I$. This implies A is invertible, with $A^{-1} = B$. □

We also have:

Theorem 14.6. *Let $A \in R^{n \times n}$. The following are equivalent:*

(i) A is invertible;

(ii) $\{\mathrm{Row}_i(A)\}_{i=1}^n$ is a basis for $R^{1 \times n}$;

(iii) $\{\mathrm{Col}_j(A)\}_{j=1}^n$ is a basis for $R^{n \times 1}$.

Proof. We first prove the equivalence of (i) and (ii). By the previous theorem, A is invertible if and only if λ_A is bijective. Also, in the previous section, we observed that λ_A is surjective if and only if $\{\mathrm{Row}_i(A)\}_{i=1}^n$ spans $R^{1 \times n}$, and that λ_A is injective if and only if $\{\mathrm{Row}_i(A)\}_{i=1}^n$ is linearly independent.

The equivalence of (i) and (iii) follows by considering the transpose of A. □

EXERCISE 14.6. Let R be a ring, and let A be a square matrix over R. Let us call B a **left inverse** of A if $BA = I$, and let us call C a **right inverse** of A if $AC = I$.

(a) Show that if A has both a left inverse B and a right inverse C, then $B = C$ and hence A is invertible.

(b) Assume that R is a field. Show that if A has either a left inverse or a right inverse, then A is invertible.

Note that part (b) of the previous exercise holds for arbitrary rings, but the proof of this is non-trivial, and requires the development of the theory of determinants, which we do not cover in this text.

EXERCISE 14.7. Show that if A and B are two square matrices over a field such that their product AB is invertible, then both A and B themselves must be invertible.

EXERCISE 14.8. Show that if A is a square matrix over an arbitrary ring, and A^k is invertible for some $k > 0$, then A is invertible.

EXERCISE 14.9. With notation as in Example 14.1, show that the matrix A is invertible if and only if $c_0 \in R^*$.

EXERCISE 14.10. With notation as in Example 14.2, show that the matrix A is invertible if and only if $x_i - x_j \in R^*$ for all $i \neq j$.

14.4 Gaussian elimination

Throughout this section, F denotes a field.

A matrix $B \in F^{m \times n}$ is said to be in **reduced row echelon form** if there exists a sequence of integers $(p_1, \ldots, p_r)$, with $0 \leq r \leq m$ and $1 \leq p_1 < p_2 < \cdots < p_r \leq n$, such that the following holds:

- for $i = 1, \ldots, r$, all of the entries in row i of B to the left of entry (i, p_i) are zero; that is, $B(i, j) = 0_F$ for $j = 1, \ldots, p_i - 1$;
- for $i = 1, \ldots, r$, all of the entries in column p_i of B above entry (i, p_i) are zero; that is, $B(i', p_i) = 0_F$ for $i' = 1, \ldots, i - 1$;
- for $i = 1, \ldots, r$, we have $B(i, p_i) = 1_F$;
- all entries in rows $r + 1, \ldots, m$ of B are zero; that is, $B(i, j) = 0_F$ for $i = r + 1, \ldots, m$ and $j = 1, \ldots, n$.

It is easy to see that if B is in reduced row echelon form, then the sequence $(p_1, \ldots, p_r)$ above is uniquely determined, and we call it the **pivot sequence** of B. Several further remarks are in order:

- All of the entries of B are completely determined by the pivot sequence, except for the entries (i, j) with $1 \leq i \leq r$ and $j > p_i$ with $j \notin \{p_{i+1}, \ldots, p_r\}$, which may be arbitrary.
- If B is an $n \times n$ matrix in reduced row echelon form whose pivot sequence is of length n, then B must be the $n \times n$ identity matrix.
- We allow for an empty pivot sequence (i.e., $r = 0$), which will be the case precisely when $B = 0_F^{m \times n}$.

