Mathematics 5150
University of Colorado
January, 2002
Professor Taylor

Here are a few comments that expand on what the book has to say in sections 0.1.3–8 (pp. 2–4), 0.2.2–3 (pp. 5–6), 0.3.3–6 (pp. 8–12), some parts of 0.4–5 (pp. 12–14), and 1.0–1.1 (pp. 33–35).

All of this material is an integral part of the syllabus, both as theoretical material (proofs) and in any practical context where it may arise. The only exception is the theoretical development of the determinant, which we will not pursue in detail. In some sense this is review from an undergraduate course. Nevertheless, it is hoped that a succinct treatment will be of some help. (In fact an undergraduate course often makes *very* hard work out of the material through Corollary 9. For example, the book by Lay used at CU takes 234 pages to get this far. No wonder people are freaked out by undergraduate linear algebra! Believe me, this way is easier.)

Spanning and linear independence.

Let V be a vector space. The *span* of a subset X of V is the set of all linear combinations of elements of X:

$$\operatorname{Span} X \;=\; \{\sum_{i=1}^{n} \lambda_i x_i \;:\; \text{any } n, \text{ any scalars } \lambda_i, \text{ any } x_i \in X\}.$$

(Convention: if $n = 0$, then the linear combination here is taken to be 0.)

Lemma 1. $\operatorname{Span} X$ *is the smallest subspace of* V *that contains* X*. In other words,* $\operatorname{Span} X$ *is a subset of* V*,* $X \subseteq \operatorname{Span} X$*, and if* W *is any subspace with* $X \subseteq W$*, then* $\operatorname{Span} X \subseteq W$*.* ∎

Now a subset X of V is said to be *(linearly) dependent* iff a relation of the form

$$\sum_{i=1}^{n} \lambda_i x_i \;=\; 0,$$

holds for some $n \geq 1$, some $x_i \in X$, and some scalars λ_i ($1 \leq i \leq n$), with not all $\lambda_i = 0$. Then X is said to be *(linearly) independent* iff it is not dependent.

Lemma 2. *X is dependent iff some $x \in X$ satisfies*

$$x \;\in\; \operatorname{Span}(X - \{x\}).$$

By a *basis* of V we mean an independent subset of V that spans V. V is called *finite-dimensional* iff there exists a finite X such that $V = \operatorname{Span} X$.

Theorem 3. *Every finite-dimensional space has (at least one) basis. In fact every minimal spanning set is a basis.*

Proof. Let $w_1, \ldots, w_m$ span V, so that no subset of these vectors spans V. If this set were dependent, then by Lemma 2 we would have some w_j in the span of

$$w_1, \ldots, w_{j-1}, w_{j+1}, \ldots, w_m.$$

Hence this proper subset would also span V, in contradiction to the minimality of $\{w_1, \ldots, w_m\}$. ∎

Remark. Provided that one has a way to decide whether a set of vectors spans V, Theorem 3 represents a practical way to obtain a basis of V. (One begins with some finite spanning set $W = \{w_1, \ldots, w_n\}$, which is provided by finite-dimensionality. Then one examines in turn each $S - \{w_j\}$. If one of these is a spanning set, one examines *its* $(n-2)$-element subsets, and so on. Finally one arrives at a spanning set, none of whose proper subsets spans V. By Theorem 3, this set is a basis.) In most familiar spaces, where the vectors are columns of real or complex numbers, the decision whether S spans V could be based on column reduction (see page 10 below). In that case, however, the column-reduction algorithm itself leads to a basis, so that the procedure mentioned here becomes unnecessary.

From now on, we will assume that V is finite-dimensional, and hence has a basis. This separate treatment of finite-dimensional spaces is traditional and natural for mathematics courses, for the following reasons:

- The purely algebraic (and combinatorial) problems encountered in vector spaces already occur for the finite-dimensional case. Moreover, the relevant algebraic methods and techniques already attain their full complexity and difficulty in the finite-dimensional case. (Linear combinations, determinants, eigenvalues, symmetry, normal forms, and so on.)
- These finite-dimensional methods could easily fill several semesters of study on their own.
- The infinite-dimensional case requires a more sophisticated approach (such as the use of Zorn's Lemma) even for elementary results such as the existence of a basis.
- With a few exceptions, the study of infinite-dimensional spaces (such as function spaces of every sort) naturally involves some topological considerations, i.e. a consideration of limits and continuity.

Corollary 4. *If S is a subset of V that spans V, then some subset of S is a basis of V.*

Lemma 5. *Suppose that $w_1, \ldots, w_n$ span V, and that $v_1, \ldots, v_m$ form an independent subset of V. Then $n \geq m$, and the w_i can be re-indexed so that*

$$v_1, \ldots, v_m, w_{m+1}, \ldots, w_n \tag{1}$$

spans V.

Proof. For fixed n, we prove the lemma by induction on m. For $m = 0$, the conclusions of the lemma are immediate. For $m \geq 1$, we will inductively apply the lemma (with m replaced by $m-1$) to the spanning vectors $w_1, \ldots, w_n$ and the independent vectors $v_1, \ldots, v_{m-1}$. For these vectors, the lemma tells us that $n \geq m-1$, and that the w_i can be re-indexed so that the vectors

$$v_1, \ldots, v_{m-1}, w_m, \ldots, w_n \tag{2}$$

span V. Therefore, v_m in particular is a linear combination of these vectors; that is,

$$v_m = \sum_{i=1}^{m-1} \lambda_i v_i + \sum_{j=m}^{n} \mu_j w_j \tag{3}$$

for some scalars $\lambda_1 \ldots \lambda_{m-1}$ and $\mu_m \ldots \mu_n$. By the linear independence of the v_i, we cannot have

$$v_m = \sum_{i=1}^{m-1} \lambda_i v_i,$$

and hence at least one of the μ_j must be non-zero. It is now immediate that $n \geq m$.

