1. Basic Matrix Algebra.

Much of linear algebra over a field K can be deduced from the following basic lemma in which, for brevity, a matrix will be called *strongly regular* if it is a product of addition or permutation type elementary matrices. In particular, such a matrix is square and has an explicit left and right inverse. On the other hand, a matrix A will be called *singular* if it has a non-zero kernel $\mathcal{N}(A)$. Obviously these two properties exclude one another.

LEMMA 1.1: Let A be an $m \times n$ matrix over K. Then there exist strongly regular matrices M and N such that

$$MAN = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{where} \qquad D = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_r \end{bmatrix} \qquad \text{with} \qquad d_i \neq 0.$$

Proof: Let $\alpha(M, N)$ stand for the entry in the first row and first column of MAN. If $\alpha(M, N) = 0$ for all M, N, then obviously $A = 0$, and we are finished. Otherwise there is a pair M_1, N_1 such that

$$M_1 A N_1 = \begin{bmatrix} d_1 & X \\ Y & A' \end{bmatrix},$$

where A' is an $(m-1) \times (n-1)$-matrix, and $d_1 \neq 0$. Multiplying on the left by addition-type elementary matrices, we make $Y = 0$. Similarly, operating from the right, we modify N_1 to get $X = 0$. The proof is finished by induction.

THEOREM 1.2: Let A be an $m \times n$ matrix with $m \leq n$. Then A is non-singular if and only if it is square and invertible.

Proof: For invertible M, N it is easy to see that A is non-singular if and only if MAN is. If the latter is as above, non-singularity clearly means $r = m = n$. But then $MAN = D$ is an invertible diagonal matrix, and $A = M^{-1}DN^{-1}$ is invertible.

COROLLARY 1.3: An independent subset of the span of r vectors cannot have more than r elements.

Proof: Suppose $W_1, \ldots, W_s$ are in the span of $V_1, \ldots, V_r$; say $W_j = a_{1j}V_1 + \cdots + a_{rj}V_r$, for $j = 1, \ldots, s$. Consider the linear combination

$$x_1W_1 + \cdots + x_sW_s = (a_{11}x_1 + \cdots + a_{1s}x_s)V_1 + \cdots + (a_{r1}x_1 + \cdots + a_{rs}x_s)V_s.$$

If $s > r$, our theorem guarantees the existence of a non-trivial s-tuple $x_1, \ldots, x_s$ such that all this is zero, because the matrix (a_{ij}) involved here has more columns than rows, hence must be singular.

Note: Let $\mathcal{V}$ be a subspace of K^n. By the Corollary, any two bases of $\mathcal{V}$ have the same cardinality $\dim \mathcal{V}$. Moreover, any independent $\{W_1, \cdots, W_s\} \subset \mathcal{V}$ is contained in a basis of $\mathcal{V}$.

To see this, start with $W_1, \ldots, W_s$ and keep adjoining more vectors $W_{s+1}, W_{s+2}, \ldots \in \mathcal{V}$ (if you can), while maintaining the independence of your collection. By the Corollary, this process cannot go beyond a total of n vectors. At some point, therefore, your set $\{W_1, \ldots, W_{s+p}\}$ must stop being enlargeable; i.e. any additional vector $V \in \mathcal{V}$ must be a linear combination of the ones you already have.

This result also shows that $\dim \mathcal{V}$ is a meaningful measure of the "size" of $\mathcal{V}$. More precisely, if $\mathcal{V}$ contains a smaller subspace $\mathcal{V}'$, we can enlarge a basis of $\mathcal{V}'$ to one of $\mathcal{V}$, thus proving that $\dim \mathcal{V}' < \dim \mathcal{V}$.