Example 14.3. The following 4×6 matrix B over the rational numbers is in reduced row echelon form:

$$B = \begin{pmatrix} 0 & 1 & -2 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The pivot sequence of B is $(2, 4, 5)$. Notice that the first three rows of B form a linearly independent family of vectors, that columns 2, 4, and 5 form a linearly independent family of vectors, and that all of other columns of B are linear combinations of columns 2, 4, and 5. Indeed, if we truncate the pivot columns to their first three rows, we get the 3×3 identity matrix. □

Generalizing the previous example, if a matrix is in reduced row echelon form, it is easy to deduce the following properties, which turn out to be quite useful:

Theorem 14.7. *If B is a matrix in reduced row echelon form with pivot sequence $(p_1, \dots, p_r)$, then:*

(i) rows $1, 2, \dots, r$ of B form a linearly independent family of vectors;

(ii) columns $p_1, \dots, p_r$ of B form a linearly independent family of vectors, and all other columns of B can be expressed as linear combinations of columns $p_1, \dots, p_r$.

Proof. Exercise—just look at the matrix! □

Gaussian elimination is an algorithm that transforms a given matrix $A \in F^{m \times n}$ into a matrix $B \in F^{m \times n}$, where B is in reduced row echelon form, and is obtained from A by a sequence of **elementary row operations**. There are three types of elementary row operations:

Type I: swap two rows;

Type II: multiply a row by a non-zero scalar;

Type III: add a scalar multiple of one row to a different row.

The application of any specific elementary row operation to an $m \times n$ matrix C can be affected by multiplying C on the left by a suitable $m \times m$ matrix X. Indeed, the matrix X corresponding to a particular elementary row operation is simply the matrix obtained by applying the same elementary row operation to the $m \times m$ identity matrix. It is easy to see that for every elementary row operation, the corresponding matrix X is invertible.

We now describe the basic version of Gaussian elimination. The input is an $m \times n$ matrix A, and the algorithm is described in Fig. 14.1.

The algorithm works as follows. First, it makes a copy B of A (this is not necessary if the original matrix A is not needed afterwards). The algorithm proceeds column by column, starting with the left-most column, so that after processing column j, the first j columns of B are in reduced row echelon form, and the current value of r represents the length of the pivot sequence. To process column j, in steps 3–6 the algorithm first searches for a non-zero element among $B(r+1, j), \dots, B(m, j)$; if none is found, then the first $j + 1$ columns of B are already in reduced row echelon form. Otherwise, one of these non-zero elements is selected as the **pivot element** (the choice is arbitrary), which is then used in steps 8–13 to bring column j into the required form. After incrementing r, the pivot element is brought into position (r, j), using a Type I operation in step 9. Then the entry (r, j) is set to 1_F, using a Type II operation in step 10. Finally, all the entries above and below entry (r, j) are set to 0_F, using Type III operations in steps 11–13. Note that because

```
$B \leftarrow A$, $r \leftarrow 0$
for $j \leftarrow 1$ to $n$ do
    $\ell \leftarrow 0$, $i \leftarrow r$
    while $\ell = 0$ and $i \leq m$ do
        $i \leftarrow i + 1$
        if $B(i, j) \neq 0_F$ then $\ell \leftarrow i$
    if $\ell \neq 0$ then
        $r \leftarrow r + 1$
        swap rows $r$ and $\ell$ of $B$
        $\mathrm{Row}_r(B) \leftarrow B(r, j)^{-1}\,\mathrm{Row}_r(B)$
        for $i \leftarrow 1$ to $m$ do
            if $i \neq r$ then
                $\mathrm{Row}_i(B) \leftarrow \mathrm{Row}_i(B) - B(i, j)\,\mathrm{Row}_r(B)$
output $B$
```

Fig. 14.1. Gaussian elimination

columns $1, \ldots, j-1$ of B were already in reduced row echelon form, none of these operations changes any values in these columns.

As for the complexity of the algorithm, it is easy to see that it performs $O(mn)$ elementary row operations, each of which takes $O(n)$ operations in F, so a total of $O(mn^2)$ operations in F.