At this point we do some further re-indexing of the w_i: if $j \neq m$, we interchange w_j and w_m. Hence we now have $\mu_m \neq 0$. We now rewrite Equation (3) so as to bring w_m to the left-hand side:

$$w_m = -\sum_{i=1}^{m-1} \frac{\lambda_i}{\mu_m} v_i + \frac{1}{\mu_m} v_m - \sum_{i=m+1}^{n} \frac{\mu_j}{\mu_m} w_j. \tag{4}$$

From this we see that the span of the vectors in (1) contains w_m. Hence this span in fact contains all the vectors of (2), and hence, by induction, contains all of V. ∎

Remarks on the proof.

- The astute reader may have noticed this apparent paradox: we have fixed n; we have m going to infinity (as is implicit in an inductive argument); and we appear to be proving that $m \leq n$. Actually, all is well. The lemma and its proof are subtler than this — $m \leq n$ is proved only when there exist m independent vectors in V.
- Equation (4) contains precise formulas for scalars that will represent w_m as a linear combination of the vectors (1). For our theoretical purposes, these formulas are of little further interest — what one needs to remember is that (assuming always the important point that $\mu_m \neq 0$), one can manipulate the equations to isolate w_m as a linear combination of other vectors. On the other hand, if we were developing a computer system to analyze linear dependence and so on, then we would probably wish to include a precise form of Equation (4) or something tantamount to it.

Corollary 6. *If X is a finite set that spans V, and Y is a linearly independent subset of V, then X contains at least as many elements as Y. (In particular, Y is finite.)* ∎

Corollary 7. *Any two bases of V have the same number of elements.*

Proof. Immediate from two applications of Corollary 6. ∎

Thus there is a single number n that is the number of elements in any basis of V. This n is called the *dimension of V*, denoted also $\dim V$.

Corollary 8. *Suppose that $\dim V = m$. Then the bases of V are precisely the m-element spanning subsets of V.*

Proof. Clearly, every basis is a spanning set with m elements. Conversely, suppose X is an m-element spanning subset. By Corollary 4, X has a subset Y that is a basis. By Corollary 7, Y also has m elements; hence $X = Y$, and so X is a basis. ∎

The following corollary appears in the textbook as 0.1.6, page 3.

Corollary 9. *Every independent subset of V can be extended to be a basis of V.*

Proof. Suppose that $\{v_1 \dots v_m\}$ is independent. Let $w_1, \dots, w_n$ be some basis of V, where of course $n = \dim V$. Then, by Lemma 5, after re-indexing the w_j we have

$$v_1, \dots, v_m, w_{m+1}, \dots, w_n$$

spanning V. By Corollary 8, these vectors form a basis of V. ∎

Corollary 10. *Suppose that $\dim V = m$. Then the bases of V are precisely the m-element independent subsets of V.*

Proof. Clearly, every basis is an independent set with m elements. Conversely, suppose X is an m-element independent subset. By Corollary 9, X is a subset of some Y that is a basis. By Corollary 7, Y also has m elements; hence $X = Y$, and so X is a basis. ∎

Corollary 11. *If V is a finite-dimensional space, then every subspace W of V is finite-dimensional. Moreover, if W is a proper subset of V, then $\dim W < \dim V$.*

Proof. Suppose that $\dim V = n$. By Corollary 6, every linearly independent subset $X = \{w_1, \dots, w_m\}$ of W has $m \leq n$. Thus we may assume that we have such an $X \subseteq W$, with m as large as possible. We claim that X spans W. So, for arbitrary $w \in W$, we need to show that w lies in the span of the vectors w_i. The result is immediate for $w = 0$ and for $w =$ some w_j, so we exclude these cases from now on. Thus $X \cup \{w\}$ is an $(m+1)$-element subset of W, and hence is linearly dependent. In other words,

$$0 \; = \; \lambda w \; + \; \sum_{j=1}^{m} \lambda_j \, w_j, \tag{5}$$

for some scalars λ, λ_j, not all zero. By the linear independence of X, we have $\lambda \neq 0$. It is now obvious that Equation (5) may be solved for w as a linear combination of the w_i. Thus W has the finite spanning set X, and is hence finite-dimensional.

In fact X is actually a basis of W, and so

$$\dim W \; = \; m \; \leq \; n \; = \; \dim V.$$

Equality of dimensions here implies $W = V$ (either by Corollary 8 or by Corollary 10), and so the final assertion of the Corollary is also proved. ∎

As for the exact dimension of the subspace W appearing in Corollary 11, in most practical cases of interest W will be given to us in one of two ways. If $V = F^n$ and W is given as the span of a set of vectors in F^n, then the method of *column reduction* (see page 10 below) will lead to the value of $\dim W$. If $V = F^n$ and W is given as the kernel of a linear map $F^n \longrightarrow F^k$, then *row reduction* (see page 10 below) will lead to the value of $\dim W$.