To test your understanding of dimension, try to prove the following identities:

$$n - \dim \mathcal{N}(A) = \dim \mathcal{C}(A) = \dim \mathcal{R}(A),$$

where $\mathcal{C}$ and $\mathcal{R}$ denote the spans of the columns and of the rows, respectively. For the first, take a basis $\{W_1, \ldots, W_k\}$ of $\mathcal{N}(A)$ and extend it to one $\{W_1, \ldots, W_n\}$ of K^n; then show that $\{AW_{k+1}, \ldots, AW_n\}$ is a basis of $\mathcal{C}(A)$. For the second, note that both dimensions are invariant under elementary row and column operations, hence equal to those of $\mathcal{C}(MAN)$ and $\mathcal{R}(MAN)$.

2. Singular Values of Matrices.

It does not seem to be widely appreciated that one of the most central theorems on real (or complex) matrices is also one of the easiest to prove. Its geometric version says that *any linear transformation has the effect of mapping some orthonormal basis of the domain onto an orthogonal subset of the range.* Here is a simple proof of the matrix version.

THEOREM 2.1: Let A be an $m \times n$ real matrix. Then there exist orthogonal matrices M and N such that

$$MAN = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{where} \qquad D = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_r \end{bmatrix} \qquad \text{with} \qquad d_i \geq d_{i+1} > 0.$$

Proof: Let $O(n)$ be the set of all $n \times n$ orthogonal matrices. For $M \in O(m)$ and $N \in O(n)$, let $\alpha(M, N)$ stand for the entry in the first row and first column of MAN. Let d_1 the greatest possible value occurring among these.

Since $O(m) \times O(n)$ is closed and bounded, there is a pair M_1, N_1 for which this value is actually attained. That is, we can obtain that

$$M_1 A N_1 = \begin{bmatrix} d_1 & X \\ Y & A' \end{bmatrix},$$

where A' is an $(m-1) \times (n-1)$-matrix. Now we claim that $X = 0$ and $Y = 0$ are *zero* rows and columns. Indeed, if X were non-trivial, the first row ρ_1 of $M_1 A N_1$ would have length $d > d_1$. Then we could multiply on the right by the reflection H which takes ρ_1 into $[d, 0, \ldots, 0]$ and create a value $\alpha(M, N) = d > d_1$. Similarly $Y = 0$. Obviously, none of the entries of A' can exceed d_1 in absolute value (otherwise *it* could be permuted to the upper left), and this is true for all the possible forms of A'. We are finished by induction.

The real numbers $d_1 \geq, \ldots, \geq d_r > 0$ are known as the *singular values* of A.

UNIQUENESS: The $n \times n$ matrix $B = A^T A$ has a very simple effect on the columns $u_1, \cdots, u_n$ of N, namely, $Bu_i = \mu_i u_i$, where $\mu_i = d_i^2$ for $i \leq r$ and 0 beyond. Indeed, $N^T B N = (MAN)^T MAN = \Delta$ is a diagonal matrix with diagonal entries μ_i as described. Now the identity $BN = N\Delta$ establishes our claim.

To prove *uniqueness* of the singular values d_i it clearly suffices to characterize the μ_i as being the only numbers such that $(B - \mu I)u = 0$ for some $u \neq 0$. But for $u = \sum a_i u_i$, we get $(B - \mu I)u = \sum a_i(\mu_i - \mu)u_i$, which is never 0, unless μ is one of the μ_i.

More geometrically, the d_i can also be retrieved from the image under A of the appropriate unit sphere.

THEOREM 2.2: Every symmetric $n \times n$ matrix A has an invariant one-dimensional subspace.

Proof: Let $u \neq 0$ be one of the columns of N, so that $A^2 u = A^T A u = \mu u$, as above. Put $\mu = \lambda^2$. Then u is annihilated by $A^2 - \mu I = (A + \lambda I)(A - \lambda I)$. If $(A - \lambda I)u = v \neq 0$, then v generates such a line; if $v = 0$ then u does.