Example 14.4. Consider the execution of the Gaussian elimination algorithm on input

$$A = \begin{pmatrix} [0] & [1] & [1] \\ [2] & [1] & [2] \\ [2] & [2] & [0] \end{pmatrix} \in \mathbb{Z}_3^{3\times 3}.$$

After copying A into B, the algorithm transforms B as follows:

$$\begin{pmatrix} [0] & [1] & [1] \\ [2] & [1] & [2] \\ [2] & [2] & [0] \end{pmatrix} \xrightarrow{\mathrm{Row}_1 \leftrightarrow \mathrm{Row}_2} \begin{pmatrix} [2] & [1] & [2] \\ [0] & [1] & [1] \\ [2] & [2] & [0] \end{pmatrix} \xrightarrow{\mathrm{Row}_1 \leftarrow [2]\,\mathrm{Row}_1} \begin{pmatrix} [1] & [2] & [1] \\ [0] & [1] & [1] \\ [2] & [2] & [0] \end{pmatrix}$$

$$\xrightarrow{\mathrm{Row}_3 \leftarrow \mathrm{Row}_3 - [2]\,\mathrm{Row}_1} \begin{pmatrix} [1] & [2] & [1] \\ [0] & [1] & [1] \\ [0] & [1] & [1] \end{pmatrix} \xrightarrow{\mathrm{Row}_1 \leftarrow \mathrm{Row}_1 - [2]\,\mathrm{Row}_2} \begin{pmatrix} [1] & [0] & [2] \\ [0] & [1] & [1] \\ [0] & [1] & [1] \end{pmatrix}$$

$$\xrightarrow{\mathrm{Row}_3 \leftarrow \mathrm{Row}_3 - \mathrm{Row}_2} \begin{pmatrix} [1] & [0] & [2] \\ [0] & [1] & [1] \\ [0] & [0] & [0] \end{pmatrix}. \quad \square$$

Suppose the Gaussian elimination algorithm performs a total of t elementary row operations. Then as discussed above, the application of the eth elementary row operation, for $e = 1, \ldots, t$, amounts to multiplying the current value of the matrix B on the left by a particular invertible $m \times m$ matrix X_e. Therefore, the final output value of B satisfies the equation

$$B = XA \text{ where } X = X_t X_{t-1} \cdots X_1.$$

Since the product of invertible matrices is also invertible, we see that X itself is invertible.

Although the algorithm as presented does not compute the matrix X, it can be easily modified to do so. The resulting algorithm, which we call **extended Gaussian elimination**, is the same as plain Gaussian elimination, except that we initialize the matrix X to be the $m \times m$ identity matrix, and we add the following steps:

- just before step 9: swap rows r and ℓ of X;
- just before step 10: $\mathrm{Row}_r(X) \leftarrow B(r, j)^{-1}\,\mathrm{Row}_r(X)$;
- just before step 13: $\mathrm{Row}_i(X) \leftarrow \mathrm{Row}_i(X) - B(i, j)\,\mathrm{Row}_r(X)$.

At the end of the algorithm we output X in addition to B.

So we simply perform the same elementary row operations on X that we perform on B. The reader may verify that the above algorithm is correct, and that it uses $O(mn(m + n))$ operations in F.

Example 14.5. Continuing with Example 14.4, the execution of the extended Gaussian elimination algorithm initializes X to the identity matrix, and then transforms X as follows:

$$\begin{pmatrix} [1] & [0] & [0] \\ [0] & [1] & [0] \\ [0] & [0] & [1] \end{pmatrix} \xrightarrow{\mathrm{Row}_1 \leftrightarrow \mathrm{Row}_2} \begin{pmatrix} [0] & [1] & [0] \\ [1] & [0] & [0] \\ [0] & [0] & [1] \end{pmatrix} \xrightarrow{\mathrm{Row}_1 \leftarrow [2]\,\mathrm{Row}_1} \begin{pmatrix} [0] & [2] & [0] \\ [1] & [0] & [0] \\ [0] & [0] & [1] \end{pmatrix}$$

$$\xrightarrow{\mathrm{Row}_3 \leftarrow \mathrm{Row}_3 - [2]\,\mathrm{Row}_1} \begin{pmatrix} [0] & [2] & [0] \\ [1] & [0] & [0] \\ [0] & [2] & [1] \end{pmatrix} \xrightarrow{\mathrm{Row}_1 \leftarrow \mathrm{Row}_1 - [2]\,\mathrm{Row}_2} \begin{pmatrix} [1] & [2] & [0] \\ [1] & [0] & [0] \\ [0] & [2] & [1] \end{pmatrix}$$

$$\xrightarrow{\mathrm{Row}_3 \leftarrow \mathrm{Row}_3 - \mathrm{Row}_2} \begin{pmatrix} [1] & [2] & [0] \\ [1] & [0] & [0] \\ [2] & [2] & [1] \end{pmatrix}. \qquad \square$$

EXERCISE 14.11. For each type of elementary row operation, describe the matrix X which corresponds to it, as well as X^{-1}.