Isomorphism of a space U with F^m.

It is our textbook's convention that, for a field F, the standard space F^n consists of all *columns* of n elements of F. Since columns are awkward in print, such a column is usually written (in this book) as $[x_1, x_2, \dots, x_n]^T$, i.e., as the *transpose* of a row. In other words, we write the *row* $[x_1, x_2, \ldots, x_n]$, and then append ^{T} to signal that we are really talking about a column. We will conform to this convention in these notes.

If U is any vector space and $u_1, \ldots, u_n \in U$, we define

$$\phi : F^n \longrightarrow U \tag{6}$$

via

$$\phi(x) \;=\; \sum_{j=1}^{n} x_j u_j, \tag{7}$$

for $x = [x_1, \ldots, x_n]^T \in F^n$. (This is the obvious linear combination of the basis elements u_j.) It is not hard to check that this map is a linear map of vector spaces. It is also not hard to check that ϕ maps onto U if and only if the vectors u_j span U, and that it is one-to-one if and only if the vectors u_j are linearly independent. We tend to consider only the case where the vectors u_j form a basis $\mathcal{B} = \mathcal{B}_U$. Obviously, in this case, the map ϕ is an isomorphism of vector spaces.

The book does not mention ϕ as such. However it does refer, first in 0.1.8, to the map

$$x \longmapsto [x]_{\mathcal{B}}, \tag{8}$$

where $\mathcal{B}$ refers to the basis $\{u_1, \ldots, u_1\}$. ($[x]_{\mathcal{B}}$ is the (uniquely determined) column of scalars that will represent x as a linear combination of the basis vectors in $\mathcal{B}$.) This is the *inverse* of ϕ.

Representation of a linear transformation by a matrix.

Let U and V be vector spaces over F of dimension m and n, respectively, and let us be given a linear transformation

$$T : U \longrightarrow V. \tag{9}$$

We wish to find a matrix representation of T. (The matrix will be defined in (11) below. The correctness of the representation is the equality of (12) and (14) below).

Select a basis $\mathcal{B}_U = \{u_1, \ldots, u_n\}$ for U and a basis $\mathcal{B}_V = \{v_1, \ldots, v_m\}$ for V. Just as was the case for U, the basis $\mathcal{B}_V$ determines an isomorphism of V with F^m, which we will denote by the same letter ϕ. We are looking for a matrix A such that the following diagram commutes:

$$\begin{array}{ccc} U & \xrightarrow{\;T\;} & V \\ {\scriptstyle\phi}\big\uparrow{\scriptstyle\cong} & & {\scriptstyle\cong}\big\uparrow{\scriptstyle\phi} \\ F^n & \xrightarrow[\;A\;]{} & F^m \end{array} \tag{10}$$

(In other words, we wish to have $T\phi(x) = \phi(Ax)$ for every vector $x \in F^n$.) Here Ax denotes the result of multiplying the vector $x = [x_1, \ldots, x_n]^T$ by the $m \times n$ matrix A to obtain a vector $y = [y_1, \ldots, y_m]^T$. The well-known formula for this multiplication is $y_i = \sum_j a_{ij} x_j$ (which appears again in (15) below.)

Definition of the matrix A associated to T. For each basis-element u_j, we know that $T(u_j) \in V$, and hence $T(u_j)$ can be expressed uniquely according to the basis $\mathcal{B}_V$:

$$T(u_j) \;=\; \sum_i a_{ij} v_i \tag{11}$$

(Incidentally, Equation (11) corresponds to the textbooks' equation $[y]_{\mathcal{B}_V} = A[x]_{\mathcal{B}_U}$ for $y = T(x)$, in 0.2.2, page 5.) We take Equation (11) as the definition of the matrix $A = [a_{ij}]$. We now calculate, for $x = [x_1, \ldots, x_n]^T \in F^n$,

$$T\phi(x) \;=\; T(\sum_j x_j u_j) \;=\; \sum_j x_j T(u_j) \tag{12}$$

$$=\; \sum_j x_j \sum_i a_{ij} v_i \;=\; \sum_i \sum_j a_{ij} x_j v_j \tag{13}$$

$$=\; \sum_i y_i v_i \;=\; \phi(Ax), \tag{14}$$

where, for each i,

$$y_i \;=\; \sum_j a_{ij} x_j. \tag{15}$$

(Notice that the second equality in (12) is where we invoke the linearity of T.) The commutativity of the diagram is now established.

We have represented the arbitrary linear map T as $\phi A \phi^{-1}$ for a matrix A. To say it in other words, the linear transformation $\phi^{-1} T \phi$ reduces to multiplication by a matrix A.

Since the maps ϕ and ϕ^{-1} are well understood and present no challenge, we may as well study the more concrete numerical object A. This is what we shall do, almost exclusively, from now on. (This is essentially the first assertion in the book's section 0.2.3: "There is no loss of generality in associating an n-dimensional vector space over F with F^n, ...")

Here are two exercises to test your understanding of Equation (11) in perhaps unfamiliar circumstances.

Exercise 1. Let V be the space of all infinitely differentiable functions that satisfy the fourth-order differential equation

$$y'''' \;-\; 4y''' \;+\; 6y'' \;-\; 4y' \;+\; y \;=\; 0.$$

Let T denote the differential operator that takes each function to its own derivative: $T : y \mapsto y'$. Prove that T maps V into itself.