For symmetric A it is trivial to show that the orthocomplement of any invariant subspace is itself invariant. Hence, by induction, Theorem 2 provides a set of n mutually orthogonal invariant lines (Spectral Theorem). Moreover, if B is symmetric and commutes with A, it can be restricted to $\ker(A - \lambda I) \neq 0$; therefore the two matrices have a *common* invariant line, hence — by induction — a complete orthogonal set of such.

All arguments on this page go through without a hitch for *complex* matrices if one changes "orthogonal" and "symmetric" to "unitary" and "hermitian", respectively, and replaces the transpose A^T by its complex conjugate A^* (adjoint). Writing a complex matrix as $C = A + iB$ with A, B hermitian, we again get a spectral theorem for C if A, B commute, i.e. if C is *normal*.

It should be noted that neither polynomials nor derivatives are involved in these proofs, not even indirectly.

3. Elementary Divisors.

An integral domain R is *principal* if all its ideals are principal. In particular, given $x, y \in R$, there is a $g \in R$ such that $(x, y) = (g)$; one easily checks that g is a greatest common divisor of x and y. For suitable $\alpha, \beta \in R$, we have $g = \alpha x + \beta y$. Putting $\gamma = -y/g$ and $\delta = x/g$, we have $\alpha\delta - \beta\gamma = 1$, and

$$\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} g \\ 0 \end{bmatrix}.$$

This little observation can be generalized as follows.

LEMMA 3.1: Let A be an $m \times n$-matrix over a principal domain R. Then there exist invertible (over R) matrices M and N such that

$$MAN = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{where} \qquad D = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_r \end{bmatrix} \qquad \text{with} \qquad 0 \neq d_i |\ d_{i+1}.$$

Proof: Let $\mathcal{S}$ be the set of all non-zero entries of MAN as M and N range over all invertible matrices of the appropriate sizes. Take $d_1 \in \mathcal{S}$ with (d_1) maximal. By definition, we then have

$$M_1 A N_1 = \begin{bmatrix} d_1 & X \\ Y & A' \end{bmatrix},$$

where X is a row, Y is a column, and A' is an $(m-1) \times (n-1)$-matrix. We claim that $X, Y \equiv 0$ modulo d_1. Indeed, if y is any non-zero entry of Y, we may assume that it is in the second row. We then multiply on the left by an invertible $m \times m$-matrix whose top left corner is the 2×2-matrix of our preamble — with d_1 playing the role of x. As a result of this move, the g.c.d. of d_1 and y pops into our matrix, contradicting the maximality of (d_1), unless $(d_1, y) = (d_1)$ as claimed. Hence we can make $Y = 0$ by elementary row operations, and similarly (using transposes) $X = 0$. Assuming this done, we can conclude that all entries of A' are divisible by d_1, because any one of them can be made to appear in the first column by a suitable addition of columns, thus playing the role of the y above. The proof is finished by induction.

NOTE 3.2: The following general observations apply to *any* commutative ring R.

(a) Let $\mathcal{C}(A)$ be the submodule of R^m generated by the columns of the matrix A. If M, N are invertible matrices over R, then $\mathcal{C}(A) = \mathcal{C}(AN) \cong \mathcal{C}(MAN)$, the isomorphism being induced by left multiplication by M.

(b) If V, W are Noetherian R-modules (i.e., every submodule finitely generated), then so is $V \oplus W$. To see this, let $U \subset V \oplus W$ be a submodule, and consider the modules $V' = \{v \in V |\ (v, 0) \in U\}$ and $W' = \{w \in W |\ (v, w) \in U \text{ for some } v \in V\}$. Generators of these can be dragged into U in an obvious way; there they generate everything.

Back to the principal domain R. Since every ideal is principal, repeated application of (b) shows that every submodule $V \subset R^m$ is finitely generated. Letting its generators form the columns of a matrix A, we get $V = \mathcal{C}(A) \cong \mathcal{C}(MAN)$, as in (a). Choosing M, N as in the Lemma we arrive at the following conclusion.

LEMMA 3.3: Every submodule of R^m is isomorphic to R^r with $r \leq m$.