EXERCISE 14.12. Given a matrix $B \in F^{m \times n}$ in reduced row echelon form, show how to compute its pivot sequence using $O(n)$ operations in F.

EXERCISE 14.13. In §4.4, we saw how to speed up matrix multiplication over $\mathbb{Z}$ using the Chinese remainder theorem. In this exercise, you are to do the same, but for performing Gaussian elimination over $\mathbb{Z}_p$, where p is a large prime. Suppose you are given an $m \times m$ matrix A over $\mathbb{Z}_p$, where $\mathrm{len}(p) = \Theta(m)$. Straightforward application of Gaussian elimination would require $O(m^3)$ operations in $\mathbb{Z}_p$, each of which takes time $O(m^2)$, leading to a total running time of $O(m^5)$. Show how to use the techniques of §4.4 to reduce the running time of Gaussian elimination to $O(m^4)$.

14.5 Applications of Gaussian elimination

Throughout this section, A is an arbitrary $m \times n$ matrix over a field F, and $XA = B$, where X is an invertible $m \times m$ matrix, and B is an $m \times n$ matrix in reduced row echelon form with pivot sequence $(p_1, \ldots, p_r)$. This is precisely the information produced by the extended Gaussian elimination algorithm, given A as input (the pivot sequence can easily be "read" directly from B—see Exercise 14.12). Also, let

$$\begin{aligned} \lambda_A : \quad F^{1 \times m} &\to F^{1 \times n} \\ v &\mapsto vA \end{aligned}$$

be the linear map corresponding to A.

Computing the image and kernel

Consider first the **row space** of A, by which we mean the subspace of $F^{1 \times n}$ spanned by $\{\mathrm{Row}_i(A)\}_{i=1}^m$, or equivalently, the image of λ_A.

We claim that the row space of A is the same as the row space of B. To see this, note that since $B = XA$, for every $v \in F^{1 \times m}$, we have $vB = v(XA) = (vX)A$, and so the row space of B is contained in the row space of A. For the other containment, note that since X is invertible, we can write $A = X^{-1}B$, and apply the same argument.

Further, note that the row space of B, and hence that of A, clearly has dimension r. Indeed, as stated in Theorem 14.7, rows $1, \ldots, r$ of B form a basis for the row space of B.

Consider next the kernel K of λ_A, or what we might call the **row null space** of A. We claim that $\{\mathrm{Row}_i(X)\}_{i=r+1}^{m}$ is a basis for K. Clearly, just from the fact that $XA = B$ and the fact that rows $r+1, \ldots, m$ of B are zero, it follows that rows $r+1, \ldots, m$ of X are contained in K. Furthermore, as X is invertible, $\{\mathrm{Row}_i(X)\}_{i=1}^{m}$ is a basis for $F^{1\times m}$ (see Theorem 14.6). Thus, the family of vectors $\{\mathrm{Row}_i(X)\}_{i=r+1}^{m}$ is linearly independent and spans a subspace K' of K. It suffices to show that $K' = K$. Suppose to the contrary that $K' \subsetneq K$, and let $v \in K \setminus K'$. As $\{\mathrm{Row}_i(X)\}_{i=1}^{m}$ spans $F^{1\times m}$, we may write $v = \sum_{i=1}^{m} c_i \,\mathrm{Row}_i(X)$; moreover, as $v \notin K'$, we must have $c_i \neq 0_F$ for some $i = 1, \ldots, r$. Setting $\tilde{v} := (c_1, \ldots, c_m)$, we see that $v = \tilde{v}X$, and so

$$\lambda_A(v) = vA = (\tilde{v}X)A = \tilde{v}(XA) = \tilde{v}B.$$

Furthermore, since $\{\mathrm{Row}_i(B)\}_{i=1}^{r}$ is linearly independent, rows $r+1, \ldots, m$ of B are zero, and $\tilde{v}$ has a non-zero entry in one of its first r positions, we see that $\tilde{v}B$ is not the zero vector. We have derived a contradiction, and hence may conclude that $K' = K$.