Next, perform calculations establishing that each of the functions $u_1 = e^x$, $u_2 = xe^x$, $u_3 = x^2e^x$ and $u_4 = x^3e^x$ belongs to V. In fact, these form a basis of V, but (for now) you are not asked to prove this. Also let v_1, ..., v_4 denote the same basis: $v_i = u_i$ for $i = 1, \ldots, 4$.

Under these bases, what is the 4×4 matrix A that is defined for T by Equation (11)?

Exercise 2. Suppose that $T : V \longrightarrow V$ is a linear self-map of V, and that W is a subspace of V. We say that W is a T-*invariant subspace* of V iff $T(W) \subseteq W$, in other words iff T maps W to itself, i.e. $T : W \longrightarrow W$. Then (under a suitable basis) $T : W \longrightarrow W$ has a representation as a $k \times k$ matrix, where $k = \dim W$; in this exercise we evaluate one such matrix.

(i) Suppose that $V = \boldsymbol{R}^4$, $U = \boldsymbol{R}^2$ and that

$$
T = \begin{bmatrix} 0 & 1 & 1 & -5 \\ 3 & 1 & 1 & 5 \\ 1 & -1 & 2 & 2 \\ 2 & 4 & 3 & 3 \end{bmatrix}, \qquad J = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 0 \end{bmatrix}.
$$

Define W to be the kernel of $J\colon V \longrightarrow U$. I save you the trouble of row-reducing J, by supplying the following basis for W:

$$
u_1 = \begin{bmatrix} 0 & -1 & 0 & 1 \end{bmatrix}^T, \quad u_2 = \begin{bmatrix} -2 & 1 & 1 & 0 \end{bmatrix}^T.
$$

Using this basis, prove that W is a T-invariant subspace of V. (For this, it is enough to show that $T(u_1)$ is a linear combination of u_1 and u_2, and that the same is true of $T(u_2)$.)

(ii) What is the matrix $[a_{ij}]$ of the map $T\colon W \longrightarrow W$, relative to the basis u_1, u_2 and the basis v_1, v_2, where $v_i = u_i$ for $i = 1, 2$? It should of course be a 2×2 matrix satisfying Equation (11). (Hint. There is not much to do here beyond what was done in Part (i). The 2×2 matrix that we are looking for simply records the coefficients that appear in the linear combinations found in Part (i). Notice that the 2×2 matrix obtained doesn't look much like the original matrix T.)

Exercise 3. (Optional) If $T\colon V \longrightarrow V$ and $W \subseteq V$ is defined as the kernel of a linear map $J\colon V \longrightarrow U$, here is a sufficient condition for W to be T-invariant:

(i) Prove that if there exists a linear map $B\colon U \longrightarrow U$ such that $JT = BJ$, then W is T-invariant.

(ii) Taking J and T as in the previous exercise, use Part (i) to re-establish the fact that $W = \ker J$ is T-invariant. (Hint. Use

$$
B = \begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix}.)
$$

Dependence on the choice of bases $\mathcal{B}_U$ and $\mathcal{B}_V$.

Suppose we imagine a second basis $\mathcal{B}'_U = \{u'_1, \dots u'_n\}$ for U. Then there are isomorphisms ϕ and ϕ', the first based on $\mathcal{B}_U$ (as above), and the other based similarly on $\mathcal{B}'_U$. We then have the diagram

(16)
$$
\begin{array}{ccccc}
U & \xrightarrow{\;1\;} & U & \xrightarrow{\;T\;} & V \\
\phi' \uparrow \cong & & \phi \uparrow \cong & & \cong \uparrow \phi \\
F^n & \xrightarrow[S]{} & F^n & \xrightarrow[A]{} & F^m
\end{array}
$$

where S is simply the non-singular matrix that represents the linear isomorphism $\phi\phi'^{-1}\colon F^n \longrightarrow F^n$. Here we have T represented by the matrix A (relative to the basis $\mathcal{B}_U$), and represented by the matrix AS (relative to the basis $\mathcal{B}'_U$). (N.b. The combined arrows

(17)
$$
F^n \xrightarrow[S]{} F^n \xrightarrow[A]{} F^m
$$

correspond to multiplication by the matrix AS.)

Similarly, a change in basis for V corresponds to modifying A to SA.

The flexibility in choice of bases is to be regarded as an opportunity. The idea is to select a basis so that the representing matrix is as nice as possible. A large part of the book (various so-called normal forms for matrices), involves a quest for such nice representations. Usually we do not allow ourselves to make independent selections of bases for U and for V. Usually $U = V$, and we must have the same basis for each. Then things become more difficult (see page 13 below). Nevertheless, there is one circumstance where one often adjusts $\mathcal{B}_U$ independently of $\mathcal{B}_V$, or vice versa, and that is in row- and column-reduction of matrices.

The column space.

Suppose A is an $n \times n$ matrix of elements of F. Then multiplication by A is of course a linear map from F^n to F^n; this map is often denoted by the same letter A:

$$\begin{aligned} A\colon F^n &\longrightarrow F^n \\ x &\mapsto Ax. \end{aligned}$$

Let e_i stand for the standard basis vector $[0, \ldots, 0, 1, 0, \ldots, 0]^T$ (with 1 in the i^{th} place). Using ordinary matrix multiplication, it is not hard to verify that *Ae_i is the i^{th} column of A.* (Always keep this fact in mind!) Since the vectors e_i $(1 \le i \le n)$ span F^n, the vectors Ae_i span the image of the map $A\colon F^n \longrightarrow F^n$. In other words, *the columns of A span the image of the map $A: F^n \longrightarrow F^n$.* Thus, this space is sometimes called the **column space of** A.