THEOREM 3.4: Every finitely generated R-module is isomorphic to $R/(d_1) \oplus \cdots \oplus R/(d_r) \oplus R^s$, with $0 \neq d_i |\ d_{i+1}$.

Proof: The fact that the module W in question is finitely generated means that there is an epimorphism $S : R^m \to W$, and hence $W \cong R^m/V$ with $V = \ker(S)$. As in the discussion preceding the proposition, let $V = \mathcal{C}(A)$. Then, with M, N as in the lemma, (a) says that M induces an isomorphism $R^m/V \cong R^m/\mathcal{C}(MAN)$, which has the desired form ($s = m - r$).

4. Uniqueness.

The finite descending chain of ideals (d_i) occurring in Theorem 3.4 turns out to be uniquely determined by the isomorphism class of the module

$$W = R/(d_1) \oplus \cdots \oplus R/(d_r) \oplus R^s. \tag{1}$$

Therefore it is also uniquely associated with any matrix A for which $R^m/\mathcal{C}(A) \cong W$. It is called the set of *elementary divisors* of A. The idea of the uniqueness proof is simple: split W (uniquely) into primary components, then prove the uniqueness of the "cyclic" decomposition for these, and finally reassemble the primary pieces, using the chain property to force uniqueness. The details of this program can create a notational nightmare, so (for once) let us indicate how to wade through them. To start with, we note that the first r terms of (1) constitute a well-defined submodule $W_t = \{w \in W |\ aw = 0 \text{ for some } 0 \neq a \in R\}$, called the *torsion* submodule, and that $(d_r) = \{a \in R |\ aW_t = 0\}$ gives an intrinsic definition of (d_r).

Since every ideal of R is principal, every ideal generated by an irreducible element $p \in R$ is maximal. Hence, for any $d \in R$, either $d \in (p)$ or $(d, p) = 1$. In particular, $K(p) = R/(p)$ is a field, and p is prime; i.e. $p |\ ab \Rightarrow p |\ a$ or $p |\ b$. A standard argument shows that, for $0 \neq x \in R$, there is a unique set of non-negative integers $v_p(x)$, almost all zero, such that $x = \epsilon \prod p^{v_p(x)}$, where ϵ is a unit, and (p) runs over all prime ideals.

Before going all out, we shall prove a weak uniqueness result, which is still strong enough for many applications.

PROPOSITION 4.1: Let p_t denote the multiplication $W_t \to W_t$ by the prime $p \in R$. Then p_t is an isomorphism if $(p, d_r) = 1$, and has a non-zero kernel iff $p |\ d_r$.

Proof: Indeed, $1 = px + d_r y$ makes $w = pxw$, for any $w \in W_t$. On the other hand, suppose $d_r = pb$, and let w be the image of b under $R \to R/(d_r) \to W$. Then $w \neq 0$, but $pw = 0$.

The next lemma will serve as the main tool for proving the uniqueness of the decomposition (1).

LEMMA 4.2: If $V = R/(d)$, then, for every prime $p \in R$ and every integer $k \geq 0$,

$$p^k V/p^{k+1}V \cong \begin{cases} K(p), & \text{if } d \in (p^{k+1}); \\ 0, & \text{otherwise.} \end{cases}$$

Proof: Since V is a cyclic R-module, $p^k V/p^{k+1}V$ is a cyclic $K(p)$-module, hence either 0 or $\cong K(p)$.

It is 0 if and only if $p^k V = p^{k+1}V$, which is equivalent to saying that $p^k \in (p^{k+1}, d)$, i.e. that $p^k = p^{k+1}a + db$ for suitable $a, b \in R$. If $d \in (p^{k+1})$, this is impossible, since the expression on the right is then divisible by p^{k+1}. If $d = p^l q$ with $(p, q) = 1$ and $l < k+1$, however, we can multiply the equation $1 = p\alpha + q\beta$ by p^k to obtain $p^k = p^{k+1}\alpha + d\beta p^{k-l}$.