Finally, note that if $m = n$, then A is invertible if and only if its row space has dimension m, which holds if and only if $r = m$, and in the latter case, B is the identity matrix, and hence X is the inverse of A.

Let us summarize the above discussion:

- The first r rows of B form a basis for the row space of A (i.e., the image of λ_A).
- The last $m - r$ rows of X form a basis for the row null space of A (i.e., the kernel of λ_A).
- If $m = n$, then A is invertible (i.e., λ_A is an isomorphism) if and only if $r = m$, in which case X is the inverse of A (i.e., the matrix of λ_A^{-1} relative to the standard basis).

So we see that from the output of the extended Gaussian elimination algorithm, we can simply "read off" bases for both the image and the kernel, as well as the inverse (if it exists), of a linear map represented as a matrix with respect to given bases. Also note that this procedure provides a "constructive" version of Theorem 13.28.

Example 14.6. Continuing with Examples 14.4 and 14.5, we see that the vectors $([1], [0], [2])$ and $([0], [1], [1])$ form a basis for the row space of A, while the vector $([2], [2], [1])$ is a basis for the row null space of A. □

Solving systems of linear equations

Suppose that in addition to the matrix A, we are given $w \in F^{1\times n}$, and want to find a solution $v \in F^{1\times m}$ (or perhaps describe all solutions), to the equation

$$vA = w. \tag{14.3}$$

Equivalently, we can phrase the problem as finding an element (or describing all elements) of the set $\lambda_A^{-1}(\{w\})$.

Now, if there exists a solution at all, say $v \in F^{1\times m}$, then $\lambda_A(v) = \lambda_A(v')$ if and only if $v \equiv v' \pmod{K}$, where K is the kernel of λ_A. It follows that the set of all solutions to (14.3) is $v + K = \{v + v_0 : v_0 \in K\}$. Thus, given a basis for K and any solution v to (14.3), we have a complete and concise description of the set of solutions to (14.3).

As we have discussed above, the last $m - r$ rows of X form a basis for K, so it suffices to determine if $w \in \operatorname{Im} \lambda_A$, and if so, determine a single pre-image v of w.

Also as we discussed, $\operatorname{Im} \lambda_A$, that is, the row space of A, is equal to the row space of B, and because of the special form of B, we can quickly and easily determine if the given w is in the row space of B, as follows. By definition, w is in the row space of B if and only if there exists a vector $\bar{v} \in F^{1\times m}$ such that $\bar{v}B = w$. We may as well assume that all but the first r entries of $\bar{v}$ are zero. Moreover, $\bar{v}B = w$ implies that for $i = 1, \ldots, r$, the ith entry of $\bar{v}$ is equal to the p_ith entry of w. Thus, the vector $\bar{v}$, if it exists, is completely determined by the entries of w at positions $p_1, \ldots, p_r$. We can construct $\bar{v}$ satisfying these conditions, and then test if $\bar{v}B = w$. If not, then we may conclude that (14.3) has no solutions; otherwise, setting $v := \bar{v}X$, we see that $vA = (\bar{v}X)A = \bar{v}(XA) = \bar{v}B = w$, and so v is a solution to (14.3).

One easily verifies that if we implement the above procedure as an algorithm, the work done in addition to running the extended Gaussian elimination algorithm amounts to $O(m(n + m))$ operations in F.

A special case of the above procedure is when $m = n$ and A is invertible, in which case (14.3) has a unique solution, namely, $v := wX$, since in this case, $X = A^{-1}$.

The rank of a matrix

We define the **row rank** of A to be the dimension of its row space, which is equal to $\dim_F(\operatorname{Im} \lambda_A)$. The **column space** of A is defined as the subspace of $F^{m\times 1}$ spanned by $\{\mathrm{Col}_j(A)\}_{j=1}^{n}$; that is, the column space of A is $\{Az : z \in F^{n\times 1}\}$. The **column rank** of A is the dimension of its column space.