Spaces in general are somewhat difficult to analyze — e.g. to find a basis, learn the dimension, and so on. Luckily, for column spaces, there is an easy algorithm to find a basis and get the dimension. That is the topic of the next section.

Column reduction.

Consider three important types of square matrix: Continuing with A the matrix of $T : U \longrightarrow V$, we wish to consider the effect of multiplying A on the right by various special $n \times n$ matrices that we will now define. First, for $1 \le j \le n$, and λ any **non-zero** scalar, let

$$M_i(\lambda) = \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ 0 & 1 & & \cdots & & 0 \\ & & \ddots & & & \\ \vdots & & & \lambda & & \vdots \\ & & & & \ddots & \\ 0 & 0 & & \cdots & & 1 \end{bmatrix} \tag{18}$$

where λ occurs as the i^{th} diagonal element, and all other diagonal elements are 1. For any i, j with $i \neq j$ and $1 \leq i, j \leq n$, and any scalar μ, let

$$(19) \qquad A_{ij}(\mu) = \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ 0 & 1 & & \cdots & & 0 \\ & & \ddots & & \mu & \\ & & & & & \\ & & & & \ddots & \\ 0 & 0 & & \cdots & & 1 \end{bmatrix}$$

where the μ appears in the i^{th} row and the j^{th} column. (In other words, $A_{ij}(\mu)$ is like the identity matrix I_n except that it has μ in one off-diagonal position.) Finally, for any $i \neq j$, let

$$(20) \qquad P_{ij} = \begin{bmatrix} 1 & & & \cdots & & 0 \\ & \ddots & & & & \\ & & 0 & & 1 & \\ \vdots & & & \ddots & & \vdots \\ & & 1 & & 0 & \\ & & & & \ddots & \\ 0 & & & \cdots & & 1 \end{bmatrix}$$

In other words, P_{ij} is like the identity matrix I_n with its i^{th} and j^{th} rows interchanged. Now the reader may check that each of the matrices we have defined is invertible, in fact:

$$(21) \qquad M_i(\lambda)^{-1} = M_i(1/\lambda), \quad A_{ij}(\mu)^{-1} = A_{ij}(-\mu) \quad \text{and} \quad P_{ij}^{-1} = P_{ij}$$

Therefore, any product whatever of these special matrices is invertible.

Now, what happens if we multiply A on the right by a product S of these special matrices? By equation (17), since S is invertible, *multiplying A on the right by S does not change the column space.*

We can therefore use multiplications of A on the right by the special matrices as a means of analyzing the column space (otherwise known as the image of T). Let us call a matrix R *column-reduced* iff it has the following properties:

- The uppermost non-zero entry of each column is 1.
- (For each $j > 1$) the uppermost non-zero entry of column j lies at least one row below the uppermost non-zero entry of column $j - 1$.
- Any column consisting entirely of zeros lies to the right of any column that does not consist entirely of zeros.
- If the uppermost non-zero entry of some column occurs in row i, then every other entry of row i is zero.

For example,

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is column-reduced, but

$$\begin{bmatrix} 2 & 0 & 0 \\ 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 3 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 \\ 3 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

are not, because they violate the first, second and fourth properties just stated.

Now it is possible to convert any matrix into a column-reduced matrix by multiplying A on the right by some of the special matrices introduced above. There is in fact an algorithm that is straightforward (except for the problem of identifying zeros in a matrix!). We will not make a detailed investigation of the algorithm in this course, but the student should be able to reduce small matrices (say, 5×5 and smaller). One should especially note the theorem that *the column-reduced form of a matrix is uniquely determined*, regardless of the path (algorithm) used to arrive at it.

One important virtue of a column-reduced matrix is that the non-zero columns are automatically linearly independent, so one immediately can tell the dimension of the image space. This dimension of the image space is known as the **rank** of A.

If A is a non-singular $n \times n$ matrix, then its rank is n, and hence its column-reduced form has n non-zero columns; in other words it is the identity matrix. In other words, one arrives at the identity matrix by multiplying A on the right by a product of matrices $M_i(\lambda)$, $A_{ij}(\mu)$ and P_{ij}. In other words, A^{-1} has this form. Since these three special matrix-types are closed under the forming of inverses, we see in fact that *every non-singular square matrix is a product of matrices $M_i(\lambda)$, $A_{ij}(\mu)$ and P_{ij}.*

Exercise 4. Column reduce each of the two matrices

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 3 & 2 & 1 \\ 1 & 4 & 3 & 2 \end{bmatrix} \qquad \begin{bmatrix} 1 & -1 & 2 & 2 \\ 2 & -3 & 2 & 1 \\ 3 & -5 & 2 & -3 \\ -4 & 12 & 8 & 10 \end{bmatrix}$$

In each case, give a basis of the column space (range of the corresponding linear transformation).

Given an $m \times n$ matrix A, suppose that S is the non-singular matrix that effects column-reduction of A; in other words, that AS is column-reduced. (S can be found by keeping track of the column-reduction algorithm.) As we have said, the columns of A are linearly dependent if and only if AS has a column of zeros. This last condition is clearly equivalent to $As_n = 0$, where s_n is the last column of S. Thus, the last column contains the scalars for an explicit linear dependence relation between the columns of A:

$$0 = A_1 s_{1n} + \cdots + A_n s_{nn}.$$

Row reduction.