We apply this lemma to the components of (1) by noting that they contribute 1 to the $K(p)$-dimension of $p^k W/p^{k+1}W$ whenever $k < v_p(d)$ or $d = 0$, and nothing otherwise. Thus we obtain the formula

$$\dim_{K(p)}\ p^k W/p^{k+1}W = \#\{i |\ k < v_p(d_i)\} + s. \tag{2}$$

Now the uniqueness of the chain $(d_1) \supseteq \cdots \supseteq (d_r)$ results from that of the sequences $v_p(d_1) \leq \cdots \leq v_p(d_r)$, which in turn follows from (2) with the help of a little combinatorial fact, to wit:

A finite non-decreasing sequence $\nu_1 \leq \cdots \leq \nu_r$ of non-negative integers can always be retrieved from the (infinite) sequence $\mu_0 \geq \cdots \geq \mu_k \geq \cdots$, where $\mu_k = \#\{i |\ k < \nu_i\}$. Note that the torsion-free rank s of W can be determined first, by using large values of k in (2).

5. Minimal Polynomial.

An $n \times n$ matrix A over a field K obviously generates an algebra $K[A]$ of finite dimension, say m. The ring-homomorphism $K[X] \to K[A]$ taking the indeterminate X to A has a non-trivial kernel consisting of all multiples of a certain polynomial $\mu_A(X) = X^m + a_{m-1}X^{m-1} + \cdots + a_0$ known as the *minimal polynomial* of A. To apply the theory developed on pp. 3-4, put $R = K[X]$ and let W denote the R-module K^n on which X acts like A. Since $\mu_A W = 0$, we have $W = W_t$ and — in the notation of p.3 — $d_r = \mu_A$. Since $K[A] \cong R/(d_r)$ is a direct summand of W, we get

(1) $\deg \mu_A \leq n$.

If $p(X) \in K[X]$ is irreducible, Proposition 4.1 says that

(2) $\ker p(A) \neq 0$ if and only if $p(X) |\ \mu_A(X)$.

In particular, $A - \lambda I$ is singular if and only if $\mu_A(\lambda) = 0$. Such λ are called *eigenvalues* of A.

For our last result we need another general fact about principal ideal domains R, namely that the obvious injection $R/(ab) \to R/(a) \oplus R/(b)$ is an *isomorphism* if $(a, b) = 1$. Indeed, $1 = ax + by$ makes $rby + sax \equiv r \bmod a$ and $\equiv s \bmod b$, with r and s given arbitrarily.

Now suppose that the module W is indecomposable over R. Then by 3.4, W must be cyclic, say $\cong R/(d)$. By what we have just seen, we cannot have $d = p^l q$ with q relatively prime to p, unless q is a unit. Hence $W \cong R/(p^l)$, with p irreducible. If $p(x) = X - \lambda$, then W has a basis of the form $w, (a - \lambda I)w, \cdots, (A - \lambda I)^{l-1}w$, and $(A - \lambda I)^l = 0$. Therefore $A - \lambda I$ is similar to the $l \times l$ matrix N with 1's just below the diagonal and 0's elsewhere. A matrix of the form $\lambda I + N$, with such an N, is called a *Jordan block*.

(3) If all the roots of $\mu_A(X)$ lie in K, then A is similar to a direct sum of Jordan blocks.

Comments.

The preceding pages contain the main theorems of of strictly linear (as opposed to multi-linear) algebra. The apparent ommissions can be easily filled in from what is there.

Lemma 3.1 could be obtained more simply if R is a Euclidean ring: one would substitute the division algorithm for the prefatory 2×2 result. This would still cover the major applications, namely $R = \mathbf{Z}$ and $R = K[X]$.

It is worth noting that the matrices M and N in 3.1 may be assumed to have determinants ± 1.

The uniqueness proof sketched on page 2 is the trickiest of these topics. Instead of wanting to know the uniqueness of the cyclic components of the decomposition (1), one is usually content with that of the elementary divisors of a given A. These can be also be obtained via determinants of submatrices — but of that later.