Now, the column space of A need not be the same as the column space of B, but from the identity $B = XA$, and the fact that X is invertible, it easily follows that these two subspaces are isomorphic (via the map that sends $y \in F^{m\times 1}$ to Xy), and hence have the same dimension. Moreover, by Theorem 14.7, the column rank of B is r, which is the same as the row rank of A.

So we may conclude: *The column rank and row rank of A are the same.*

Because of this, we may define the **rank** of a matrix to be the common value of its row and column rank.

The orthogonal complement of a subspace

So as to give equal treatment to rows and columns, one can also define the **column null space** of A to be the kernel of the linear map defined by multiplication on the left by A; that is, the column null space of A is $\{z \in F^{n\times 1} : Az = 0_F^{m\times 1}\}$. By applying the results above to the transpose of A, we see that the column null space of A has dimension $n - r$, where r is the rank of A.

Let $U \subseteq F^{1\times n}$ be the row space of A, and let $U^\perp \subseteq F^{1\times n}$ denote the set of all vectors $\bar{u} \in F^{1\times n}$ whose transpose $\bar{u}^\top$ belongs to the column null space of A. Now, U is a subspace of $F^{1\times n}$ of dimension r and $U^\perp$ is a subspace of $F^{1\times n}$ of dimension $n - r$. The space $U^\perp$ consists precisely of all vectors $\bar{u} \in F^{1\times n}$ that are "orthogonal" to all vectors $u \in U$, in the sense that the "inner product" $u\bar{u}^\top$ is zero. For this reason, $U^\perp$ is sometimes called the "orthogonal complement of U."

Clearly, $U^\perp$ is determined by the subspace U itself, and does not depend on the particular choice of matrix A. It is also easy to see that the orthogonal complement of $U^\perp$ is U; that is, $(U^\perp)^\perp = U$. This follows immediately from the fact that $U \subseteq (U^\perp)^\perp$ and $\dim_F((U^\perp)^\perp) = n - \dim_F(U^\perp) = \dim_F(U)$.

Now suppose that $U \cap U^\perp = \{0\}$. Then by Theorem 13.11, we have an isomorphism of $U \times U^\perp$ with $U + U^\perp$, and since $U \times U^\perp$ has dimension n, it must be the case that $U + U^\perp = F^{1\times n}$. It follows that every element of $F^{1\times n}$ can be expressed uniquely as $u + \bar{u}$, where $u \in U$ and $\bar{u} \in U^\perp$.

We emphasize that the observations in the previous paragraph hinged on the assumption that $U \cap U^\perp = \{0\}$, which itself holds provided U contains no non-zero "self-orthogonal vectors" u such that $uu^\top$ is zero. If F is the field of real numbers, then of course there are no non-zero self-orthogonal vectors, since $uu^\top$ is the sum of the squares of the entries of u. However, for other fields, there may very well be non-zero self-orthogonal vectors. As an example, if $F = \mathbb{Z}_2$, then any vector u with an even number of 1-entries is self orthogonal.

So we see that while much of the theory of vector spaces and matrices carries over without change from familiar ground fields, like the real numbers, to arbitrary ground fields F, not everything does. In particular, the usual decomposition of a vector space into a subspace and its orthogonal complement breaks down, as does any other procedure that relies on properties specific to "inner product spaces."

For the following three exercises, as above, A is an arbitrary $m \times n$ matrix over a field F, and $XA = B$, where X is an invertible $m \times m$ matrix, and B is in reduced row echelon form.

EXERCISE 14.14. Show that the column null space of A is the same as the column null space of B.

EXERCISE 14.15. Show how to compute a basis for the column null space of A using $O(r(n-r))$ operations in F, given A and B.

EXERCISE 14.16. Show that the matrix B is uniquely determined by A; more precisely, show that if $X'A = B'$, where X' is an invertible $m \times m$ matrix, and B' is in reduced row echelon form, then $B' = B$.