By multiplying on the left by the special matrices introduced above, we perform **row operations** on A. These are what the book calls *elementary operations* in 0.3.3. Any combination of row operations corresponds to multiplying A on the left by a non-singular matrix S. An algorithm analogous to column-reduction shows how to arrange things so that the final outcome SA is row-reduced. (In the textbook, a row-reduced matrix is said to be in *row-reduced echelon form* (RREF) — see 0.3.4 on page 10.) This is very important for the solving of linear equations. For example, if x is regarded as unknown, we may study the homogeneous linear equation

$$Ax = 0. \tag{22}$$

Clearly, for S non-singular, the solutions to (22) are exactly the same as the solutions to

$$SAx = 0. \tag{23}$$

If this last set of equations is row-reduced, then the solution is very easy. In class we will illustrate the simple algorithm (no computer required!) that leads from a row-reduced matrix (such as the SA here) to a basis of its null space.

Exercise 5. Row reduce each of the two 4×4 matrices that you previously column-reduced. For each of these two matrices, give a basis of the null space.

Row rank = column rank.

As before, let A be the matrix corresponding to $T : U \longrightarrow V$. By the **kernel** of T, we mean the subspace of U defined by

$$\text{Ker } T = \{u \in U : Tu = 0\}. \tag{24}$$

A synonym for kernel is **null space**. and by the **image** of T we mean the subspace of V defined by

$$\text{Im } T = \{v \in V : v = Tu \text{ for some } u \in U\}. \tag{25}$$

A synonym for image is **range**. Null space and range are briefly mentioned in the book's section 0.2.3.

Theorem 12. *Let $T : U \longrightarrow V$ be a linear map. If $u_1, \ldots, u_k$ form a basis of Ker T, and if T maps $w_1, \ldots, w_s$ bijectively to a basis of Im T, then*

$$u_1, \ldots, u_k,\ w_1, \ldots, w_s$$

form a basis of V.

Exercise 6. Prove (a) that the vectors displayed in Theorem 12 are linearly independent; (b) that they span U. (In doing the full exercise (parts (a) and (b)), you surely need four facts: the independence of the u_j, that of the vectors $T(w_j)$, the fact that the u_j span Ker T, and the fact that the vectors $T(w_j)$ span the image. If you don't use all four pieces, then your proof is bound to be incomplete at best.)

Corollary 13. *Let $T : U \longrightarrow V$ be a linear map.*

$$\dim U = \dim Im\ T + \dim Ker\ T. \tag{26}$$

We rewrite Equation (26) as

$$\dim U - \dim \text{Ker } T = \dim \text{Im } T, \tag{27}$$

It has been said above that the right-hand side of Equation (27) is the column rank of the matrix A, i.e., the number of non-zero columns of A after column reduction. It is also not hard to check that the left-hand side of Equation (27) is the row rank of the matrix A, i.e., the number of non-zero rows of A after row reduction.

Therefore, *the row rank of A is equal to*[1] *the column rank of A*. The **rank of** A refers to this common integer. It is also equal, of course, to either side of Equation (27). At this point, we have seen the equivalence of items (a), (e) and (g) of 0.4.4 (the textbook's section on rank).

The determinant of a square matrix.

The determinant of A is defined by these three laws:

(i) The determinant function is linear in each row; also in each column. (In short, the function is *multilinear*.)
(ii) The matrix P_{ij} defined above changes the determinant by a factor of -1
(iii) The determinant of the identity matrix is 1.

The textbook has these three axioms in 0.3.6.

From these axioms one first deduces from (ii) that if two rows or columns are identical, then the determinant is zero. The same holds, by (i), if one row is a scalar multiple of another row. From this and multilinearity, one then deduces that *Multiplication by $A_{ij}(\mu)$ does not change the determinant of any matrix.* Clearly, by (i), multiplication by $M_i(\lambda)$ multiplies the determinant by λ. We now know (step by step) the effect that row reduction has on the value of the determinant. A row-reduced square matrix is either the identity, in which case its determinant is 1 (by (iii)), or it contains a row of zeros, in which case its determinant is zero. Thus, step by step, all determinants can be calculated on the basis of Axioms (i)–(iii).

(More explicitly, one begins with the original square matrix A, together with the equation $D = 1$. One then proceeds to row- or column-reduce A. Each time one multiplies A by $M_i(\lambda)$ (resp. P_{ij}) the value of D is multiplied by $1/\lambda$ (resp. -1). If the reduced form of A turns out to be the identity matrix, then the determinant is given by the final value of D. Otherwise, the determinant is zero.)

Exercise 7. Calculate the determinant of

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

by row reduction. Also calculate the determinant of

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

by row reduction.

[1]In 0.4.1, page 12, the textbook calls this a "remarkable fact," but no proof is given.

It is also possible to derive the Laplace expansion (0.3.1), also known as the expansion by minors, from Axioms (i)–(iii) — see Exercise 9 below. The student should be aware of the Laplace expansion of the determinant (0.3.1), and the alternating-sum formula for the determinant (0.3.2).

Exercise 8. Calculate the two 4×4 determinants done previously, this time using the Laplace expansion.