6. Determinants.

(following Lang following Artin)

Let R be a commutative ring and consider a map $D : \mathcal{M}_{n\times n}(R) \to R$ defined on $n \times n$ matrices, such that

(i) D is n-linear with respect to the matrix *columns*,

(ii) $D(A) = 0$ if any two adjacent columns of A are equal.

It easily follows that D changes sign whenever two adjacent columns are switched. Hence $D(A) = 0$ also when non-adjacent columns are equal. The reason for formulating (ii) so modestly will be apparent in the existence proof below.

THEOREM: There is exactly one function D with (i) and (ii) and such that $D(I_n) = 1$, namely

$$D(A) = \sum_{i_1,\dots,i_n} \epsilon(i_1,\dots,i_n) a_{i_1 1} \cdots a_{i_n n}, \tag{1}$$

where $i_1, \dots, i_n$ runs over all permutations of $\{1, \dots, n\}$, and $\epsilon = \pm 1$ depending on whether the permutation at hand requires an even or odd number of switches.

Proof: We proceed by induction, the case $n = 1$ being trivial. Given an $n \times n$ matrix $A = (a_{ij})$, let A_{ij} denote the submatrix obtained by deleting row i and column j. Fixing a *row* index i, put

$$\Delta(A) = \sum_j (-1)^{i+j} a_{ij} D(A_{ij}). \tag{2}$$

The D appearing on the right is the unique function with the desired properties for $(n-1)\times(n-1)$ matrices, which exists by induction hypothesis. The Δ on the left — which so far depends on i — will be seen to satisfy (i) and (ii). As to (i), we note that each term $a_{ij}D(A_{ij})$ is n-linear in the columns of A, because each column (in truncated form) occurs exactly once in it.

Now suppose that columns k and $k+1$ are equal. For $j \neq k, k+1$ both these columns are present in A_{ij}, and hence $D(A_{ij}) = 0$. Therefore the expression (2) boils down to

$$\Delta(A) = (-1)^{i+k} a_{ik} D(A_{ik}) + (-1)^{i+k+1} a_{i,k+1} D(A_{i,k+1}).$$

This is 0 because $a_{ik} = a_{i,k+1}$ and $A_{ik} = A_{i,k+1}$, see?

Finally, if $A = I_n$, then $A_{ii} = I_{n-1}$, and (2) says that $\Delta(I_n) = D(I_{n-1})$, which is 1 by induction.

To prepare for the proof of uniqueness, let B be an $n \times n$ matrix with columns $B^1, \dots, B^n$, and Δ be *any* old function satisfying (i),(ii), and $\Delta(I_n) = 1$. By multilinearity, we have

$$\Delta(BA) = \Delta(\sum_i a_{i1} B^i, \dots, \sum_i a_{in} B^i) = \sum_{i_1,\dots,i_n} a_{i_1 1} \cdots a_{i_n n} \Delta(B^{i_1}, \dots, B^{i_n}),$$

where $i_1, \dots, i_n$ ranges over all n-tuples of indices. However, $\Delta(B^{i_1}, \dots, B^{i_n}) = 0$ if the n-tuple contains repeats; hence we need only consider permutations, and for these we can reshuffle the columns of B into their original order, at the price of the factor ϵ. Altogether,

$$\Delta(BA) = D(A)\Delta(B), \tag{3}$$

with $D(A)$ as in (1). Setting $B = I_n$ shows $\Delta = D$ and proves the desired uniqueness.

REMARK: Two more basic facts need to be recorded. A hard look at (1) reveals that

$$D(A^T) = D(A). \tag{4}$$

Finally, forming the matrix $\tilde{A}$ by setting $\tilde{a}_{ij} = (-1)^{i+j} D(A_{ij})$, we obtain from (2) that

$$A\tilde{A} = D(A) I_n. \tag{5}$$