In the following two exercises, the theory of determinants could be used; however, they can all be solved directly, without too much difficulty, using just the ideas developed so far in the text.

EXERCISE 14.17. Let p be a prime. A matrix $A \in \mathbb{Z}^{m \times m}$ is called **invertible modulo** p if there exists a matrix $B \in \mathbb{Z}^{m \times m}$ such that $AB \equiv BA \equiv I \pmod{p}$, where I is the $m \times m$ integer identity matrix. Here, two matrices are considered congruent with respect to a given modulus if their corresponding entries are congruent. Show that A is invertible modulo p if and only if A is invertible over $\mathbb{Q}$, and the entries of A^{-1} lie in $\mathbb{Q}^{(p)}$ (see Example 7.26).

EXERCISE 14.18. You are given a matrix $A \in \mathbb{Z}^{m \times m}$ and a prime p such that A is invertible modulo p (see previous exercise). Suppose that you are also given $w \in \mathbb{Z}^{1 \times m}$.

(a) Show how to efficiently compute a vector $v \in \mathbb{Z}^{1 \times m}$ such that $vA \equiv w \pmod{p}$, and that v is uniquely determined modulo p.

(b) Given a vector v as in part (a), along with an integer $e \geq 1$, show how to efficiently compute $\hat{v} \in \mathbb{Z}^{1 \times m}$ such that $\hat{v}A \equiv w \pmod{p^e}$, and that $\hat{v}$ is uniquely determined modulo p^e. Hint: mimic the "lifting" procedure discussed in §12.5.2.

(c) Using parts (a) and (b), design and analyze an efficient algorithm that takes the matrix A and the prime p as input, together with a bound H on the absolute value of the numerator and denominator of the entries of the vector v' that is the unique (rational) solution to the equation $v'A = w$. Your algorithm should run in time polynomial in the length of H, the length of p, and the sum of the lengths of the entries of A and w. Hint: use rational reconstruction, but be sure to fully justify its application.

Note that in the previous exercise, one can use the theory of determinants to derive good bounds, in terms of the lengths of the entries of A and w, on the size of the least prime p such that A is invertible modulo p (assuming A is invertible over the rationals), and on the length of the numerator and denominator of the entries of rational solution v' to the equation $v'A = w$. The interested reader who is familiar with the basic theory of determinants is encouraged to establish such bounds.

The next two exercises illustrate how Gaussian elimination can be adapted, in certain cases, to work in rings that are not necessarily fields. Let R be an arbitrary ring. A matrix $B \in R^{m\times n}$ is said to be in **row echelon form** if there exists a pivot sequence $(p_1, \ldots, p_r)$, with $0 \le r \le m$ and $1 \le p_1 < p_2 < \cdots < p_r \le n$, such that the following holds:

- for $i = 1, \ldots, r$, all of the entries in row i of B to the left of entry (i, p_i) are zero;
- for $i = 1, \ldots, r$, we have $B(i, p_i) \neq 0_R$;
- all entries in rows $r + 1, \ldots, m$ of B are zero.

EXERCISE 14.19. Let R be the ring $\mathbb{Z}_{p^e}$, where p is prime and $e > 1$. Let $\pi := [p] \in R$. The goal of this exercise is to develop an efficient algorithm for the following problem: given a matrix $A \in R^{m\times n}$, with $m > n$, find a vector $v \in R^{1\times m}$ such that $vA = 0_R^{1\times n}$ but $v \notin \pi R^{1\times m}$.

(a) Show how to modify the extended Gaussian elimination algorithm to solve the following problem: given a matrix $A \in R^{m\times n}$, compute $X \in R^{m\times m}$ and $B \in R^{m\times n}$, such that $XA = B$, X is invertible, and B is in row echelon form. Your algorithm should run in time $O(mn(m + n)e^2 \operatorname{len}(p)^2)$. Assume that the input includes the values p and e. Hint: when choosing a pivot element, select one divisible by a minimal power of π; as in ordinary Gaussian elimination, your algorithm should only use elementary row operations to transform the input matrix.

(b) Using the fact that the matrix X computed in part (a) is invertible, argue that none of its rows belong to $\pi R^{1\times m}$.