Exercise 9. (Optional.) The Laplace expansion, inductive proof from the axioms. To begin with, let us simplify what is written on page 7 (0.3.1) to expansion along the first row:

$$\det A \;=\; \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j}, \tag{28}$$

where A_{1j} denotes the matrix the results from deleting the first row and the j^{th} column from A. For an inductive proof, we assume the truth of (28) for all $n \times n$ matrices. In particular, this implies that Axioms (i)–(iii) uniquely specify $\det A$ for all such A. The exercise here is to carry out the inductive step; in other words, using the $n \times n$ result, to prove (28) for all $(n+1) \times (n+1)$ matrices. [Hint: The key element is to prove that

$$\det \begin{bmatrix} 1 & 0 \\ 0 & B \end{bmatrix} \;=\; \det B \tag{29}$$

for any $n \times n$ matrix B. (This is a special case of (28); the full statement of (28) follows readily by multilinearity.) So how does one prove (29)? By uniqueness, once one checks that both sides of equation (29) satisfy Axioms (i)–(iii).]

Square matrices — change of basis.

Suppose that we have a linear transformation from a finite-dimensional space V into *itself*, $T : V \longrightarrow V$. According to the plans we have already worked out, this may be represented by taking V to be F^n, and by defining T as left multiplication by a matrix A:

$$F^n \xrightarrow{\;A\;} F^n. \tag{30}$$

Thus, the matrix of T is A, relative to the basis $e_1 = [1, 0, \ldots, 0]^T$, $e_2 = [0, 1, 0, \ldots, 0]^T$, $\ldots e_n = [0, \ldots, 0, 1]^T$. By Equation (11) above, this means that

$$T(e_j) \;=\; Ae_j \;=\; \sum_{i=1}^{n} a_{ij} e_i. \tag{31}$$

(Notice that this is the first place we have used the same basis both for elements of the domain (e_j plays this role in (31)) and for elements of this codomain (the e_i play this role in (31)). It is essential to do this if we wish to speak meaningfully of eigenvalues and eigenvectors.

Suppose that S is a non-singular $n \times n$ matrix. We wish to know the matrix representation of T relative to the basis $Se_1, \ldots, Se_n$. Thus let $B = [b_{ij}]$ be the matrix $S^{-1}AS$. Using $AS = SB$, we calculate

$$A(Se_j) = SBe_j = S(\sum_{i=1}^{n} b_{ij}e_i) \tag{32}$$

$$= \sum_{i=1}^{n} b_{ij}Se_i \tag{33}$$

These equations obviously tell us that $B = S^{-1}AS$ is the matrix of T relative to the new basis $Se_1, \ldots, Se_n$. (The textbook says this in 1.0.1, page 33.) The meaning of this for our work is twofold:

(i) Any property or characteristic of a matrix A that is really a property of the map $T : V \longrightarrow V$ — that is to say, not special to a choice of basis — must be invariant under the map $A \longmapsto S^{-1}AS$. (Example: A and $S^{-1}AS$ have the same characteristic polynomial, as we shall see.)

(ii) A frequent project is, for a given matrix A, to find S so that $B = S^{-1}AS$ is relatively simple (for example, triangular). This is no different than finding a different basis $Se_1, \ldots, Se_n$, relative to which the matrix has this simpler form. In fact, this could be seen as the central project of the course. The project begins in earnest in Theorem 1.3.9 on page 48, and gets into high gear with Schur's Theorem on page 79.

Exercise 10. Suppose

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}; \qquad S = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

What is the matrix B that represents A relative to the basis Se_1, Se_2?

`Further homework:`

`p. 37  1, 3, 4, 5, 6`

`p. 42  2, 3, 5, 7, 12 (optional)`

Here ends the review material.

A comment on eigenvalues. In this section I want to point a couple of things that can be read between the lines in the textbook, but which I believe are important enough to be stated clearly in their own right.

The most important single fact about eigenvalues and eigenvectors is this: when one has complex scalars, *every square matrix has at least one eigenvector*. The textbook does not proclaim this fact as thoroughly as I would like (the result seems to be buried in Exercise 7, page 37, and in 1.2.6, page 40).

The extension to finite-dimensional spaces is almost immediate: over the complex field, *every linear self-map of a finite-dimensional vector space W possesses at least one eigenvector in W* (regardless of whether the map happens to be given to us in the form of a matrix). The proof is simple: If $T : W \longrightarrow W$, with W n-dimensional, then there is an isomorphism $\phi : \boldsymbol{C}^n \longrightarrow W$ (see page 5) such that the composite map $\phi^{-1}T\phi : \boldsymbol{C}^n \longrightarrow \boldsymbol{C}^n$ is multiplication by a certain square matrix A. The matrix

A has an eigenvector v; clearly $\phi(v) \in W$ is an eigenvector of T. (This result — actually an unnecessarily special case — lies buried in the "Exercise" on page 51, with a hint toward our proof here.)

This result is needed, for example, in the proof of the interesting Lemma 1.3.17 on page 51 of the textbook.

Here is an exercise where one can go through the above proof with down-to-earth matrices.

Exercise 11. Consider $T : V \to V$, with $T = C^4$ and T given by the matrix of Exercise 2. Let W be the invariant subspace defined as $W = \ker J$, as in Exercise 2. Using the method outlined in the above proof, find an eigenvector of T in W. (Hint. The 2×2 matrix found in Part (ii) of Exercise 2 is obviously relevant here.) Your answer will of course be a non-zero element w of C^4. Don't forget to check these two things: (a) $Tw = \lambda w$ for some scalar λ, and (b) $w \in W$.