(c) Argue that if $m > n$ and the matrix B computed in part (a) has pivot sequence $(p_1, \ldots, p_r)$, then $m - r > 0$ and if v is any one of the last $m - r$ rows of X, then $vA = 0_R^{1\times n}$.

(d) Give an example that shows that $\{\mathrm{Row}_i(B)\}_{i=1}^{r}$ need not be linearly independent, and that $\{\mathrm{Row}_i(X)\}_{i=r+1}^{m}$ need not span the kernel of the linear map λ_A corresponding to A.

EXERCISE 14.20. Let R be the ring $\mathbb{Z}_\ell$, where $\ell > 1$ is an integer. You are given a matrix $A \in R^{m\times n}$. Show how to efficiently compute $X \in R^{m\times m}$ and $B \in R^{m\times n}$

such that $XA = B$, X is invertible, and B is in row echelon form. Your algorithm should run in time $O(mn(m+n)\operatorname{len}(\ell)^2)$. Hint: to zero-out entries, you should use "rotations"—for integers a, b, d, s, t with

$$d = \gcd(a, b) \neq 0 \text{ and } as + bt = d,$$

and for row indices r, i, a rotation simultaneously updates rows r and i of a matrix C as follows:

$$(\operatorname{Row}_r(C), \operatorname{Row}_i(C)) \leftarrow (s \operatorname{Row}_r(C) + t \operatorname{Row}_i(C), -\frac{b}{d} \operatorname{Row}_r(C) + \frac{a}{d} \operatorname{Row}_i(C));$$

observe that if $C(r, j) = [a]_\ell$ and $C(i, j) = [b]_\ell$ before applying the rotation, then $C(r, j) = [d]_\ell$ and $C(i, j) = [0]_\ell$ after the rotation.

EXERCISE 14.21. Consider again the setting in Exercise 14.3. Show that $A \in F^{m\times(m-\ell)}$ is a parity check matrix for U if and only if $\{\operatorname{Col}_j(A)^\top\}_{i=1}^{m-\ell}$ is a basis for the orthogonal complement of $\operatorname{Vec}_S(U) \subseteq F^{1\times m}$.

EXERCISE 14.22. Let $\{v_i\}_{i=1}^n$ be a family of vectors, where $v_i \in \mathbb{R}^{1\times\ell}$ for each $i = 1, \ldots, n$. We say that $\{v_i\}_{i=1}^n$ is **pairwise orthogonal** if $v_i v_j^\top = 0$ for all $i \neq j$. Show that every pairwise orthogonal family of non-zero vectors over $\mathbb{R}$ is linearly independent.

EXERCISE 14.23. The purpose of this exercise is to use linear algebra to prove that any pairwise independent family of hash functions (see §8.7) must contain a large number of hash functions. More precisely, let $\{\Phi_r\}_{r\in R}$ be a pairwise independent family of hash functions from S to T, with $|T| \geq 2$. Our goal is to show that $|R| \geq |S|$. Let $n := |S|$, and $m := |T|$, and $\ell := |R|$. Write $R = \{r_1, \ldots, r_\ell\}$ and $S = \{s_1, \ldots, s_n\}$. Without loss of generality, we may assume that T is a set of non-zero real numbers that sum to zero (e.g., $T = \{1, \ldots, m-1, -m(m-1)/2\}$). Now define the matrix $A \in \mathbb{R}^{n\times\ell}$ with $A(i, j) := \Phi_{r_j}(s_i)$. Show that $\{\operatorname{Row}_i(A)\}_{i=1}^n$ is a pairwise orthogonal family of non-zero vectors (see previous exercise). From this, deduce that $\ell \geq n$.

14.6 Notes

While a trivial application of the defining formulas yields a simple algorithm for multiplying two $n \times n$ matrices over a ring R that uses $O(n^3)$ operations in R, this algorithm is not the best, asymptotically speaking. The currently fastest algorithm for this problem, due to Coppersmith and Winograd [28], uses $O(n^\omega)$ operations in R, where $\omega < 2.376$. We note, however, that the good old $O(n^3)$ algorithm is still the only one used in almost any practical setting.