Nilpotent transformations and the Jordan form The textbook shows us how (beginning with Schur's Theorem) to begin with a matrix A and find a block-diagonal matrix such that each block B_j has the form $\lambda_j I + M_j$, where M_j is a nilpotent matrix. (The steps are recapitulated on page 122 of the textbook.) Finally the textbook proves Theorem 3.1.5 on page 123, which serves to analyze each individual nilpotent block. Here, we give an alternate statement and proof of that analysis.

Theorem 14. *Let* $f : \mathbf{V} \longrightarrow \mathbf{V}$ *be a linear transformation, where* $\dim \mathbf{V} = n$, *and suppose that* f *satisfies* $f^m = 0$ *for some* m. *Then there is a basis*

$$\left\{ v_i^j \,:\, 0 \le j < N;\;\; 0 \le i < m(j) \right\} \tag{34}$$

of $\mathbf{V}$, *such that*

$$f(v_{i+1}^j) = v_i^j \tag{35}$$

$$f(v_0^j) = 0 \tag{36}$$

for all $j < N$ *and all* $i < m(j) - 1$. *Conversely, if* $\mathbf{V}$ *has such a basis, then* $f^m = 0$ *for* m *the largest of the* $m(j)$.

(The integers $m(j)$ are arbitrary, subject to $\sum m(j) = \dim \mathbf{V}$ and $m(j) \le n$ for each j.)

Proof. The final assertion is easily proved. Thus we concentrate on the main assertion, namely, that if $f^m = 0$, then such a basis exists. We prove it by induction on the dimension of $\mathbf{V}$. If $\dim \mathbf{V} = 0$, then there is nothing to prove, because the empty basis is of the required type (with $N = 0$). On the other hand, if $\dim \mathbf{V} > 0$, then we consider the subspace $\mathbf{V}_0 = f[\mathbf{V}]$ of $\mathbf{V}$. Clearly, $\mathbf{V}_0$ has dimension less than that of $\mathbf{V}$, and f maps $\mathbf{V}_0$ into itself. By induction, $\mathbf{V}_0$ has a basis (34) that satisfies Equations (35) and (36). We extend this to a basis of $\mathbf{V}$ as follows. Each $v_{m(j)-1}^j$ lies in $\mathbf{V}_0 = f[\mathbf{V}]$; we therefore choose $v_{m(j)}^j$ so that $f(v_{m(j)}^j) = v_{m(j)-1}^j$ for each j. (This involves increasing each $m(j)$ to $m(j) + 1$.) It is now obvious that Equation (35) holds for all appropriate values of i and j, and that Equation (36) continues to hold for $j < N$.

The vectors $\{v_0^j : j < N\}$ form a linearly independent subset of $\mathbf{KER}f$. Thus, by Corollary 9, vectors $\{v_0^j : N \le j < M\}$ can be found to make $\{v_0^j : j < M\}$ a basis of $\mathbf{KER}f$. (We are thus increasing N to M, and taking $m(j) = 1$ for $N \le j < M$.) Then it is evident that Equation (36) holds for all $j < M$. It now remains only to show that

$$\{\, v_i^j \ : \ 0 \le j < M; \ \ 0 \le i < m(j) + 1 \,\}$$

is a basis of $\mathbf{V}$. Using Theorem 12, this follows directly from the facts that f maps

$$\{\, v_i^j \ : \ 0 \le j < N; \ \ 1 \le i < m(j) + 1 \,\}$$

one-to-one onto the basis (34) of $f[\mathbf{V}]$, and that

$$\{\, v_0^j \ : \ 0 \le j < M \,\}$$

is a basis of $\mathbf{KER}f$. ∎

Theorem 15. *Up to order of appearance, the sequence of integers $\langle m(j) \rangle$ is determined by f.*

Proof. The number of $m(j)$ that are greater than or equal to k is

$$\dim \mathbf{KER} f^k \ \ - \ \ \dim \mathbf{KER} f^{k-1}$$

Subtracting two such numbers obviously yields that number of $m(j)$ that are equal to k. ∎

In fact, we have proved that there exists a basis of $\mathbf{V}$ under which f has the matrix

$$T \ = \ \begin{bmatrix} 0 & a_1 & 0 & & 0 \\ 0 & 0 & a_2 & & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & & a_{n-1} \\ 0 & 0 & 0 & & 0 \end{bmatrix} \tag{37}$$

where each a_i is either 0 or 1. (Each block of r contiguous 1's among the entries a_i of T corresponds to $m(j) = r + 1$ for some j.) The converse for matrices, namely that every such matrix T satisfies $T^n = 0$, is almost obvious.

As we said, the exact pattern of 0's and 1's occurring among the a_i of (37) is determined by numbers $m(j)$ of Theorem 14; equivalently, by the number $M - N$ of vectors added to a basis of the kernel in each inductive step of that proof; equivalently, by the dimensions mentioned in the proof of Theorem 15. Probably the easiest method to discern this pattern from the start is to find the rank of each A^n and then refer to the proof of Theorem 15. One can also follow through the proof of Theorem 14, quite constructively, to find a basis of the type described there. (Some row and column reduction may be required!)

Exercise. Find a basis of the desired type for

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

Exhibit an S such that $S^{-1}AS$ has the form (37).